

**MODUL KULIAH**  
**STATISTIKA DASAR**



Oleh:  
Dr. Molli Wahyuni, S.Si, M.Pd

**PROGRAM STUDI GIZI**  
**FAKULTAS ILMU KESEHATAN**  
**UNIVERSITAS PAHLAWAN TUANKU TAMBUSAI**  
**2020**

# **BAB I**

## **PENDAHULUAN STATISTIKA**

### **1.1 Beberapa istilah dasar**

Kita sangat akrab dengan kata “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. Berikut definisi beberapa istilah tersebut:

**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 dikumpulkan. Data dapat berupa hasil pengukuran; misalnya data tinggi dan berat badan, hasil pengelompokan; misalnya jenis kelamin, hasil jawaban responden terhadap suatu questioner; 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 sebagian koleksi anggota yang dipilih dari 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 INFERENSI** adalah statistika yang mencoba untuk membuat suatu deduksi atau kesimpulan pada populasi dengan menggunakan sampel dari populasi tersebut.

Sebagai contoh, suatu lembaga survey melakukan wawancara terhadap 2350 penduduk Indonesia untuk mengetahui tingkat kepuasan terhadap kinerja pemerintah. Dalam hal ini sebanyak 2350 penduduk merupakan **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 kelakuan 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 mengetahui aspirasi dari semua pemilih. Jadi pemilu merupakan proses sensus untuk populasi pemilih; walaupun kenyataannya tidak semua data populasi dapat diperoleh karena banyaknya “golput”. Sedangkan, lembaga survey yang melakukan perhitungan 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 diambil 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 merupakan 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 mempunyai peluang kejadian sebaliknya. Sangat dimungkinkan beberapa lembaga survey perhitungan cepat pilkada memberikan kesimpulan yang berbeda satu sama lainnya; terutama bila keadaan sesungguhnya hanya memberikan selisih yang sangat tipis. Masih ingat dengan kasus pilkada Jawa Timur beberapa waktu yang lalu?

## 1.2 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 diperoleh 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 contoh di atas, terdapat pula data dalam bentuk kategori. Kedua bentuk data ini didefinisikan secara formal sebagai berikut :

**DATA KUANTITATIF** adalah data yang menggambarkan hasil perhitungan atau hasil pengukuran.

**DATA KUALITATIF** atau **DATA KATEGORI** adalah 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 banyak kemungkinannya takterbilang.

## CONTOH

1. **Data diskrit:** jam kerja dalam sehari (kemungkinannya adalah 1, 2, 3, ... , 24), banyak telur 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).

### 1.3 Level Pengukuran

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

**LEVEL 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 level 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 level ini tidak dapat diurutkan. Data ini tidak dapat digunakan untuk kalkulasi, misalnya Ya + tidak tahu = ???, merah + hitam = ??? tidak dapat dilakukan.

**LEVEL ORDINAL** data yang diperoleh pada level 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 level 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 ranking kedua, Banda Aceh pada ranking ketiga dan seterusnya sampai Kupang pada ranking ke 50 atau terkorup. Data ranking di sini merupakan level pengukuran ordinal. Walaupun ada angka di sini namun selisih antara ranking 2 dan ranking 1, bila ditulis dalam bentuk  $2-1 = 1$  tidak mempunyai makna sama sekali.

**LEVEL INTERVAL** seperti level ordinal dengan sifat tambahannya adalah selisih antara dua data mempunyai makna. Tetapi level ini tidak mempunyai **titik nol** alami sebagai titik awal.

**CONTOH:** Berikut ini data dalam level interval

1. **TEMPERATUR:** suhu badan 36 derajat celsius dan 37 derajat celsius merupakan contoh data dalam level interval. Nilai-nilai ini dapat diurutkan dan selisihnya dapat ditentukan dengan jelas, 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 merupakan data dalam level interval. Data ini dapat diurutkan dan dapat diketahui 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”.

**LEVEL RASIO** seperti level interval namun ia mempunyai titik nol alami sebagai titik awal. Data dari level rasio, data data dapat dibandingkan (selisih) dan dirasioakan (pembagian).

**CONTOH:** Berikut inidata dalam level rasio

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

#### **SOAL-SOAL LATIHAN:**

1. [STATISTIK dan PARAMETER] Identifikasilah 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.
2. [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 disurvei ditemukan 38% dari mereka adalah perokok aktif.
  - c. Suatu sampel terdiri dari sejumlah mobil, ditemukan bahwa rata-rata beratnya adalah 1500 kg.
3. [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.
4. [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,
5. [INTERPRETASI POOLING POLITIK] Misalkan 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 ?
6. [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 choice.

**LEMBAR KERJA :**

1).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

2).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

3).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

4).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

5).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

6).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

## BAB II

### DISTRIBUSI FREKUENSI

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 alfabet 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 distribusi 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.

#### 2.1 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 merupakan 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 interval 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  $(66+68)/2 = 67$ .

## 2.2 Prosedur umum membuat tabel frekuensi

Berikut langkah-langkah untuk membuat tabel frekuensi:

1. Tetapkan data terbesar dan data terkecil, kemudian tentukan rangenya.
2. Bagilah range ini ke dalam sejumlah interval kelas yang mempunyai 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 merupakan data observasi sesungguhnya. Hal ini untuk mengurangi apa yang disebut dengan *grouping-error*. Namun batas kelas sebaiknya tidak sama dengan data observasi.
3. Hitung lebar interval kelas  $d = \frac{\text{range}}{\text{banyak interval kelas}}$ . Kalau diperlukan dapat dibulatkan.

4. **Starting point:** mulailah dengan bilangan limit bawah untuk kelas interval pertama. Dapat dipilih sebagai data terkecil dari observasi atau bilangan di bawahnya.
5. Dengan menggunakan limit bawah interval kelas pertama dan lebar interval kelas, tentukan limit bawah interval kelas lainnya.
6. Susunlah semua limit bawah interval kelas secara vertikal, kemudian tentukan limit atas yang bersesuaian.
7. 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 diambil 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 distribusi 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  $\bar{x}_d = \frac{\sum f_i d_i}{\sum f_i}$  dimana  $f_i$  adalah frekuensi kelas ke  $i$  dan  $d_i$  adalah tanda

kelas ke  $i$ .

- 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 4<sup>th</sup> ed” oleh Murray Spiegel dan Larry Stephens harus sudah dimiliki dan soal-soalnya dikerjakan dan pahami.

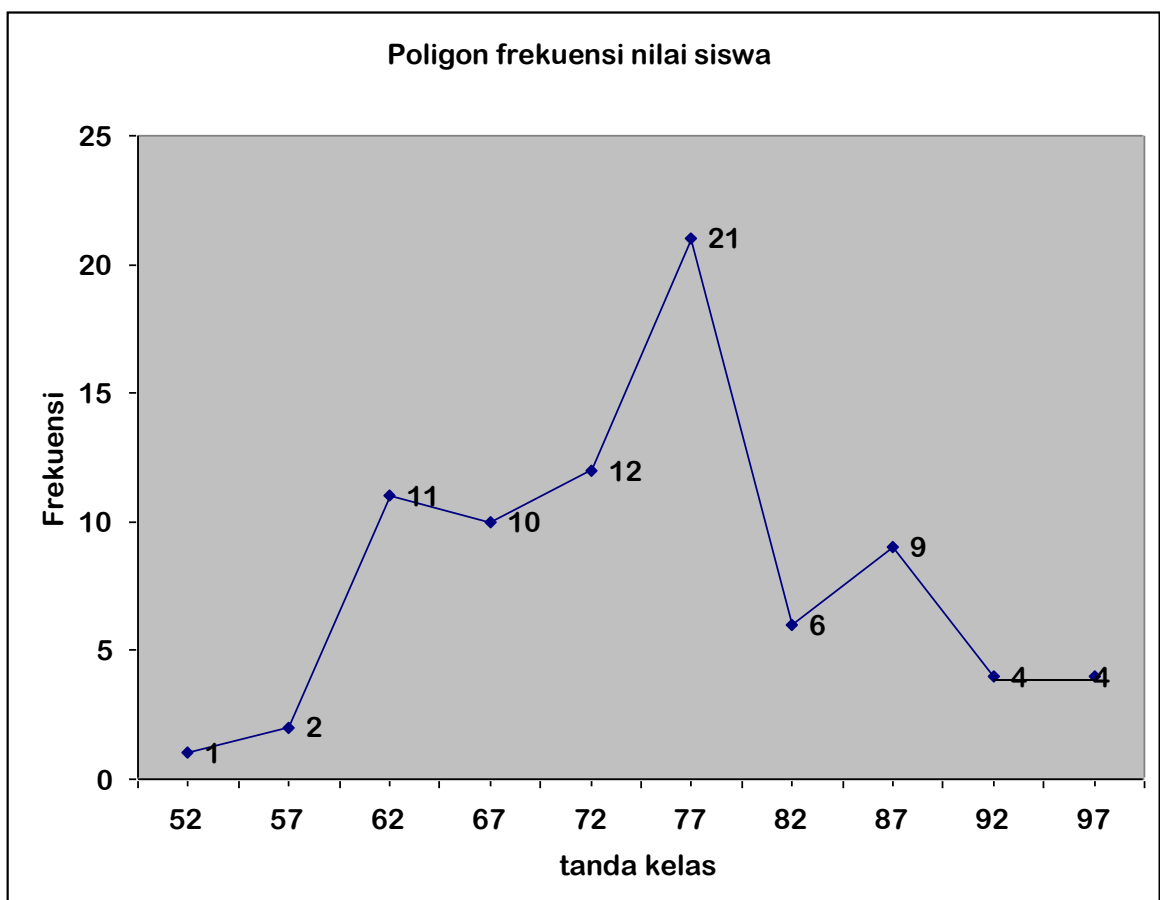
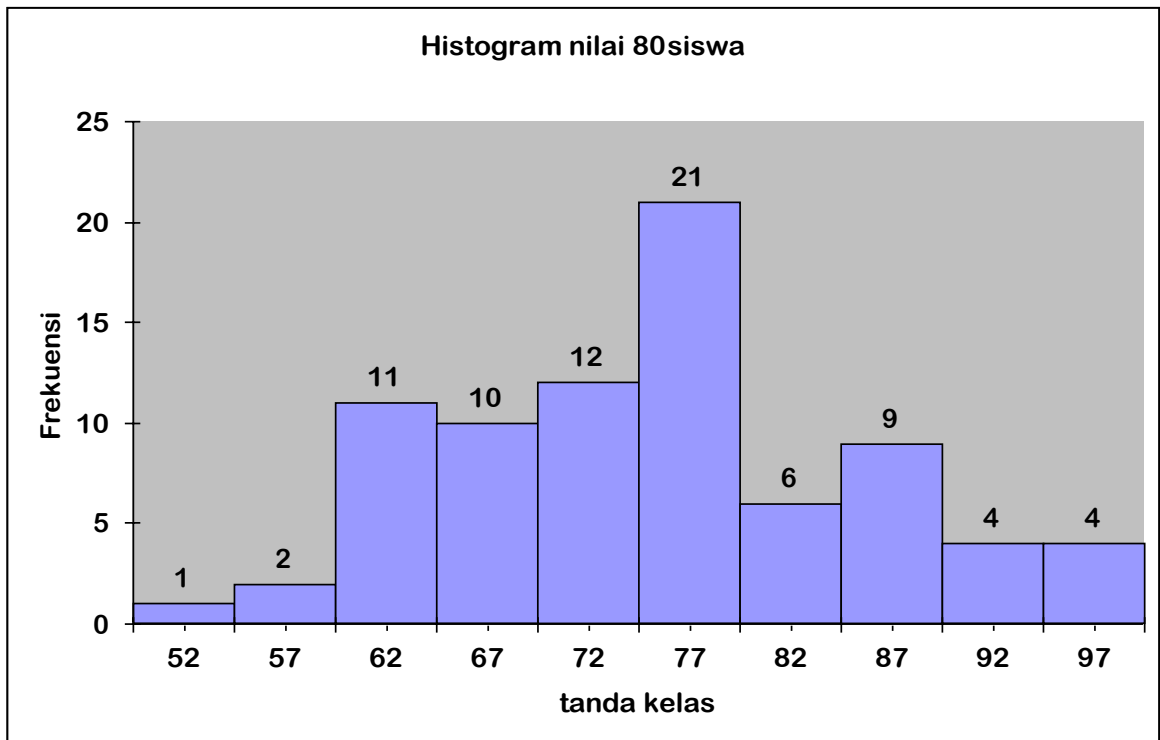
### 2.3 Histogram dan poligon frekuensi

Histogram dan poligon frekuensi merupakan 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 merupakan 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.



**PENGUNAAN 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 dengan 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 diinstall paket statistiknya yang biasanya sebagai opsional pada CD asalnya. Langkah-langkah mengupdate excel dengan paket analisis data statistik:

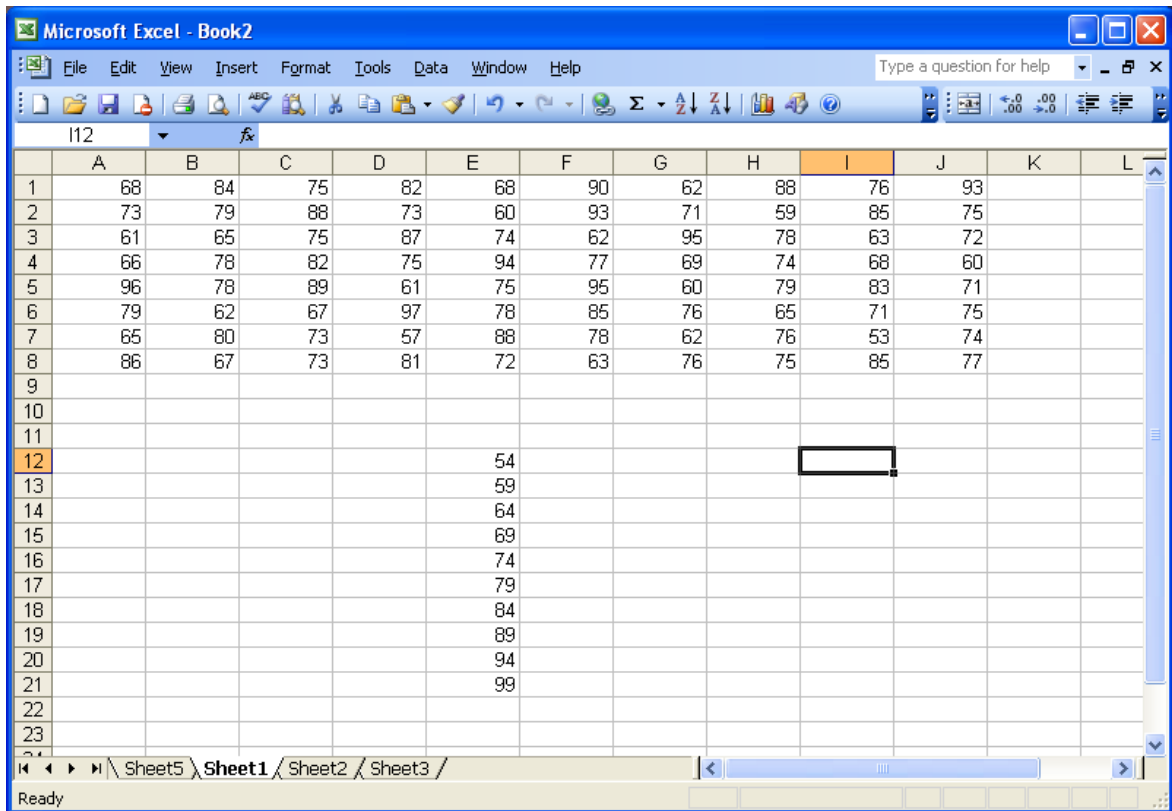
1. Siapkan CD dari mana office diinstall
2. Pada menu pilih **Tools - Add-Ins**.
3. Pada papan dialog Add-Ins, pilih **Analysis ToolPak** dan **Analysis ToolPak-VBA**.
4. Klik OK.

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.

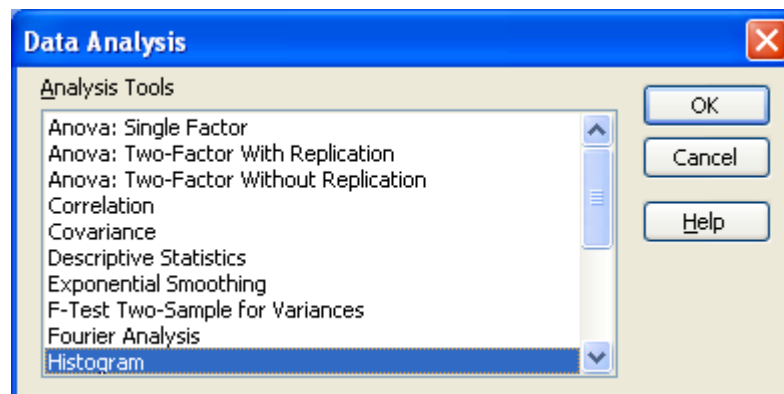
Dengan asumsi anda sudah bisa bekerja dengan excel paling tidak sudah dapat memasukkan data ke dalam lembar kerja excel, maka berikut ini diberikan langkah-langkah membuat tabel distribusi frekuensi, histogram dan poligon frekuensi dalam satu paket.

#### **Langkah-langkah:**

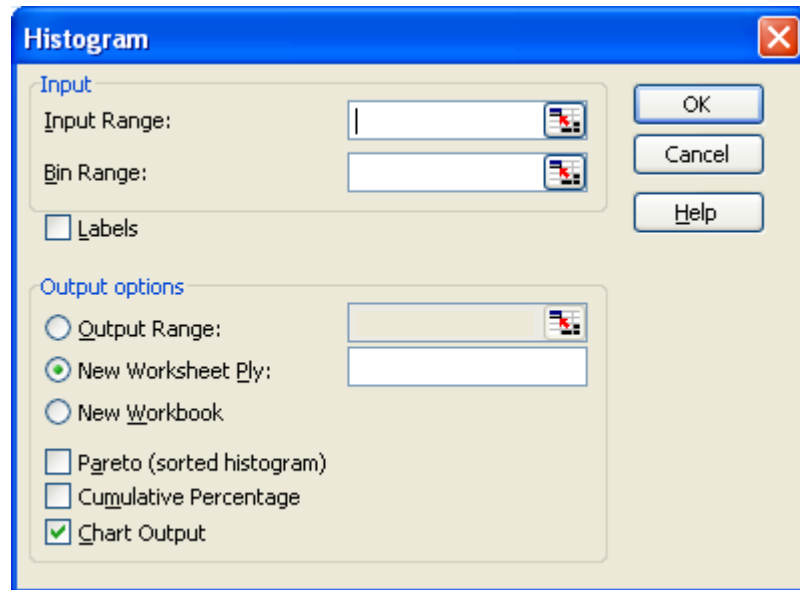
1. Buka Excel dan masukkan data mentah ke dalam sel-sel yang tersedia, misalnya terlihat pada tampilan berikut.
2. Buat array terpisah untuk memasukkan limit atas masing-masing interval seperti terlihat pada tampilan berikut.



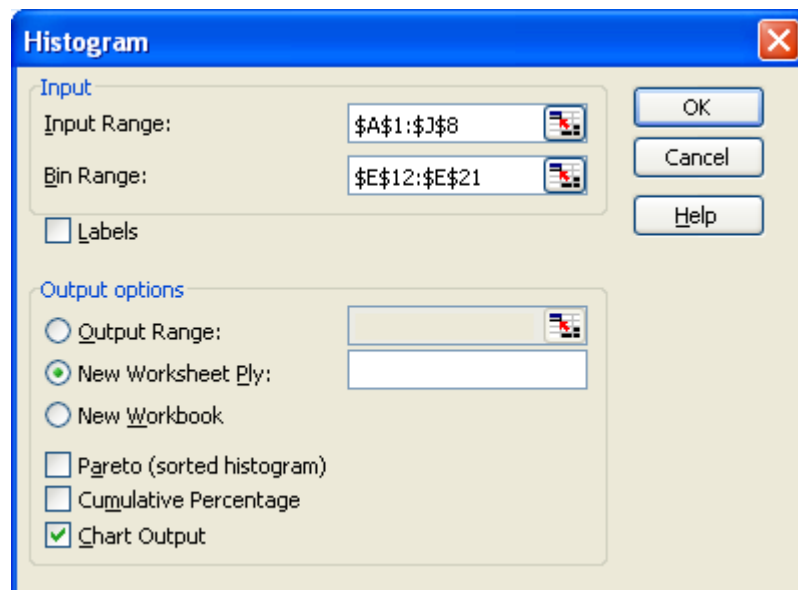
3. Lakukan langkah-langkah analisis data sebagai berikut:
  - a. Melalui menu Tools, pilih Data analysis, kemudian muncul pilihan berikut.
  - b. Pilih Histogram, klik Ok.



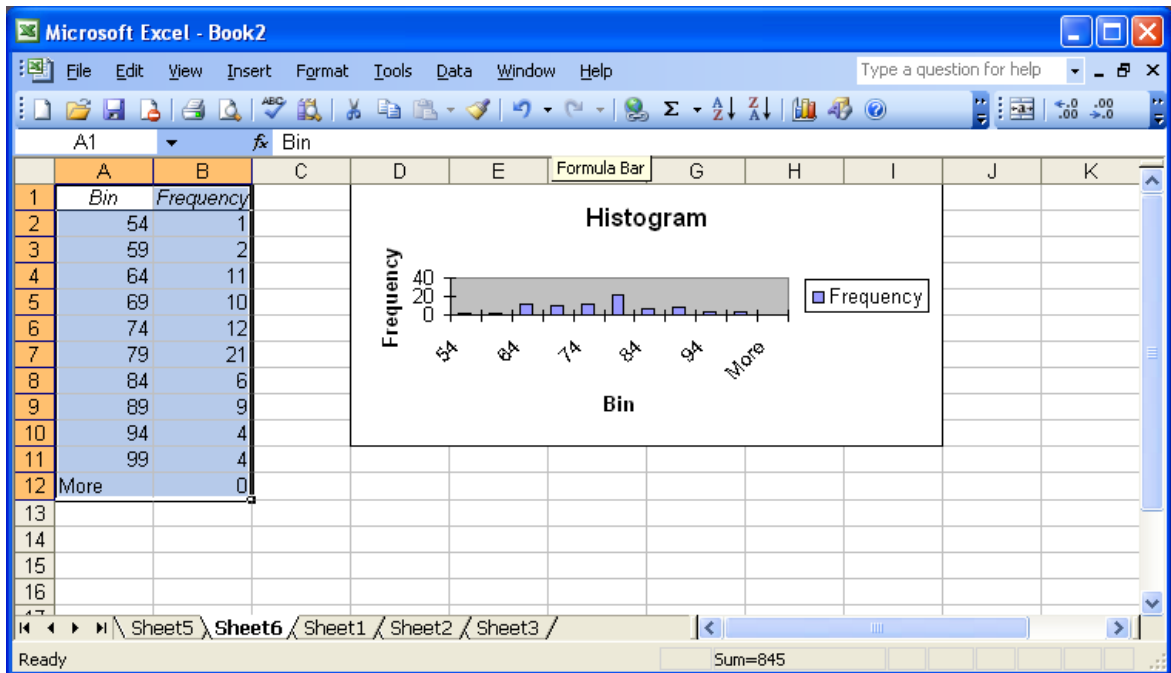
Diperolehlah tampilan berikut:



- 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 berikut.

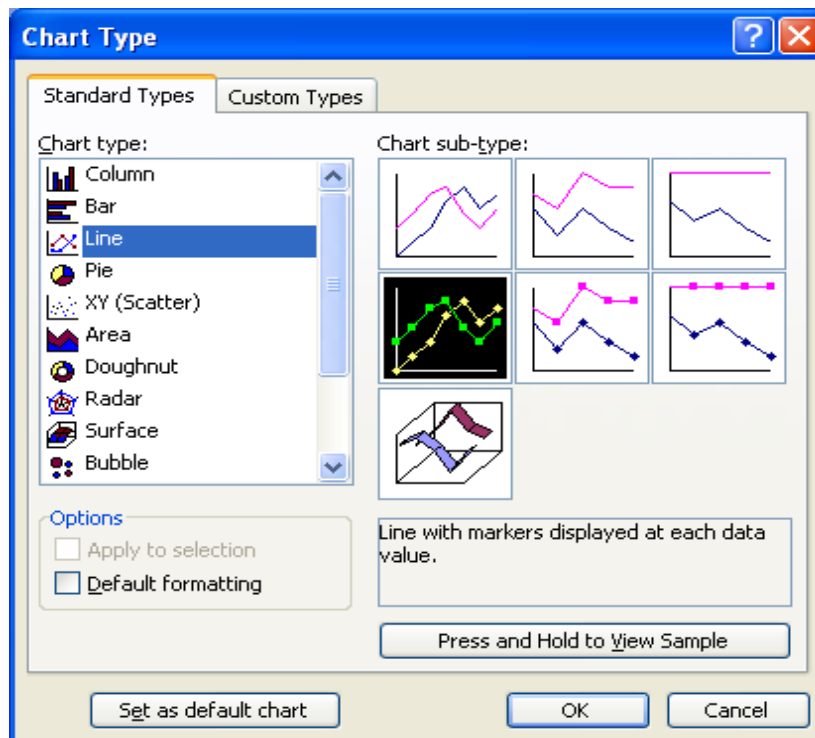


- e. Klik Ok. Setelah itu akan muncul tampilan berikut:



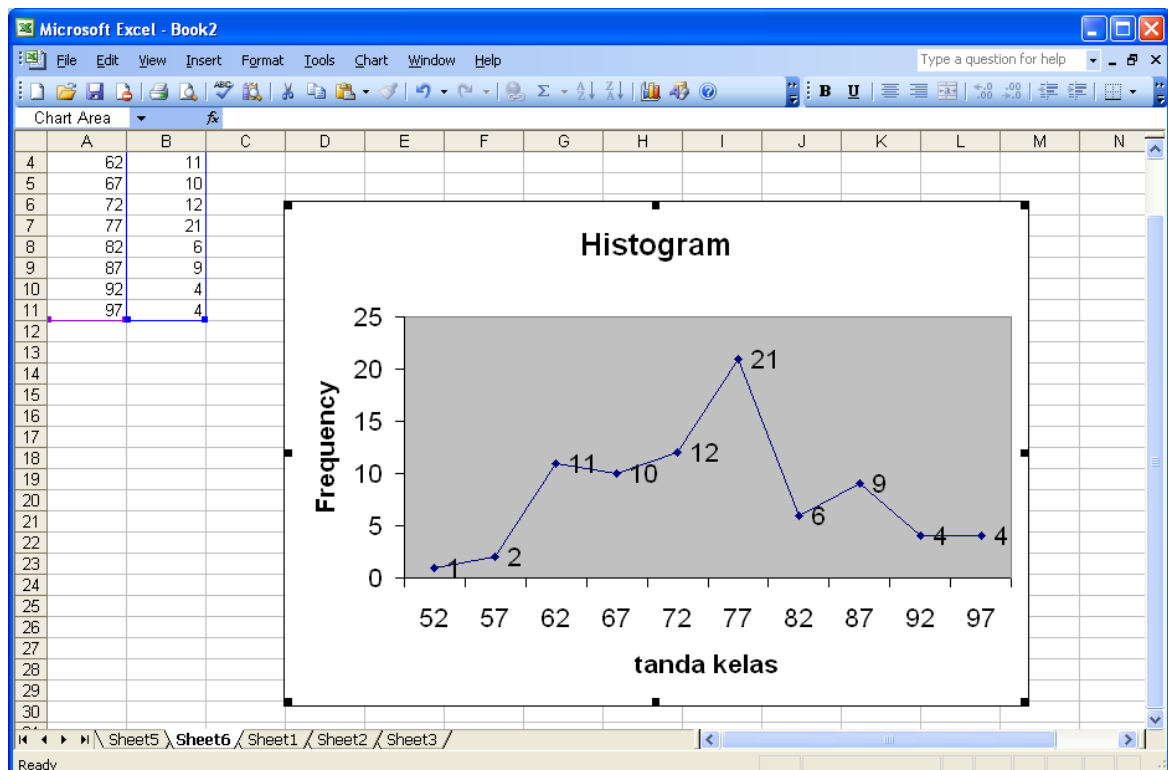
- f. Dengan melakukan editing pada histogram yang ada seperti memperbesar, menggeser, mengubah label, font, warna dan lain-lain maka akan diperoleh histogram yang diinginkan. Lakukanlah dengan coba-coba sambil mempelajari materi excel lebih lanjut.
- g. Untuk menampilkan tanda kelas (titik tengah interval) 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.



## 2.4 Distribusi frekuensi kumulatif, relatif dan ogive

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

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

**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%

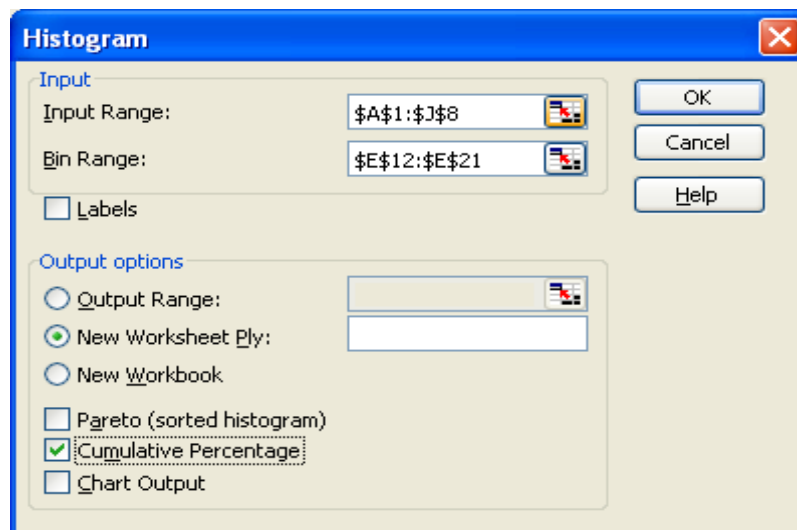
Tabel 2.6 Distribusi frekuensi kumulatif nilai matematika 80 siswa SMA XYZ

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		

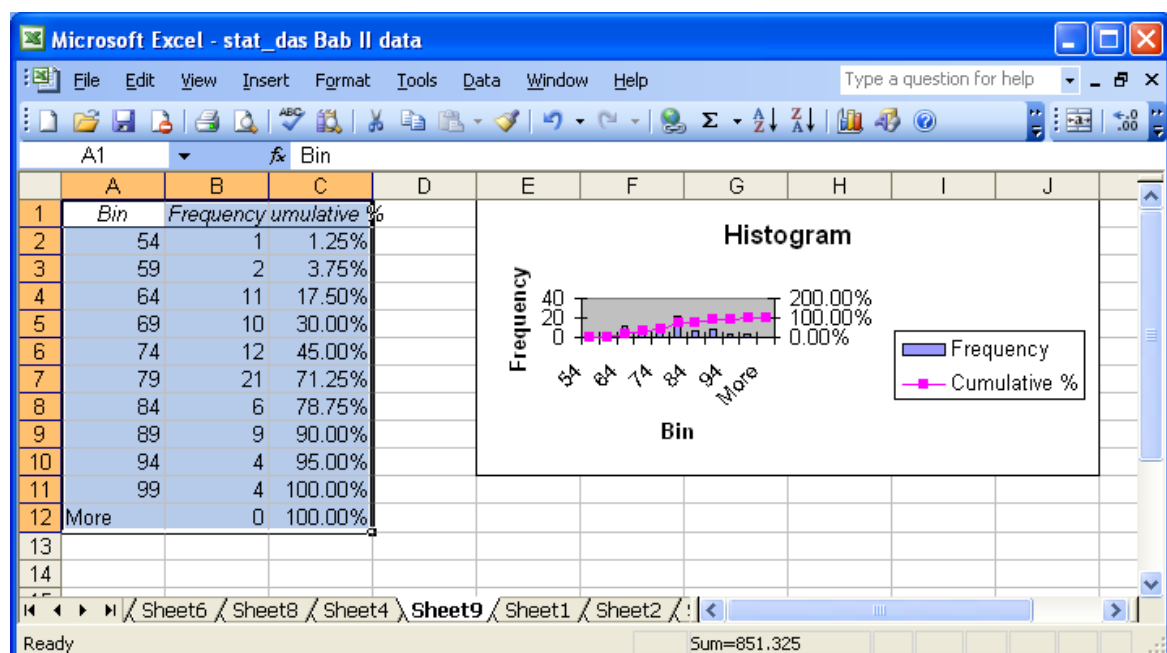


Diperhatikan bahwa frekuensi kumulatif 24 pada kelas 65-69 diperoleh dari  $1+2+11+10$ . Grafik yang menyajikan distribusi kumulatif ini disebut **ogive**. Untuk membuat ogive dengan excel, ikuti langkah-langkah berikut:

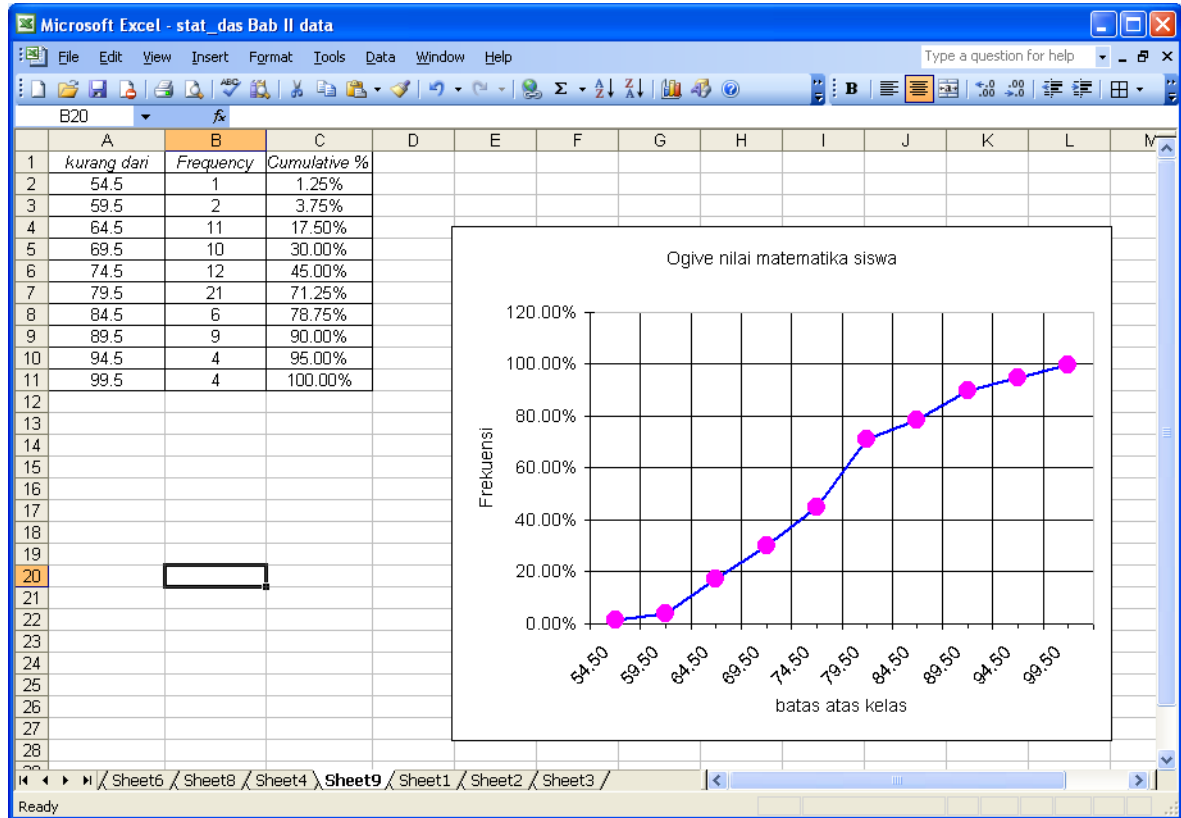
1. Anggaplah data nilai matematika siswa dan limit atas semua kelas sudah dimasukkan ke dalam workshet. Langkah ini sudah dipelajari ketika membuat histogram.
2. Melalui menu **Tools**, pilih **Data Analysis**, pilih **histogram**.
3. Pada **Output option**, pilih **Cumulative Percentage** dan **Chart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



Diperolehlah output sebagai berikut:



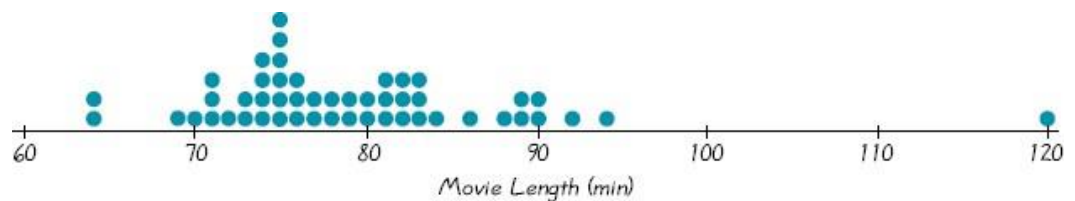
Setelah dilakukan editing, seperti membuang baris More pada Tabel, menghapus histogram frekuensi, menggeser, memperbesar, mengganti judul header, judul sumbu koordinat, dan lain-lain maka diperoleh tampilan yang lebih menarik berikut.



## 2.4 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 terpercil 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.

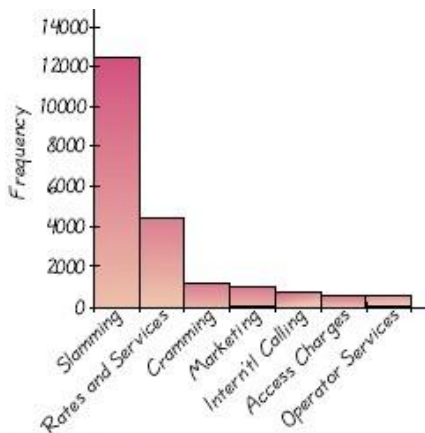


FIGURE 2-6 Pareto Chart of Phone Company Complaints

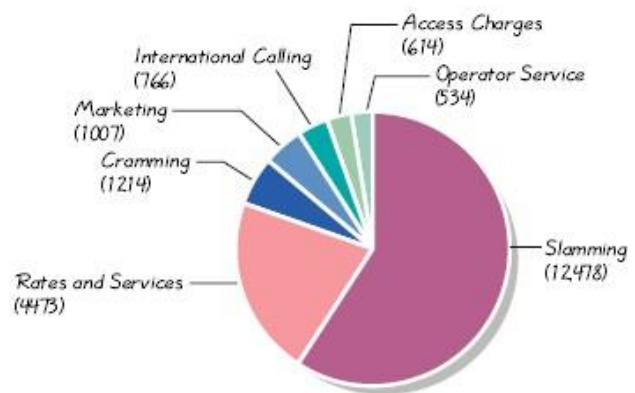


FIGURE 2-7 Pie Chart of Phone Company Complaints

## 3. Diagram kue (Pie)

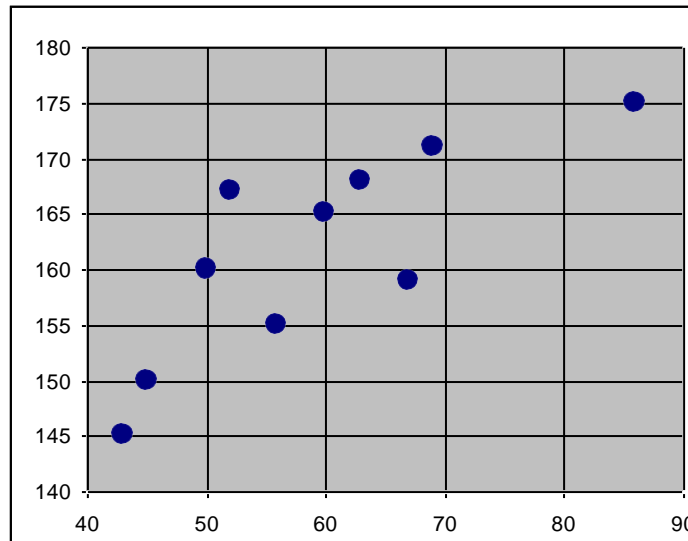
Ini adalah bentuk penyajian data kualitatif dalam bentuk potongan kue. Potongan kue dibuat proposional. Lihat contoh di atas.

## 4. Diagram pencar (scatter)

Diagram pencar ini digunakan untuk menyajikan pasangan data (x,y). Dengan melihat tampilan pada diagram pencar maka dapat diketahui secara umum bentuk hubungan antara dua kelompok data. Misalkan 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:

- nilai tertinggi.
- nilai terendah.
- rentang nilai.
- nilai-nilai yang menduduki ranking 5 terbesar.
- nilai-nilai yang menduduki ranking 5 terkecil.
- banyak siswa yang mendapat nilai tidak kurang dari 75.
- banyak siswa yang mendapat nilai kurang dari 85.
- prosentasi siswa yang mendapat nilai lebih dari 65 tetapi tidak lebih dari 85.
- 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

- limit bawah kelas ke 4.
- limit atas kelas ke 5.
- tanda kelas kelas ke 3.
- batas-batas kelas ke 6.
- lebar kelas ke 5.
- frekuensi kelas ke 2.
- frekuensi relatif kelas ketiga.
- kelas interval yang mempunyai 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

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

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

4. Misalkan 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 tampak dan 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 tabel 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

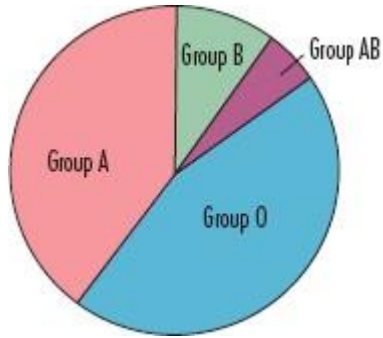
With this reference of table determine:

- The upper limit of the fifth class
- The lower limit of the eighth class
- The class mark of the seventh class
- The class boundaries of the last class
- The class-interval size
- The frequency of the fourth class
- The relative frequency of the sixth class
- The percentage of students whose weekly viewing time does not exceed 600 minutes
- The percentage of students with viewing times greater than or equal to 900 minutes
- 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 approximate 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 chart is based on sample of 500 people, approximately how many of those 500 people have group B blood?

**LEMBAR KERJA :**

1).....

.....

.....

.....

.....

.....

.....

.....

.....

.....

2).....

.....



.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

3).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

4).....  
.....

.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

5).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

6).....  
.....  
.....

.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

7).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

8).....  
.....

.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

## BAB III UKURAN PEMUSATAN DATA

Misalkan kita mempunyai data mentah dalam bentuk array  $X = X_1, X_2, \dots, X_n$ . Pada Bab ini kita akan mempelajari beberapa ukuran yang dapat memberikan informasi tentang bagaimana data-data ini mengumpul atau memusat.

### 3.1 Notasi sigma dan sifat-sifatnya

Sebelumnya kita pahami dulu notasi jumlah berikut:

$$\sum_{j=1}^n X_j = X_1 + X_2 + \dots + X_n.$$

Notasi lainnya adalah  $\sum X = \sum_{j=1}^n X_j$ .

**CONTOH :** Misalkan diberikan array  $X = 2, 3, 5, 7$ . Disini  $X_1=2, X_2=3, X_3=5, X_4=7$ . Diperoleh  $\sum X = 2 + 3 + 5 + 7 = 17$ .

#### Sifat-sifat :

$$1. \sum_{j=1}^N X_j Y_j = X_1 Y_1 + X_2 Y_2 + X_3 Y_3 + \dots + X_N Y_N$$

$$2. \sum_{j=1}^N a X_j = a X_1 + a X_2 + a X_3 + \dots + a X_N = a \sum_{j=1}^N X_j$$

$$\text{Jadi } \sum_{j=1}^N a X_j = a \sum_{j=1}^N X_j$$

### 3.2 RATA-RATA ATAU UKURAN PEMUSATAN DATA

#### 3.2.1 Mean Aritmatika (Rata-rata Aritmatika/Rata-rata hitung)

➤ Mean aritmatika dari  $N$  data tunggal yaitu  $X_1, X_2, X_3, \dots, X_N$

dinotasikan  $\bar{X}$  dibaca (  $\bar{X}$  bar ) dan didefinisikan :

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N} = \frac{\sum_{j=1}^N X_j}{N}$$

Keterangan :  $\bar{X}$  = Rata-rata

$X_j$  = Data ke-  $j$  dengan  $j = 1, 2, 3, \dots, N$

Contoh : Carilah Mean Aritmatika dari 8, 3, 5, 12, dan 10 !

$$\bar{X} = \frac{8 + 3 + 5 + 12 + 10}{5} = \frac{38}{5} = 7,6$$

➤ Mean Aritmatika Terbobot

Mean aritmatika dari N data tunggal berfrekuensi yaitu  $X_1, X_2, X_3, \dots, X_N$  dengan frekuensi  $f_1, f_2, f_3, \dots, f_N$  disebut mean aritmatika terbobot

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + f_3 X_3 + \dots + f_N X_N}{f_1 + f_2 + f_3 + \dots + f_N} = \frac{\sum_{j=1}^N f_j X_j}{\sum_{j=1}^N f_j}$$

Keterangan :  $\bar{X}$  = Rata-rata

$X_j$  = Data ke- j

$f_j$  = frekuensi ke- j

Contoh :

$X_j$	$f_j$	$X_j f_j$
70	5	350
69	6	414
45	3	135
80	1	80
56	1	56
Jumlah	16	1035

$$\bar{X} = \frac{70 \times 5 + 69 \times 6 + 45 \times 3 + 80 \times 1 + 56 \times 1}{5 + 6 + 3 + 1 + 1} = \frac{1035}{16} = 64,6$$

➤ Mean Aritmatika dari data berdaftar distribusi frekuensi

$$\bar{X} = \frac{\sum_{j=1}^N A_j \cdot f_j}{\sum_{j=1}^N f_j}$$

Keterangan :  $\bar{X}$  = Rata-rata

$A_j$  = Tanda kelas ke- j

$f_j$  = frekuensi kelas ke- j

Untuk membantu menghitung biasanya digunakan table tambahan sebagai berikut :

Rentang nilai	$f_j$	$A_j$	$A_j \cdot f_j$
Jumlah	...	-	...

Contoh :

Tabel 2.4 Distribusi nilai matematika 80 siswa SMA XYZ

Rentang nilai	frekuensi	$A_j$	$A_j \cdot f_j$
50-54	1	52	52
55-59	2	57	114
60-64	11	62	682
65-69	10	67	670
70-74	12	72	864
75-79	21	77	1617
80-84	6	82	492
85-89	9	87	783
90-94	4	92	368
95-99	4	97	388
Jumlah	80	-	6030

$$\bar{X} = \frac{\sum_{j=1}^N A_j \cdot f_j}{\sum_{j=1}^N f_j} = \frac{6030}{80} = 75,375$$

- Cara Sandi

$$\bar{X} = A_0 + d \cdot \left( \frac{\sum f_j \cdot c_j}{\sum f_j} \right)$$

Keterangan :  $\bar{X}$  = Rata-rata

$A_0$  = Tanda kelas dengan  $c = 0$

$d$  = lebar interval kelas

$c_j$  = sandi ( 0,  $\pm 1$ ,  $\pm 2$ , ...)

$f_j$  = frekuensi kelas ke-j

Untuk membantu menghitung biasanya digunakan table tambahan sebagai berikut :

Rentang nilai	$f_j$	$A_j$	$c_j$	$f_j \cdot c_j$
---------------	-------	-------	-------	-----------------

Jumlah	...	-	-	...

Contoh :

Tabel 2.4 Distribusi nilai matematika 80 siswa SMA XYZ

Rentang nilai	frekuensi	$A_j$	$c_j$	$f_j \cdot c_j$
50-54	1	52	-9	-9
55-59	2	57	-8	-16
60-64	11	62	-7	-77
65-69	10	67	-6	-60
70-74	12	72	-5	-60
75-79	21	77	-4	-84
80-84	6	82	-3	-18
85-89	9	87	-2	-18
90-94	4	92	-1	-4
95-99	4	97	0	0
Jumlah	80	-	-	-346

$$\bar{X} = A_0 + d \cdot \left( \frac{\sum f_j \cdot c_j}{\sum f_j} \right) = 97 + 5 \frac{-346}{80} = 97 + -21,625 = 75,375$$

### 3.2.2 Mean Geometrik (G)

Rata-rata Geometrik (G) dari data  $X_1, X_2, X_3, \dots, X_N$  di definisikan :

$$G = \sqrt[N]{X_1 \cdot X_2 \cdot \dots \cdot X_N}$$

Ex : Mean Geometric dari 2, 4, dan 8 adalah  $\sqrt[3]{2 \cdot 4 \cdot 8} = \sqrt[3]{64} = 4$

### 3.2.3 Mean Harmonik (H)

Rata-rata Harmonik (H) dari data  $X_1, X_2, X_3, \dots, X_N$  di definisikan



$$H = \frac{N}{\sum_{j=1}^N \frac{1}{x_j}}$$

Ex : Mean Harmonik dari 2, 4, dan 8 adalah  $\frac{3}{\frac{1}{2} + \frac{1}{4} + \frac{1}{8}} = \frac{3}{\frac{7}{8}} = 3,43$

### 3.2.4 Hubungan antara G, H, $\bar{X}$

Hubungan antara Mean Aritmatika, Mean Geometrik, dan Mean Harmonik adalah :

$$H \leq G \leq \bar{X}$$

## 3.3 MEDIAN

Median menentukan letak data setelah data itu disusun menurut urutan nilainya. Median dari sekumpulan data adalah data tengah setelah seluruh data di susun dari yang terkecil sampai yang terbesar dari seluruh data.

### ➤ Median data tunggal

- Median data tunggal dengan banyak data ganjil

Misal  $X_1, X_2, \dots, X_n, \dots, X_{2n-1}$                       n = bilangan bulat.

$$Me = X_n$$

Contoh : Median dari 3, 7, 6, 5, 4, 3, 3, 2, 5 adalah ... .

$$2, 3, 3, 3, 4, 5, 5, 6, 7 \quad Me = 4$$

- Median data tunggal dengan banyak data genap

Misal  $X_1, X_2, \dots, X_n, X_{n+1}, \dots, X_{2n}$                       n = bilangan bulat

$$Me = \frac{X_n + X_{n+1}}{2}$$

Contoh : Median dari 2, 3, 7, 5, 6, 4, 3, 2 adalah ... .

$$2, 2, 3, 3, 4, 5, 6, 7 \quad Me = \frac{3+4}{2} = 3,5$$

### ➤ Median data berdaftar distribusi frekuensi

$$Me = L_1 + d \left( \frac{\frac{N}{2} - \sum F}{f_{Median}} \right)$$

Keterangan : Me = Median

$L_1$  = Batas bawah kelas median

$d$  = lebar interval kelas

$N$  = banyak data

$\sum F$  = Jumlah frekuensi sebelum interval kelas median

$f_{median}$  = frekuensi kelas median

Contoh :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
Jumlah	80

$$\frac{N}{2} = 40$$

Kelas Median = 75 - 79

$$L_1 = 74,5$$

$$d = 5$$

$$\sum F = 36$$

$$f_{median} = 21$$

$$Me = L_1 + d \left( \frac{\frac{N}{2} - \sum F}{f_{Median}} \right) = 74,5 + 5 \left( \frac{40 - 36}{21} \right) = 74,5 + 0,952 =$$

75,452

### 3.4 MODUS

Modus dari sekumpulan data adalah data yang paling sering muncul/mempunyai frekuensi tertinggi.

➤ Modus dari data tunggal

Example :

- The set 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, and 18 has mode 9 disebut uni modal.
  - The set 1, 1, 1, 1, 1, has mode 1
  - The set 3, 5, 8, 10, 12, 15, and 16 has no mode.

- The set 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, and 9 has two modes, 4 and 7, and is called bimodal

➤ Modus dari data berdistribusi frekuensi

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

Keterangan : Mo = Modus

$L_1$  = Batas bawah kelas modus

$d$  = lebar interval kelas

$\Delta_1$  = frekuensi kelas modus dikurangi frekuensi kelas sebelumnya

$\Delta_2$  = frekuensi kelas modus dikurangi frekuensi sesudahnya

Contoh :

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
Jumlah	80

Kelas Modus 75 - 79

$$L_1 = 74.5$$

$$d = 5$$

$$\Delta_1 = 21 - 12 = 9$$

$$\Delta_2 = 21 - 6 = 15$$

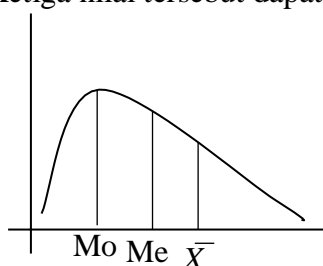
$$Mo = 74.5 + 5 \left( \frac{9}{9+15} \right) = 76,375$$

### 3.5 HUBUNGAN ANTARA MEAN, MEDIAN DAN MODUS

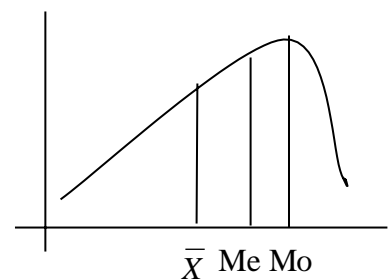
Hubungan antara Mean, Median dan Modus adalah :

$$\text{Mean} - \text{modus} = 3 (\text{Mean} - \text{Median})$$

Ketiga nilai tersebut dapat dilihat sebagai berikut :



Kurva Positif



Kurva negatif

### 3.6 Kuartil, Desil, Persentil

#### ➤ KUARTIL

Jika sekumpulan data dibagi menjadi empat bagian yang sama banyak, sesudah disusun menurut urutan nilainya, maka bilangan pembagiannya disebut **KUARTIL**. Ada tiga buah kuartil yaitu kuartil pertama, kuartil kedua, kuartil ketiga yang masing-masing disimbolkan dengan  $Q_1, Q_2, \text{ dan } Q_3$ .

Untuk menentukan nilai kuartil caranya :

- Susun data menurut urutan nilainya
- Tentukan letak kuartil

$$\text{Letak } Q_j = \text{data ke } \frac{j(n+1)}{4} \text{ dengan } j=1,2,3$$

- Tentukan nilai kuartil

Untuk data berdaftar distribusi frekuensi nilai Kuartil :

$$Q_j = L_1 + d \left( \frac{\frac{j \cdot N}{4} - \sum F}{f_{Q_j}} \right) \text{ dengan } j=1,2,3$$

Contoh :

1. Carilah  $Q_1, Q_2, \text{ dan } Q_3$  dari data 75, 82, 66, 57, 64, 56, 92, 94, 86, 52, 60, 70

!

Data diurutkan dulu menjadi :

✓  $Q_1$

Letak  $Q_1 =$

Nilai  $Q_1$

✓  $Q_2$

Letak  $Q_2 =$

Nilai  $Q_2$

✓  $Q_3$

Letak  $Q_3 =$

Nilai  $Q_3$

2. Carilah  $Q_1, Q_2$ , dan  $Q_3$  dari data

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
Jumlah	80

$$\frac{N}{4} =$$

Kelas  $Q_1 =$

$$L_1 =$$

$$d =$$

$$\sum F =$$

$$f_{Q1} =$$

$$Q_1 =$$

PENYELESAIAN :

$$\rightarrow 2 \frac{N}{4} = 40$$

$$\rightarrow \text{Kelas } Q_2 = 75 - 79$$

$$\rightarrow L_1 = 74,5$$

$$\rightarrow d = 5$$

$$\rightarrow \sum F = 36$$

$$\rightarrow f_{Q2} = 21$$

$$\rightarrow Q_2 = 74,5 + 5 \left( \frac{2 \cdot \frac{80}{4} - 36}{21} \right) = 74,5 + 0,95 = 75,45$$

$$3 \frac{N}{4} = 60$$

$$\text{Kelas } Q_3 = 80-84$$

$$L_1 = 79,5$$

$$d = 5$$

$$\sum F = 57$$

$$f = 6$$

## DESIL

Jika kumpulan data itu dibagi menjadi 10 bagian yang sama, maka di dapat sembilan pembagi dan tiap pembagi dinamakan DESIL. Dinotasikan  $D_1, D_2, D_3, \dots, D_9$ .

Letak  $D_j = \text{data ke } \frac{j(n+1)}{10}$  dengan  $j=1,2,3,4,\dots,9$

Untuk data dalam distribusi frekuensi nilai Desil :

$$D_j = L_1 + d \left( \frac{\left( \frac{j(n+1)}{10} - \sum F \right)}{f} \right) \quad \text{dengan } j=1,2,3,\dots,9$$

### ➤ PERSENTIL

Jika kumpulan data itu dibagi menjadi 100 bagian yang sama, maka di dapat sembilan puluh sembilan pembagi dan tiap pembagi dinamakan PERSENTIL. Dinotasikan  $P_1, P_2, P_3, \dots, P_{99}$ .

Letak  $P_j = \text{data ke } \frac{j(n+1)}{100}$  dengan  $j=1,2,3,4,\dots,99$

Untuk data dalam distribusi frekuensi nilai Desil :

$$P_j = L_1 + d \left( \frac{\left( \frac{j(n+1)}{100} - \sum F \right)}{f} \right) \quad \text{dengan } j=1,2,3,\dots,99$$

## 3.7 PENGGUNAAN EXCEL

### Fungsi Statistika

FUNGSI	SINTAKSIS	KETERANGAN
Rerata aritmatika	AVERAGE	Rata-rata (aritmatika) data
Rerata deviasi mutlak	AVEDEV	Rata-rata harga mutlak deviasi,
Jumlah	SUM	Jumlah data

Maksimum,	MAX	Data terbesar
Minimum	MIN	Data terkecil
Median	MEDIAN	Median data
Modus	MODE	Modus data
Kuartil	QUARTILE	Kuartil ke d data dimana d=0 menghasilkan data terkecil, d=1 kuartil pertama, d=2 kuartil kedua, d=3 kuartil ketiga dan d=4 menghasilkan data terbesar.
Persentil	PERCENTILE	Persentil ke d dimana d= 0 s.d. 1. Contoh: d=0.6 menghasilkan data ke 60%.
Standar deviasi	STDEV(data)	Standar deviasi untuk sample
	STDEVP(data)	Standar tu deviasi untuk populasi
Rerata geometri	GEOMEAN(x1, x2, . . . xn)	Rerata geometri

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
Jumlah	80

$$P_{25}$$

$$\frac{25 \cdot 80}{100} = 20$$

$$\text{Kelas } P_{25} = 65 - 69$$

$$L = 64,5$$

$$d = 5$$

$$\sum F = 14$$

$$f = 10$$

$$P_{25} = 64,5 + 5 \left( \frac{\frac{25 \cdot 80}{100} - 14}{10} \right) = 67,5$$





### SOAL-SOAL LATIHAN

1. Write out the terms in each of the following indicated sums:

a)  $\sum_{j=1}^6 X_j$       b).  $\sum_{j=1}^4 (X_j - a)$       c).  $\sum_{k=1}^5 f_k \cdot X_k$       d).  $\sum_{j=1}^4 (j^2)$

2. The grades of a student on six examinations were 84, 91, 72, 68, 87, and 78. Find the arithmetic mean of the grades!
3. Find the arithmetic mean of the numbers 5, 3, 6, 5, 4, 5, 2, 8, 6, 5, 4, 8, 3, 4, 5, 4, 8, 2, 5, 4!
4. Out of 100 numbers, 20 were 4's, 40 were 5's, 30 were 6's and the remainder were 7's. Find the arithmetic mean of the numbers!
5. 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

Hitunglah : Mean, Median, Modus !

**LEMBAR KERJA :**

1).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

2).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

3).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

4).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

5).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

## BAB IV

### UKURAN PENYEBARAN, DISPERSI DAN VARIANSI

Kecuali ukuran gejala pusat masih ada lagi ukuran lain yaitu ukuran penyebaran atau ukuran dispersi. Ukuran ini kadang-kadang dinamakan pula ukuran variansi, yang menggambarkan bagaimana berpencahnya data kuantitatif. Beberapa ukuran dispersi yang terkenal antara lain range, deviasi mean, range semi-interkuartil, range persentil 10-90, standar deviasi, variansi.

#### 4.1 RANGE (Rentang)

**Range = data terbesar – data terkecil**

EXAMPLE 1.

Range dari data 2, 3, 3, 5, 5, 5, 8, 10, 12 is  $12 - 2 = 10$ .

Terkadang range ditulis secara sederhana yaitu 2 to 12, or 2–12.

#### 4.2 MEAN DEVIATION /AVERAGE DEVIATION/Deviasi Mean (Rata-rata Simpangan)

Mean Deviation (MD) dari data tunggal yaitu  $X_1, X_2, X_3, \dots,$

$X_N$  didefinisikan :

$$MD = \frac{\sum_{j=1}^N |X_j - \bar{X}|}{N} = \frac{\sum |X - \bar{X}|}{N} = |\overline{X - \bar{X}}|$$

Dengan : MD = Mean Deviation

$X_j$  = Data ke-j, dengan  $j = 1, 2, 3, \dots$

$\bar{X}$  = Mean Aritmatika

$|X - \bar{X}|$  = Jarak antara tiap data dengan mean/rata-rata

EXAMPLE 2 : Hitunglah MD dari data 2, 3, 6, 8, 11!

$$\bar{X} = \frac{2 + 3 + 6 + 8 + 11}{5} = 6$$

$$MD = \frac{(2-6) + |3-6| + |6-6| + |8-6| + |11-6|}{5} = \frac{4+3+0+2+5}{5} = 2,8$$

Jika data tunggal berfrekuensi yaitu  $X_1, X_2, X_3, \dots, X_N$  dengan frekuensi  $f_1, f_2, f_3, \dots, f_N$  maka

$$MD = \frac{\sum_{j=1}^N f_j |X_j - \bar{X}|}{N} = \sum f \frac{|X - \bar{X}|}{N}$$

### 4.3 RANGE SEMI-INTERKUARTIL/DEVIASI KUARTIL (Rentang Semi antar Kuartil)

RAK (Rentang Antar Kuartil) =  $Q_3 - Q_1$

Range Semi Interkuartil dari sekumpulan data adalah  $Q = \frac{Q_3 - Q_1}{2}$

### 4.4 RANGE PERSENTIL 10-90

Range Persentil 10-90 dari sekumpulan data adalah  $P_{90} - P_{10}$

Range Semi Persentil 10-90 dari sekumpulan data adalah  $\frac{P_{90} - P_{10}}{2}$

### 4.5 STANDAR DEVIASI (Simpangan Baku)

Standar Deviasi dari data tunggal  $X_1, X_2, X_3, \dots, X_N$  yang berasal dari populasi didefinisikan

$$\sigma = \sqrt{\frac{\sum_{j=1}^N (X_j - \bar{X})^2}{N}}$$

Standar Deviasi dari data tunggal  $X_1, X_2, X_3, \dots, X_N$  yang berasal dari sampel didefinisikan

$$s = \sqrt{\frac{\sum_{j=1}^N (X_j - \bar{X})^2}{N-1}}$$

Contoh : Diberikan sample dengan data 6, 7, 8, 9, 10

Hitunglah simpangan bakunya !

$$\bar{x} = 8$$

$X_j$	$X_j - \bar{X}$	$(X_j - \bar{X})^2$
6	-2	4
7	-1	1
8	0	0
9	1	1
10	2	4

Jumlah	10
$s = \sqrt{\frac{10}{5-1}} = \sqrt{2.5}$	

$$s = \sqrt{\frac{N \sum X^2 - (\sum X)^2}{N(N-1)}} = \sqrt{\frac{5.330 - 40^2}{5.4}} = \sqrt{\frac{1650 - 1600}{20}} = \sqrt{2.5}$$

$X_j$	$X_j^2$
6	36
7	49
8	64
9	81
10	100
40	330

Standar Deviasi dari data berdaftar distribusi frekuensi yang berasal dari sampel didefinisikan :

$$s = \sqrt{\frac{\sum_{j=1}^N f_j (A_j - \bar{X})^2}{N-1}}$$

- Dengan :
- s = Standar Deviasi
  - $f_j$  = Frekuensi kelas ke-j
  - $A_j$  = Tanda kelas ke-j
  - $\bar{X}$  = Rata-rata
  - N = Banyaknya data

Untuk membantu menghitung biasanya digunakan table tambahan sebagai berikut :

Rentang nilai	$f_j$	$A_j$	$A_j \cdot f_j$	$A_j - \bar{X}$	$(A_j - \bar{X})^2$	$f_j (A_j - \bar{X})^2$
Jumlah	...	-	...	-	-	...

Contoh :Tabel 2.4 Distribusi nilai matematika 80 siswa SMA XYZ

Rentang nilai	frekuensi	$A_j$	$A_j \cdot f_j$	$A_j - \bar{X}$	$(A_j - \bar{X})^2$	$f_j (A_j - \bar{X})^2$
50-54	1	52	52	-23,375	546,39	546.39
55-59	2	57	114	-18,375	337,64	675.28

60-64	11	62	682	-13,375	178,89	1967.79
65-69	10	67	670	-8,375	70,14	701.4
70-74	12	72	864	-3,375	11,39	136.68
75-79	21	77	1617	1,625	2,64	55.45
80-84	6	82	492	6,625	43,89	263.34
85-89	9	87	783	11,625	135,14	1216.26
90-94	4	92	368	16,625	276,39	1105.56
95-99	4	97	388	21.625	467,64	1870.56
Jumlah	80	-	6030	-	-	8538.71

$$\bar{X} = 75,375$$

$$s = \sqrt{\frac{8538.71}{80 - 1}} = \sqrt{108.08} = 10.39$$

$$s = \sqrt{\frac{N \sum f_j \cdot A_j^2 - (\sum f_j \cdot A_j)^2}{N(N-1)}}$$

Untuk membantu menghitung biasanya digunakan table tambahan sebagai berikut :

Rentang nilai	$f_j$	$A_j$	$A_j^2$	$f_j A_j$	$f_j A_j^2$
Jumlah	...	-	-	...	...

Contoh :Tabel 2.4 Distribusi nilai matematika 80 siswa SMA XYZ

Rentang nilai	frekuensi	$A_j$	$A_j^2$	$f_j A_j$	$f_j A_j^2$
50-54	1	52	2704	52	2704
55-59	2	57	3249	114	6498
60-64	11	62	3844	682	42284
65-69	10	67	4489	670	44890
70-74	12	72	5184	864	62208
75-79	21	77	5929	1617	124509
80-84	6	82	6724	492	40344
85-89	9	87	7569	783	68121
90-94	4	92	8464	368	33856
95-99	4	97	9409	388	37636
Jumlah	80	-	-	6030	463050



$$s = \sqrt{\frac{N \sum f_j \cdot A_j^2 - (\sum f_j \cdot A_j)^2}{N(N-1)}} = \sqrt{\frac{80 \cdot 463050 - (6030)^2}{80 \times 79}} = \sqrt{108,08} = 10.39$$

$$\bullet \quad s = \sqrt{d^2 \frac{N \sum f_j \cdot c_j^2 - (\sum f_j \cdot c_j)^2}{N(N-1)}}$$

- Dengan :
- s = Standar Deviasi
  - d = lebar kelas interval
  - c = sandi ,  $c = 0, \pm 1, \pm 2, \dots$
  - $f_j$  = Frekuensi kelas ke-j
  - N = Banyaknya data

Untuk membantu menghitung biasanya digunakan table tambahan sebagai berikut :

Rentang nilai	$f_j$	$c_j$	$c_j^2$	$f_j c_j$	$f_j c_j^2$
Jumlah	...	-	-	...	...

Contoh : Tabel 2.4 Distribusi nilai matematika 80 siswa SMA XYZ

Rentang nilai	frekuensi	$c_j$	$c_j^2$	$f_j c_j$	$f_j c_j^2$
50-54	1	-5	25	-5	25
55-59	2	-4	16	-8	32
60-64	11	-3	9	-33	99
65-69	10	-2	4	-20	40
70-74	12	-1	1	-12	12
75-79	21	0	0	0	0
80-84	6	1	1	6	6
85-89	9	2	4	18	36
90-94	4	3	9	12	36
95-99	4	4	16	16	64
Jumlah	80	-	-	- 26	350

$$s = \sqrt{d^2 \frac{N \sum f_j \cdot c_j^2 - (\sum f_j \cdot c_j)^2}{N(N-1)}} = \sqrt{5^2 \frac{80 \times 350 - (-26)^2}{80 \times 79}} = \sqrt{108,08} = 10.39$$

#### 4.6 Varians

Varians dari suatu data adalah kuadrat dari standar deviasi.

## 4.7 PENGGUNAAN EXCEL

### Fungsi Statistika

FUNGSI	SINTAKSIS	KETERANGAN
Mean Deviation	AVEDEV	Rata-rata simpangan
Standar Deviation (Sampel)	STDEV	Simpangan baku dari data yang berasal dari sampel
Varians	VAR	Varians dari sebuah data
Standar Deviation (Populasi)	STDEVP	Simpangan baku dari data yang berasal dari populasi

TUGAS DIKUMPULKAN DALAM BENTUK PRINT OUT !

Dibawah ini data mengenai kematian per 1000 penduduk yang terdapat di beberapa kota di jawa sebagai sampel :

13.6	17.7	10.8	21.5	11.3	16.4	14.1	21.2
18.6	15.9	12.8	12.7	16.5	13.4	19.3	7.3
17.1	9.5	23.3	21.5	10.6	14.0	14.1	9.3
17.5	13.5	11.3	10.6	20.5	12.2	7.4	19.7
9.0	24.6	17.8	17.3	15.5	13.6	10.9	10.4
15.3	10.9	10.7	10.9	11.1	15.9	13.2	
19.9	14.2	14.1	14.7	13.5	17.7	14.1	
9.8	8.8	9.1	12.9	13.7	17.3	19.4	
14.8	15.9	16.1	12.6	15.1	11.3	14.6	
10.7	9.0	13.0	19.8	9.9	13.2	18.7	

Dengan menggunakan excel Hitunglah :

1. Data Minimal
2. Data Maximal
3. Mean Aritmatika, Mean Geometrik, Mean Harmonik
4. Median
5. Modus
6. Q1,Q2,Q3
7. D3, D7, D9
8. P35, P75, P99
9. Mean Deviation
10. Standar Deviasi
11. Varians





## BAB V.

### PENDUGAAN PARAMETER

- ☺ Untuk mempelajari populasi, tidak mungkin kita mengamati seluruh anggota populasi, tetapi dengan menggunakan contoh (sampel)
- ☺ Populasi → Parameter →  $\mu$  dan  $\sigma^2$
- ☺ Sampel → Statistik →  $\bar{X}$  dan  $S^2$

#### PENDUGA TITIK

- ☺ Misalkan kita mempunyai sebuah peubah acak  $X$  dengan nilai sampel pengamatan  $X_1, X_2, X_3, \dots, X_n$ , maka nilai tengahnya:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

dinamakan penduga titik (point estimate) dari nilai tengah populasi  $\mu$ .

- ☺ Sedangkan ragam contohnya:

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

dinamakan penduga titik dari ragam populasi  $\sigma^2$ .

#### PENDUGA SELANG

- ☺ Dalam berbagai keadaan, nilai duga titik belum memberikan informasi yang cukup tentang parameter populasi, karena nilai tengahnya tergantung pada contoh yang diambil.
- ☺ Misalkan saja untuk menduga hasil rata-rata per hektar hasil padi pada suatu musim tanam, tidak cukup kita mengatakan 3 ton/ha gabah kering karena bila contoh yang diambil berbeda, rata-ratanya belum tentu sama dengan 3 ton/ha, mungkin lebih atau kurang dari itu.
- ☺ Untuk menentukan penduga selang dari sesuatu parameter yang tidak diketahui (dilambangkan  $\theta$ , bisa untuk  $\mu$  ataupun  $\sigma^2$ ), sehingga:

$$P(B_1 < \theta < B_2) = 1 - \alpha$$

di sebut selang kepercayaan  $100(1 - \alpha)\%$  untuk parameter  $\theta$ .

- ☺  $B_1$  dan  $B_2$  masing-masing disebut batas kepercayaan bawah dan atas.  $B_1$  dan  $B_2$  didapatkan dari hasil perhitungan
- ☺  $1 - \alpha$  adalah taraf kepercayaan kebenaran selang tersebut, atau dengan kata lain  $\alpha$  disebut tingkat kesalahan

### **Pendugaan Selang Untuk Nilai Tengah $\mu$**

a. Untuk Ragam Populasi Diketahui

$$P(\bar{X} - Z_{\alpha/2} \sigma / \sqrt{n} < \mu < \bar{X} + Z_{\alpha/2} \sigma / \sqrt{n}) = 1 - \alpha$$

di mana  $B_1 = \bar{X} - Z_{\alpha/2} \sigma / \sqrt{n}$  dan  $B_2 = \bar{X} + Z_{\alpha/2} \sigma / \sqrt{n}$

$Z_{\alpha/2}$  adalah nilai dari Tabel normal baku

$n$  adalah besarnya ukuran sampel

Contoh:

Untuk menentukan rata-rata pendapatan per kapita dilakukan survei terhadap 150 keluarga yang ditentukan secara acak. Dari hasil survei tersebut diperoleh rata-rata pendapatan

Rp 60.000 per kapita per bulan. Dari hasil sensus, diperoleh bahwa simpangan baku adalah Rp 12.500, tentukan selang kepercayaan 0,90 untuk pendapatan tersebut.

Jawab:

Dari sini diperoleh:  $n = 150$ ,  $\bar{X} = 60.000$  dan  $\sigma = 12.500$

Dengan  $1 - \alpha = 0,90$ , berarti  $\alpha = 0,10$  dan  $\alpha/2 = 0,05$ .

Dari tabel normal baku, taraf 0,05 memiliki besar  $Z = 1,64$

$$P(60.000 - 1,64 \times 12.500 / \sqrt{150} < \mu < 60.000 + 1,64 \times 12.500 / \sqrt{150}) = 0,9$$

$$P(58.326 < \mu < 61.674) = 0,90$$

Jadi selang kepercayaan 0,90 untuk rata-rata pendapatan adalah  $58.326 < \mu < 61.674$

b. Untuk Ragam Populasi Tidak Diketahui

$$P(\bar{X} - t_{\alpha/2(n-1)} s / \sqrt{n} < \mu < \bar{X} + t_{\alpha/2(n-1)} s / \sqrt{n}) = 1 - \alpha$$

$t_{\alpha/2(n-1)}$  adalah nilai dari Tabel t dengan derajat bebas (db) =  $n - 1$

Contoh:

Suatu percobaan dilakukan untuk mempelajari pengaruh pemberian obat perangsang (procaine) terhadap laju jantung

13 ekor kucing. Tiap kucing diberi 10 mg procaine. Setelah beberapa saat, tekanan jantungnya diukur, sebagai berikut:

170; 126; 105; 135; 186; 198; 140; 160; 138; 120; 150; 168; 123

Tentukan selang kepercayaan 0,95 untuk nilai tengah laju jantung tersebut.

Jawab:

Dari data di atas,  $n = 13$ , dihitung dulu  $\bar{X}$  dan  $S$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = 147,6$$

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}} = 27,46$$

Dari tabel t diketahui,  $\alpha/2 = 0,025$  (karena  $1 - \alpha = 0,95$ ), dengan derajat bebas =  $n - 1 = 13 - 1 = 12$

$$t_{(\alpha/2)(12)} = 2,179$$

$$P(147,6 - 2,179 \times 27,46 / \sqrt{13} < \mu < 147,6 + 2,179 \times 27,46 / \sqrt{13}) = 0,95$$

$$P(131,0 < \mu < 164,2) = 0,95$$

### **Pendugaan Selang Untuk Ragam $\sigma^2$**

☺ Selang kepercayaan  $(1 - \alpha)$  bagi ragam populasi  $\sigma^2$  adalah:

$$P\left(\frac{(n-1)S^2}{\chi^2_{k(\alpha/2)}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{k(1-\alpha/2)}}\right) = 1 - \alpha$$

$\chi^2_{k(\alpha/2)}$  adalah nilai dari Tabel khi kuadrat dengan db =  $k = n - 1$  dengan taraf  $\alpha/2$ , dan

$\chi^2_{k(1-\alpha/2)}$  adalah nilai dari Tabel khi kuadrat dengan db =  $k = n - 1$  dengan taraf  $(1 - \alpha/2)$

Contoh:

Tentukan selang kepercayaan 0,95 untuk ragam laju jantung kucing pada soal di atas!

$$n = 13; \bar{X} = 147,6; S = 27,46$$

Dari tabel khi kuadrat

$$\chi^2_{(0,025)(12)} = 23,3 \text{ dan } \chi^2_{(0,975)(12)} = 4,40$$

$$P\left(\frac{(13-1)27,46^2}{23,3} < \sigma^2 < \frac{(13-1)27,46^2}{4,4}\right) = 0,95$$

Selang kepercayaan untuk ragam

$$P(388,35 < \sigma^2 < 2056,50) = 0,95$$

Selang kepercayaan untuk simpangan baku

$$P(19,71 < \sigma < 45,35) = 0,95$$



## BAB VI.

### PENGUJIAN HIPOTESIS

- ☺ Hipotesis adalah suatu proses dari pendugaan parameter dalam populasi, yang membawa kita pada perumusan segugus kaidah yang dapat membawa kita pada suatu keputusan akhir, yaitu menolak atau menerima pernyataan tersebut.

Contoh:

1. Seorang peneliti masalah kedokteran diminta untuk memutuskan, berdasarkan bukti-bukti hasil percobaan, apakah suatu vaksin baru lebih baik daripada yang sekarang beredar di pasaran.
  2. Berdasarkan data, apakah ada perbedaan ketelitian antara dua jenis alat ukur;
  3. Seorang ahli sosiologi ingin mengumpulkan data yang memungkinkan ia menyimpulkan apakah jenis darah dan warna seseorang ada hubungannya atau tidak.
- ☺ Hipotesis Statistika: suatu proses untuk menentukan apakah dugaan tentang nilai parameter/karakteristik populasi didukung kuat oleh data sampel atau tidak
- ☺ Alur dalam pengujian hipotesis:
- DATA (KUANTITATIF) → HIPOTESIS → PENGUJIAN → DECISION RULE  
→ KEPUTUSAN → KESIMPULAN
- ☺ Dalam statistika, dikenal 2 macam hipotesis:
1. Hipotesis nol ( $H_0$ ), berupa suatu pernyataan tidak adanya perbedaan karakteristik/parameter populasi (selalui ditandai dengan tanda =)
  2. Hipotesis alternatif ( $H_1$ ), berupa suatu pernyataan yang bertentangan dengan  $H_0$ .
- ☺ Ingat, yang diuji dalam hipotesis adalah parameter, maka notasi yang digunakan dalam hipotesis statistika adalah parameter  $\mu$  (untuk nilai tengah),  $\sigma$  (untuk simpangan baku), dan  $p$  (untuk proporsi).
- ☺ Contoh: Suatu obat baru lebih baik dari obat yang selama ini digunakan jika persentase orang yang sembuh setelah meminum obat baru ini lebih dari 60%.  
Dalam permasalahan ini, maka dapat dibentuk hip statistik:
- $H_0 : p = 0,6$  (obat baru tidak lebih baik)
- $H_1 : p > 0,6$  (obat baru lebih baik)

☺ Terdapat 2 tipe hipotesis:

1. Hipotesis satu arah (atau hipotesis satu sisi)

Jika hipotesis alternatif menunjukkan tanda  $>$  atau  $<$ . Hal ini dikarenakan si peneliti atau si perancang hipotesis, menginginkan suatu perubahan satu arah, misalnya apakah meningkat, apakah terjadi penurunan, dan sebagainya.

Contoh: sebuah perusahaan rokok menyatakan bahwa kadar nikotin rata-rata rokok yang diproduksinya *tidak melebihi 2,5* miligram (tidak melebihi berarti kurang dari, berarti satu arah saja,  $H_1 : \mu < 2,5$ ).

2. Hipotesis dua arah (atau hipotesis dua sisi)

Jika hipotesis alternatif menunjukkan tanda  $\neq$ .

Misalkan  $H_0 : \mu = 20$ , lawan  $H_1 : \mu \neq 20$

Ini berarti hipotesis alternatifnya memiliki dua definisi,  $H_1 : \mu > 20$  dan/atau  $H_1 : \mu < 20$ . Hal ini dikarenakan si peneliti menginginkan suatu perbedaan, yaitu apakah berbeda atau tidak (entah berbeda itu meningkat, atau menurun).

Contoh: sebuah pabrik sereal ingin mengetes unjuk kerja dari mesin pengisinya. Mesin tersebut dirancang untuk mengisi 12 ons setiap boksnya. (karena hanya ingin menguji apakah rata-rata mesin pengisi tersebut dapat mengisi 12 ons setiap boksnya atau tidak,  $H_0 : \mu = 12$ , dan  $H_1 : \mu \neq 12$ )

☺ Langkah pengujian hipotesis:

1. Tentukan hipotesis

Misal:  $H_0 : \mu = c$ , lawan  $H_1 : \mu \neq c$  (uji dua sisi)

Atau:  $H_0 : \mu = c$ , lawan  $H_1 : \mu > c$  (uji satu sisi)

2. Tentukan tingkat signifikansi  $\alpha$

Biasanya kalau tidak diketahui, maka hal yang biasa digunakan adalah tingkat kesalahan  $\alpha$  sebesar 5%.

3. Statistik Uji

4. Daerah kritik,  $H_0$  diterima bila dan  $H_0$  ditolak bila.

5. Keputusan,  $H_0$  diterima atau ditolak

6. Kesimpulan

Latihan: tentukan hipotesis nol dan alternatifnya!

1. Rata-rata curah salju di Danau Toba selama bulan Februari 21,8 cm.
2. Banyaknya staf dosen di suatu PT yang menyumbang dalam suatu acara pengumpulan dana sosial tidak lebih dari 20%.
3. Secara rata-rata anak-anak di St. Louis, berangkat dari rumah ke sekolah menempuh jarak tidak lebih dari 6,2 km.
4. Di tahun mendatang, sekurang-kurangnya 70% dari mobil baru termasuk dalam kategori kompak dan subkompak.
5. Dalam pemilu mendatang, proporsi yang memilih calon lama adalah 0,58.
6. Di Restoran X, rata-rata steak yang dihidangkan sekurang-kurangnya 340 gram.

## PENGUJIAN HIPOTESIS SATU POPULASI

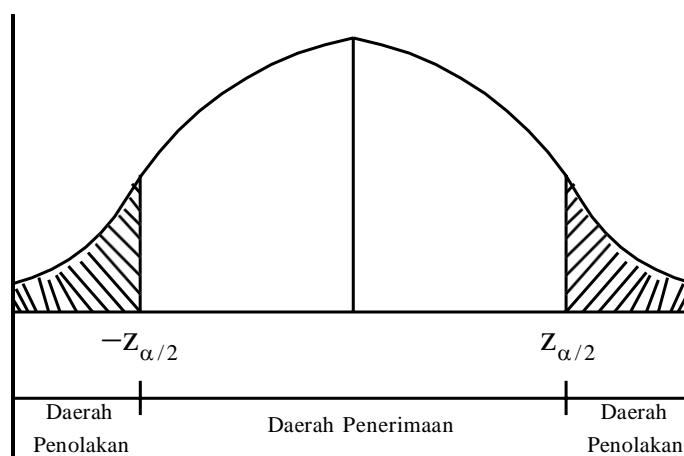
### PENGUJIAN $\mu$ UNTUK RAGAM DIKETAHUI

☺ Statistik Uji: Z

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

☺ Untuk hipotesis dua sisi:

$H_0: \mu = c$  lawan  $H_1: \mu \neq c$



Daerah Penerimaan  $H_0$

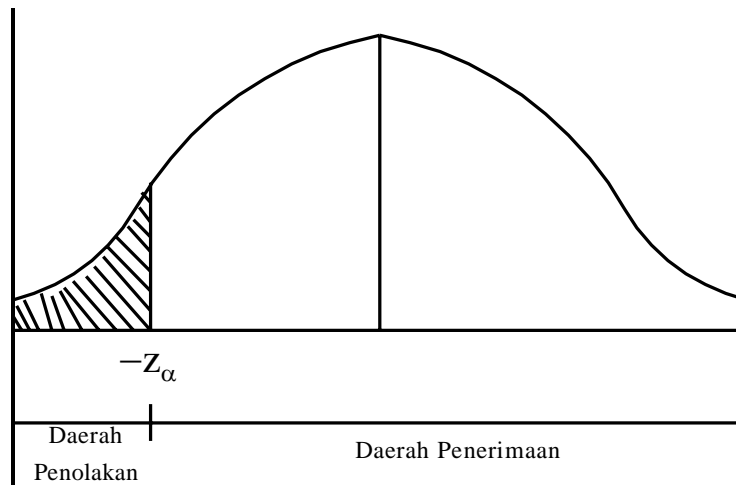
$$-Z_{\alpha/2} < Z < Z_{\alpha/2}$$

Daerah Penolakan  $H_0$

$$Z > Z_{\alpha/2} \text{ atau } Z < -Z_{\alpha/2}$$

☺ Untuk hipotesis satu sisi:

$H_0: \mu = c$  lawan  $H_1: \mu < c$



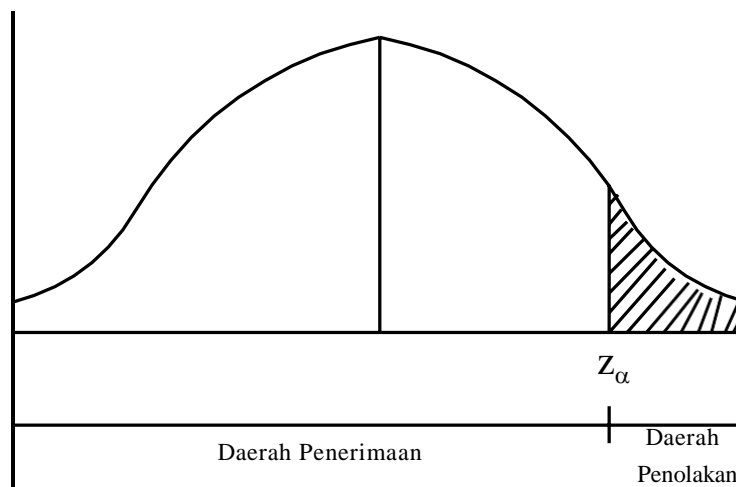
Daerah Penerimaan  $H_0$

$$Z > -Z_\alpha$$

Daerah Penolakan  $H_0$

$$Z < -Z_\alpha$$

$H_0: \mu = c$  lawan  $H_1: \mu > c$



Daerah Penerimaan  $H_0$

$$Z < Z_\alpha$$

Daerah Penolakan  $H_0$

$$Z > Z_\alpha$$

### PENGUJIAN $\mu$ UNTUK RAGAM TIDAK DIKETAHUI

☺ Statistik Uji:  $t$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

Dibandingkan dengan  $t_{\alpha/2}$  (dua sisi) &  $t_\alpha$  (satu sisi) dg db=n-1

Metode daerah penerimaan maupun penolakan  $H_0$  sama dengan di atas.

☺ Contoh:

Sebuah perusahaan alat olahraga mengembangkan jenis batang pancing sintetis, ingin menguji apakah alat pancing tersebut memiliki kekuatan dengan nilai tengah 8 kg. Diketahui bahwa simpangan baku adalah 0,5 kg. Ujilah hipotesis tersebut, bila suatu contoh acak 50 batang pancing itu setelah di tes memberikan nilai tengah 7,8 kg. Gunakan taraf nyata 0,01.

Jawab:

1. Hipotesis

$H_0 : \mu = 8$ , lawan  $H_1 : \mu \neq 8$  (uji dua sisi)

2. Tingkat signifikansi  $\alpha = 0,01$

$$Z_{\alpha/2} = Z_{0,005} = 2,575$$

3. Statistik Uji

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{7,8 - 8}{0,5 / \sqrt{50}} = -2,83$$

4. Daerah kritik

$$H_0 \text{ diterima} : -Z_{\alpha/2} < Z < Z_{\alpha/2} \rightarrow -2,575 < Z < 2,575$$

$$H_0 \text{ ditolak} : Z > Z_{\alpha/2} \text{ atau } Z < -Z_{\alpha/2} \rightarrow Z > 2,575 \text{ atau } Z < -2,575$$

5. Keputusan

Karena  $Z < -Z_{\alpha/2}$  ( $-2,83 < -2,575$ ), maka  $H_0$  ditolak

6. Kesimpulan

Bahwa rata-rata kekuatan batang pancing tidak sama dengan 8 kg, tetapi kurang dari 8 kg.

☺ Contoh:

Seorang peneliti ingin melakukan suatu penelitian mengenai tinggi badan mahasiswa yang mengikuti mata kuliah Statistika. Untuk itu dilakukan suatu penelitian terhadap sepuluh mahasiswa yang mengikuti mata kuliah tsb.

Mhs ke-	1	2	3	4	5	6	7	8	9	10
TB (cm)	185	150	156	171	160	160	165	171	166	150

Ujilah hipotesis:

- Apakah tinggi badan mahasiswa tersebut adalah 155 cm?
- Apakah tinggi badan mahasiswa tersebut di atas 155 cm?
- Apakah tinggi badan mahasiswa tersebut di bawah 155 cm?

*Penyelesaian:*

- $H_0 : \mu = 155$  vs  $H_1 : \mu \neq 155$

$$\bar{X} = 163,40$$

$$s = 10,69$$

$$t = \frac{\bar{X} - \mu}{s / \sqrt{n}} = \frac{163,40 - 155}{10,69 / \sqrt{10}} = 2,48$$

$$t_{0,025(9)} = 2,262$$

$$t > t_{0,025(9)}$$

Keputusan: tolak  $H_0$ , terima  $H_1$

b.  $H_0 : \mu = 155$

$$H_1 : \mu > 155$$

$$t_{0,05(9)} = 1,833$$

$$t > t_{0,05(9)}$$

Keputusan: tolak  $H_0$ , terima  $H_1$

c.  $H_0 : \mu = 155$

$$H_1 : \mu < 155$$

$$-t_{0,05(9)} = -1,833$$

$$t > -t_{0,05(9)}$$

Keputusan: terima  $H_0$

## PENGUJIAN UNTUK PROPORSI

☺ Hipotesisnya:

$$H_0 : p = c \text{ lawan } H_1 : p > c \text{ (satu sisi)}$$

☺ Statistik Uji: Z

$$Z = \frac{X - nc}{\sqrt{nc(1-c)}}$$

Metode daerah penerimaan maupun penolakan  $H_0$  sama dengan pengujian hipotesis nilai tengah untuk ragam diketahui.

☺ Contoh:

Seorang pemborong menyatakan bahwa di 70% rumah-rumah yang baru dibangun di kota X dipasang suatu alat pemompa udara panas. Ingin diuji pernyataan tersebut di atas, dengan dilakukan suatu penelitian, diperoleh 15 rumah baru yang diambil secara acak, terdapat 8 rumah yang menggunakan pompa udara panas. Gunakan taraf nyata 0,10.

Penyelesaian:

1. Hipotesis

$H_0 : p = 0,7$  dan  $H_1 : p \neq 0,7$  (dua sisi)

2. Tingkat signifikansi  $\alpha = 0,01$

$$Z_{\alpha/2} = Z_{0,005} = 2,575$$

3. Statistik Uji

$$Z = \frac{x - nc}{\sqrt{nc(1-c)}} = \frac{8 - 15 \times 0,7}{\sqrt{15 \times 0,7 \times 0,3}} = -1,41$$

4. Daerah kritik

$H_0$  diterima :  $-Z_{\alpha/2} < Z < Z_{\alpha/2} \rightarrow -2,575 < Z < 2,575$

$H_0$  ditolak :  $Z > Z_{\alpha/2}$  atau  $Z < -Z_{\alpha/2} \rightarrow Z > 2,575$  atau  $Z < -2,575$

5. Keputusan

Karena  $-Z_{\alpha/2} < Z < Z_{\alpha/2}$  ( $-2,575 < -1,41 < 2,575$ ),  $H_0$  diterima

6. Kesimpulan

Bahwa tidak ada alasan yang kuat untuk meragukan pernyataan pemborong di atas.

## PENGUJIAN UNTUK RAGAM

☺ Hipotesisnya:

$H_0 : \sigma^2 = c$  lawan  $H_1 : \sigma^2 \neq c$  (dua sisi)

☺ Statistik Uji:  $\chi^2$

$$\chi^2 = \frac{(n-1)s^2}{c}$$

Dibandingkan dengan  $\chi^2_{\alpha/2}$  (dua sisi) &  $\chi^2_{\alpha}$  (satu sisi) dg db=n-1

☺ Untuk hipotesis dua sisi  $H_0 : \sigma^2 = c$  lawan  $H_1 : \sigma^2 \neq c$

Daerah penolakan  $H_0 \rightarrow \chi^2 < \chi^2_{1-\alpha/2}$  dan  $\chi^2 > \chi^2_{\alpha/2}$

☺ Untuk hipotesis satu sisi  $H_0 : \sigma^2 = c$  lawan  $H_1 : \sigma^2 < c$

Daerah penolakan  $H_0 \rightarrow \chi^2 < \chi^2_{1-\alpha}$

☺ Untuk hipotesis satu sisi  $H_0 : \sigma^2 = c$  lawan  $H_1 : \sigma^2 > c$

Daerah penolakan  $H_0 \rightarrow \chi^2 > \chi^2_{\alpha}$

☺ Contoh:

Sebuah perusahaan aki mobil mengatakan bahwa umur aki yang diproduksinya mempunyai simpangan baku 0,9 tahun. Bila suatu contoh acak 10 aki

menghasilkan simpangan baku  $s = 1,2$  tahun, apakah menurut anda simpangan baku tersebut lebih besar dari 0,9 tahun? Gunakan taraf nyata 0,05.

Penyelesaian:

1. Hipotesis

$$H_0 : \sigma^2 = 0,9^2 = 0,81 \text{ dan } H_1 : \sigma^2 > 0,81 \text{ (satu sisi)}$$

2. Tingkat signifikansi  $\alpha = 0,05$

$$\chi^2_{\alpha} = \chi^2_{0,05} = 16,919$$

3. Statistik Uji

$$\chi^2 = \frac{(n-1)s^2}{c} = \frac{9 \times 1,44}{0,81} = 16,0$$

4. Daerah kritik

$$H_0 \text{ ditolak : } \chi^2 > \chi^2_{\alpha} \rightarrow \chi^2 > 16,919$$

5. Keputusan

Karena  $\chi^2 < \chi^2_{\alpha}$  ( $16,0 < 16,919$ ),  $H_0$  diterima

6. Kesimpulan

Bahwa tidak ada alasan untuk meragukan bahwa simpangan bakunya adalah 0,9 tahun.



## SOAL-SOAL

1. Tinggi rata-rata mhs tingkat awal di suatu PT adalah 162,5 cm dengan simpangan baku 6,9 cm. Apakah ada alasan untuk mempercayai bahwa telah terjadi perubahan dalam tinggi rata-rata, bila suatu contoh acak 50 mhs tingkat awal mempunyai tinggi rata-rata 165,2 cm? Gunakan taraf nyata 0,02.
2. Ujilah bahwa isi kaleng rata-rata suatu jenis minyak pelumas adalah 10 liter bila isi suatu contoh acak 10 kaleng adalah 10,2; 9,7; 10,1; 10,3; 10,1; 9,8; 9,9; 10,4; 10,3; dan 9,8 liter. Gunakan taraf nyata 0,01.
3. Tahun lalu karyawan dinas kebersihan kota menyumbang rata-rata \$8 pada korban bencana alam. Ujilah hipotesis bahwa sumbangan rata-rata tahun ini akan meningkat bila suatu contoh acak 12 karyawan menunjukkan sumbangan rata-rata \$8,9 dengan simpangan baku \$1,75.
4. Pengalaman lalu menunjukkan bahwa waktu yang diperlukan oleh siswa kelas 3 SMA untuk menyelesaikan suatu ujian memiliki simpangan baku 6 menit. Ujilah hipotesis bahwa simpangan baku tersebut saat ini menjadi lebih kecil, jika suatu contoh acak 20 siswa menghasilkan simpangan baku 4,51
5. Suatu obat penenang ketegangan saraf diduga hanya 60% efektif. Hasil percobaan dengan obat baru terhadap 100 orang dewasa penderita ketegangan syaraf, yang diambil secara acak, menunjukkan bahwa obat baru itu 70% efektif. Apakah ini merupakan bukti yang cukup untuk menyimpulkan bahwa obat baru itu lebih baik daripada yang beredar sekarang?

**LEMBAR KERJA :**

1).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

2).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

3).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

4).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

5).....

.....

.....

.....

.....

.....

.....

.....

.....

.....

## BAB VII.

### PENGUJIAN HIPOTESIS DUA POPULASI

- ☺ Misalkan kita tertarik untuk membandingkan efisiensi 2 mesin, mesin A dan mesin B, mana yang lebih baik,
- ☺ Atau kita tertarik untuk membandingkan potensi tanaman pada varietas A dan varietas B, apakah terdapat perbedaan hasil panen varietas A dan B, maka hipotesis yang akan di uji adalah:  
 $H_0 : \mu_A = \mu_B$  (tidak terdapat perbedaan pada kedua varietas tersebut)  
 $H_1 : \mu_A \neq \mu_B$  (terdapat perbedaan pada kedua varietas tersebut)
- ☺ Atau kita ingin menguji apakah varietas A lebih baik daripada varietas B? maka hipotesisnya:  
 $H_0 : \mu_A = \mu_B$  versus  $H_1 : \mu_A < \mu_B$

#### PENGUJIAN DUA $\mu$ UNTUK RAGAM POP DIKETAHUI

- ☺ Statistik Uji yang digunakan:

$$Z = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{(\sigma_A^2 / n_A + \sigma_B^2 / n_B)}}$$

- ☺ Decision rule (kaidah keputusannya) sama dengan sebelumnya
- ☺ Contoh:

Dari suatu survei di dua daerah yang masing-masing dengan contoh berukuran 30 dan 36 berturut-turut diperoleh nilai tengah pendapatan per kapita per bulan Rp 45.000 di daerah A dan Rp 47.500 untuk daerah B. Jika diketahui bahwa ragam pendapatannya sebesar  $(Rp.6.000)^2$  dan  $(Rp.7.500)^2$  berturut-turut, dengan taraf kepercayaan 95%, tentukan apakah pendapatan rata-rata di A berbeda dengan di B atau tidak!

Jawab:

1. Hipotesis

$H_0 : \mu_A = \mu_B$  versus  $H_1 : \mu_A \neq \mu_B$  (uji dua sisi)

2. Tingkat signifikansi  $\alpha = 0,05$

$Z_{\alpha/2} = Z_{0,025} = 1,96$

3. Statistik Uji

$$Z = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{(\sigma^2_A / n_A + \sigma^2_B / n_B)}} = \frac{45000 - 47500}{\sqrt{(6000^2 / 30 + 7500^2 / 36)}} = -1,504$$

4. Daerah kritik

$H_0$  diterima :  $-Z_{\alpha/2} < Z < Z_{\alpha/2} \rightarrow -1,96 < Z < 1,96$

$H_0$  ditolak :  $Z > Z_{\alpha/2}$  atau  $Z < -Z_{\alpha/2} \rightarrow Z > 1,96$  atau  $Z < -1,96$

5. Keputusan

Karena  $-Z_{\alpha/2} < Z < Z_{\alpha/2}$  ( $-1,96 < Z < 1,96$ ), maka  $H_0$  diterima

6. Kesimpulan

Bahwa pendapatan perkapita dua daerah tersebut adalah sama.

### **PENGUJIAN DUA $\mu$ UNTUK RAGAM POP TDK DIKETAHUI**

- ☺ Sama seperti uji satu populasi, jika ragam tidak diketahui, statistik uji yang digunakan adalah statistik t.
- ☺ Bila ragam populasi untuk kedua populasi tersebut tidak diketahui, kita harus menyelidiki contoh A (dari populasi A) dan contoh B (dari populasi B) apakah populasi tersebut berpasangan atau tidak.
  1. Jika contoh A, yang diambil bebas terhadap contoh B. Artinya, kita mengambil secara acak contoh A berukuran  $n_A$  dan kita juga mengambil contoh B secara acak berukuran  $n_B$ . Jenis pengujian ini dinamakan uji t tidak berpasangan.
  2. Jika pada setiap pengukuran contoh A dan B diambil secara berpasangan. Dengan demikian, ukuran untuk contoh A dan B adalah sama, yaitu katakanlah n. Jenis pengujian ini dinamakan uji t berpasangan.

## A. UJI t TIDAK BERPASANGAN

☺ Terdapat permasalahan dalam uji t tidak berpasangan, yaitu apakah dua populasi tersebut berasal dari ragam yang sama atau tidak?

☺ Untuk itu kita harus mengujinya apakah  $\sigma^2_A$  sama dengan  $\sigma^2_B$  atau tidak. Hipotesis untuk menguji hal itu adalah sebagai berikut:

$H_0 : \sigma^2_A = \sigma^2_B$  (artinya kedua populasi berasal dari ragam yang sama)

$H_1 : \sigma^2_A \neq \sigma^2_B$  (artinya kedua populasi berasal dari ragam yang sama)

Statistik uji yang digunakan adalah statistik F.

$$F = \frac{s^2_1}{s^2_2}$$

di mana  $s^2_1$  adalah ragam terbesar dari dua populasi tersebut (apakah  $s^2_A$  atau  $s^2_B$ ) dan  $s^2_2$  adalah ragam terkecil di antara keduanya.

F tersebut dibandingkan dengan  $F_\alpha$  dengan  $db1 = n_1 - 1$  dan  $db2 = n_2 - 1$ . Jika  $F < F_\alpha$  maka  $H_0$  diterima, artinya ragam populasi sama, sedangkan bila  $F > F_\alpha$  maka  $H_0$  ditolak, artinya ragam populasi berbeda.

### 1. Untuk ragam populasi sama

Karena kedua ragam sama, maka ragamnya dapat di gabung:

$$s^2 = \frac{(n_A - 1)s^2_A + (n_B - 1)s^2_B}{(n_A - 1) + (n_B - 1)}$$

statistik uji nya:

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{s^2 \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

di bandingkan dengan  $t_\alpha$  (untuk satu sisi) dan  $t_{\alpha/2}$  (untuk dua sisi) dengan  $db = n_A + n_B - 2$

### 2. Untuk ragam populasi tidak sama

Karena kedua ragam tidak sama, maka kita tidak dapat menggabungkan kedua ragam populasi tersebut.

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\left( \frac{s^2_A}{n_A} + \frac{s^2_B}{n_B} \right)}}$$

di bandingkan dengan  $t_{\alpha}$  (untuk satu sisi) dan  $t_{\alpha/2}$  (untuk dua sisi) dengan

$$db = \frac{(s_A^2 / n_A + s_B^2 / n_B)^2}{[(s_A^2 / n_A)^2 / (n_A - 1)] + [(s_B^2 / n_B)^2 / (n_B - 1)]}$$

Contoh:

Kemampuan mahasiswa dari jalur PSB dan SPMB akan diperbandingkan dalam hal kemampuan mereka terhadap mata kuliah statistika. Pada masing-masing kelompok diambil secara acak 14 mahasiswa dari PSB (dinamakan kelompok A) dan 18 mahasiswa dari SPMB (dinamakan kelompok B).

Dari data yang diperoleh, setelah dilakukan perhitungan, ternyata bahwa  $\bar{X}_A = 68,5$ ;  $\bar{X}_B = 66,0$ ;  $s_A^2 = 110,65$  dan  $s_B^2 = 188,59$ . Dengan tingkat kesalahan 5%, ingin ditentukan apakah kemampuan kedua kelompok tersebut sama atau tidak.

Jawab:

Hipotesis yang akan di uji:

$H_0 : \mu_A = \mu_B$  versus  $H_1 : \mu_A \neq \mu_B$

Untuk menentukan apakah ragam kedua populasi itu sama atau tidak dilakukan uji F

$$F = \frac{s_B^2}{s_A^2} = \frac{188,59}{110,65} = 1,70$$

Dengan  $F_{0,05}$  dengan  $db_1 = 18 - 1 = 17$  dan  $db_2 = 14 - 1 = 13$  sebesar 2,357. Karena  $F <$

$F_{0,05}$  maka ragam kedua populasi adalah sama. Maka ragam gabungannya:

$$s^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{(n_A - 1) + (n_B - 1)} = \frac{(14 - 1)110,65 + (18 - 1)188,59}{(14 - 1) + (18 - 1)} = 154,82$$

Statistik uji t yang digunakan:

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{s^2 \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}} = \frac{68,5 - 66,0}{\sqrt{154,82 \left( \frac{1}{14} + \frac{1}{18} \right)}} = 0,56$$

Dengan  $t_{0,025}$  dan  $db = n_A + n_B - 2 = 14 + 18 - 2 = 30$  adalah sebesar 2,045.

Karena t terletak di antara  $-t_{0,025} < t < t_{0,025}$  maka  $H_0$  diterima, artinya tidak terdapat



perbedaan kemampuan statistika antara mahasiswa asal PSB dengan SPMB.

Contoh:

Suatu penelitian terhadap suatu populasi mengambil 2 contoh masing-masing berukuran 15 dan 10. Berdasarkan hasil pengukuran diperoleh  $\bar{X}_A = 2$ ,  $\bar{X}_B = 1$ ,  $s^2_A = 10$  dan  $s^2_B = 35$ . Tentukan apakah kedua contoh di atas berasal dari populasi dengan nilai tengah sama atau tidak!

Jawab

Hipotesis yang akan di uji:

$$H_0 : \mu_A = \mu_B \text{ versus } H_1 : \mu_A \neq \mu_B$$

Untuk menentukan apakah ragam kedua populasi itu sama atau tidak dilakukan uji F

$$F = \frac{s^2_B}{s^2_A} = \frac{35}{10} = 3,5$$

Dengan  $F_{0,05}$  dengan  $db_1=10-1 = 9$  dan  $db_2=15-1=14$  sebesar 2,65. Karena  $F > F_{0,05}$  maka ragam kedua populasi adalah tidak sama.

Statistik uji t yang digunakan:

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\left(\frac{s^2_A}{n_A} + \frac{s^2_B}{n_B}\right)}} = \frac{2 - 1}{\sqrt{\left(\frac{10}{15} + \frac{35}{10}\right)}} = 0,46$$

$$db = \frac{(s_A^2 / n_A + s_B^2 / n_B)^2}{[(s_A^2 / n_A)^2 / (n_A - 1)] + [(s_B^2 / n_B)^2 / (n_B - 1)]} = 10,73 \approx 11$$

Dengan  $t_{0,025(11)} = 2,201$

Karena t terletak di antara  $-t_{0,025} < t < t_{0,025}$  maka  $H_0$  diterima, artinya nilai tengah kedua populasi sama.

**B. UJI t BERPASANGAN**

- ☺ Dua sampel yang diamati secara berpasangan, artinya dalam setiap pengukuran yang diukur adalah pasangan [A,B].
- ☺ Karena pengamatannya secara berpasangan maka dalam setiap pengamatan  $X_A$  dan  $X_B$  tidak lagi bebas sesamanya meski bebas antara pasangan yang satu dengan pasangan yang lain.

☺ Sebagai contoh,  $X_A$  dan  $X_B$  masing-masing kadar auksin ruas pertama dan kedua dari pucuk burung tanaman teh

Atau  $X_A$  dan  $X_B$  berturut-turut kadar vitamin C bagian ujung dan pangkal dari sebuah buah mangga.

Atau lebih ekstrim lagi, yaitu detak jantung seseorang pada saat biasa, dan pada saat dekat dengan belahan jiwa.

☺ Metode uji t berpasangan ini adalah sama dengan pengujian hipotesis satu populasi, yaitu data selisih dari kedua populasi tersebut.

$$D_j = X_{Aj} - X_{Bj}$$

$$t = \frac{\bar{D}}{\sqrt{s^2_D / n}}$$

Dibandingkan dengan  $t_{\alpha/2}$  (dua sisi) dan  $t_\alpha$  (satu sisi) dengan derajat bebas =  $n - 1$

☺ Contoh:

Suatu penelitian ditujukan untuk mempelajari apakah ada perbedaan antara banyaknya biji per bunga dari bunga bagian atas dan bagian bawah 10 tanaman bakau.

	1	2	3	4	5	6	7	8	9	10
Atas	1,4	3,3	2,0	0,4	2,1	1,9	1,1	0,1	0,9	3,0
Bawah	1,1	1,7	1,8	0,3	0,8	1,4	1,0	0,4	0,7	0,9

Pengamatan di atas jelas pengamatan berpasangan, dan kita memandang baik bagian atas ( $X_A$ ) maupun bagian bawah ( $X_B$ ) pada setiap pasangan tidak bebas sesamanya. Yang harus dicari adalah selisih antara bagian atas dan bawah:

	1	2	3	4	5	6	7	8	9	10
Atas	1,4	3,3	2,0	0,4	2,1	1,9	1,1	0,1	0,9	3,0
Bawah	1,1	1,7	1,8	0,3	0,8	1,4	1,0	0,4	0,7	0,9
D	0,3	1,6	0,2	0,1	1,3	0,5	0,1	-0,3	0,2	2,1

Hipotesis yang akan diuji:

$$H_0 : \mu_A = \mu_B \text{ versus } H_1 : \mu_A \neq \mu_B$$

$$\text{Atau } H_0 : \mu_D = 0 \text{ versus } H_1 : \mu_D \neq 0$$

Di mana kita peroleh  $\bar{D} = 0,61$  dan  $s^2 = 0,6077$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{0,61 - 0}{\sqrt{0,6077 / 10}} = 2,474$$

Dengan db =  $n - 1 = 10 - 1 = 9$ , diperoleh  $t_{0,025} = 2,262$ . Karena  $t > t_{0,025}$  maka  $H_0$  ditolak, artinya terdapat perbedaan antara banyak biji yang

dihasilkan oleh bunga bagian atas tanaman dan bunga bagian bawah tanaman.

Karena  $\bar{D} > 0$ , di mana D adalah selisih bagian atas dengan bagian bawah, maka bagian atas memiliki jumlah biji yang lebih banyak.

### PENGUJIAN DUA PROPORSI

- ☺ Misalkan pada dua populasi, populasi A berukuran N dengan karakteristik x sebanyak  $N_x$ , dan populasi B yang berukuran M dengan karakteristik x sebanyak  $M_x$ . Dari contoh berukuran n dan m yang diambil secara acak dari populasi pertama dan kedua berturut-turut ternyata dengan karakteristik x sebanyak  $n_x$  dan  $m_x$ .
- ☺ Penduga proporsi untuk kedua populasi tersebut adalah:

$$\hat{p}_A = \frac{n_x}{n} \text{ dan } \hat{p}_B = \frac{m_x}{m}$$

- ☺ Hipotesis yang akan di uji:  $H_0 : p_A = p_B$  dan  $H_1 : p_A \neq p_B$
- ☺ Statistik uji Z:

$$Z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\left( \frac{\hat{p}_A (1 - \hat{p}_A)}{n} + \frac{\hat{p}_B (1 - \hat{p}_B)}{m} \right)}}$$

- ☺ Contoh:

Suatu penelitian dilakukan untuk mempelajari pengaruh merokok pada saat seorang ibu mengandung terhadap kondisi anak setelah lahir. Untuk itu diambil contoh acak 200 dan 250 orang ibu yang pada saat mengandung anaknya adalah perokok dan bukan perokok berturut-turut. Setelah dilakukan pengujian ternyata banyak anak lahir cacat adalah 90 dan 60 orang berturut-turut. Dengan tingkat kesalahan 5%, tentukan apakah ada pengaruh merokok saat mengandung pada kondisi fisik anak atau tidak!

Jawab:

Jika ada pengaruh merokok, maka proporsi bayi tersebut cacat antara kelompok ibu perokok dan tidak perokok adalah berbeda. Maka hipotesis yang akan di uji adalah:  
 $H_0 : p_A = p_B$  (artinya proporsi bayi tersebut cacat untuk kedua kelompok adalah sama)

$H_1 : p_A \neq p_B$  (artinya proporsi bayi tersebut cacat untuk kedua kelompok adalah berbeda)

Kita samakan persepsi, kelompok A adalah kelompok ibu perokok, dan B adalah kelompok ibu tidak perokok.

$$n = 200 \quad m = 250$$

$$n_x = 90 \quad m_x = 60$$

maka

$$\hat{p}_A = \frac{n_x}{n} = 90/200 = 0,45$$

$$\hat{p}_B = \frac{m_x}{m} = 60/250 = 0,24$$

$$Z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\left(\frac{\hat{p}_A(1-\hat{p}_A)}{n} + \frac{\hat{p}_B(1-\hat{p}_B)}{m}\right)}} = \frac{0,45 - 0,24}{\sqrt{\left(\frac{0,45 * 0,55}{200} + \frac{0,24 * 0,76}{250}\right)}} = 4,82$$

Berdasarkan tabel normal baku,  $Z_{0,025} = 1,96$ . Karena  $Z > Z_{0,025}$  maka  $H_0$  ditolak, artinya merokok pada saat mengandung berpengaruh pada kondisi fisik bayi yang dilahirkan.

#### Latihan

- Dua jenis plastik A dan B dapat digunakan untuk komponen elektronik. Tegangan luluh (*breaking strength*) dari kedua plastik tersebut sangat penting dalam menentukan kualitasnya. Diketahui bahwa simpangan baku tegangan luluh plastik A dan B adalah sama yaitu sebesar 10 psi. Untuk menguji jenis plastik tersebut, diambil contoh acak berukuran 10 untuk jenis plastik A, dan 12 untuk jenis plastik B, didapatkan nilai tengah berturut-turut 162,5 psi dan 155,0 psi. Ujilah apakah kedua jenis plastik di atas berkekuatan/berkualitas sama atau tidak!
- Suatu contoh berukuran 20 keluarga diambil secara acak dari kota A dan 25 keluarga dari kota B. Dari hasil pengamatan diperoleh hasil:
  - Rata-rata pengeluaran di kota A adalah Rp 148.000 per bulan dengan simpangan baku Rp 13.200.
  - Rata-rata pengeluaran di kota B adalah Rp 133.760 per bulan dengan simpangan baku Rp 11.100.
 Ujilah apakah rata-rata pengeluaran di kota A paling tidak sedikit lebih tinggi daripada rata-rata pengeluaran di kota B.
- Dua cara fermentasi pucuk teh diperbandingkan untuk ditentukan cara mana yang memberikan persentase teh hancur (*broken tea*) yang paling sedikit. Untuk masing-

masing cara diambil contoh acak sebanyak 10 dan 16 kali berturut-turut. Dari data yang dikumpulkan (persentase teh hancur) diperoleh nilai tengah dan ragam sebesar 25% dan 35% untuk fermentasi I, sedangkan untuk fermentasi II dengan nilai tengah dan ragam 20% dan 25%. Ujilah pernyataan tersebut.

4. Data berikut ini adalah hasil pengukuran skala I/E (*internal external locus of control scale*) dari dua kelompok orang, yaitu kelompok A yang terdiri dari 20 orang perokok yang ingin menghentikan kebiasaan merokoknya dan kelompok B yang terdiri dari 20 orang perokok yang tidak ingin menghentikan kebiasaan rokoknya.

No	1	2	3	4	5	6	7	8	9	10
A	8	5	10	9	8	8	9	10	6	11
B	13	18	11	17	8	10	10	16	14	15
No	11	12	13	14	15	16	17	18	19	20
A	14	15	18	8	9	16	9	9	5	15
B	15	10	11	10	10	13	10	13	16	10

Ujilah apakah ada perbedaan skala I/E di antara kedua kelompok tersebut.

5. Suatu kelompok terdiri dari 10 orang diberi suatu zat perangsang. Hasil pengukuran terhadap tekanan darah pada saat sebelum (A) dan sesudah (B) perlakuan sebagaimana dalam tabel berikut ini:

No	1	2	3	4	5	6	7	8	9	10
A	118	120	128	124	136	130	130	140	140	128
B	127	128	136	131	138	132	131	141	132	120

Apa kesimpulan saudara?

6. Data berikut ini adalah tegangan permukaan (dalam dyness) dari cairan rumen yang diambil pada 8 hari yang berbeda untuk sebelum dan sesudah diberi makan.

Sebelum	52,9	49,1	50,9	51,2	49,2	48,5	51,7	53,8
Sesudah	56,7	51,4	54,7	50,9	56,7	55,8	54,4	53,5

Ujilah bahwa tidak ada perbedaan tegangan permukaan pada saat sebelum dan sesudah diberi makan!

7. Empat ratus klon suatu jenis rumput dipelajari ketahanannya terhadap penyakit karat di dua tempat. Pada tempat pertama (A) ternyata 372 klon terserang karat. Sedangkan di tempat kedua (B) terdapat 230 klon yang terserang. Apakah terdapat perbedaan di antara kedua tempat tersebut!
8. Untuk mempelajari perilaku laki-laki dan perempuan dalam suatu pemilihan, masing-masing kelompok diambil contoh acak berukuran 500. Dari contoh acak ternyata 420 laki-laki dan 360 perempuan yang ikut berpartisipasi dalam pemilihan. Ujilah apakah ada perbedaan perilaku antara laki-laki dan perempuan!
9. Dua kelompok anak-anak penderita asma dipergunakan untuk mempelajari perilaku anak yang menderita penyakit tersebut. Satu kelompok (A) terdiri dari 160 anak yang diperlakukan di sebuah rumah sakit, dan satu kelompok lagi (B) terdiri dari 100 anak di suatu tempat yang terisolir. Setelah perlakuan masing-masing kelompok didiagnosa untuk ditentukan siapa yang berperilaku anti sosial (sisopatik) dan tidak. Ternyata 25 anak dan 30 anak dari kelompok A dan B berturut-turut adalah sosiopatik. Apakah terdapat perbedaan sosiopatik di kedua kelompok tersebut.

**LEMBAR KERJA :**

1).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

2).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

3).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

4).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....



**LEMBAR KERJA :**

5).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

6).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

7).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

8).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

9).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

## BAB VIII.

### ANALISIS RAGAM

- ☺ Pada dasarnya analisis ragam adalah pengujian hipotesis lebih dari dua populasi. Tetapi analisis ragam ini bisa saja digunakan untuk uji hipotesis dua populasi, tetapi hasilnya identik (persis sama) dengan uji t dua populasi.
- ☺ Misalkan terdapat 5 populasi. Atau dalam istilah penelitian, terdapat 5 populasi dikarenakan terdapat 5 perlakuan (*treatment*) yang berbeda terhadap kelima kelompok tersebut. Misalkan tiap perlakuan tersebut berukuran sampel sebanyak 3 buah (dinamakan ulangan), maka dapat ditabulasikan sebagai berikut:

Perlakuan	A	B	C	D	E
	$Y_{11}$	$Y_{21}$	$Y_{31}$	$Y_{41}$	$Y_{51}$
	$Y_{12}$	$Y_{22}$	$Y_{32}$	$Y_{42}$	$Y_{52}$
	$Y_{13}$	$Y_{23}$	$Y_{33}$	$Y_{43}$	$Y_{53}$
Jumlah	$Y_{1.}$	$Y_{2.}$	$Y_{3.}$	$Y_{4.}$	$Y_{5.}$

di mana  $Y_{ij}$  adalah data hasil pengamatan perlakuan ke  $i$  dan ulangan ke  $j$ .

Hipotesis yang ingin diuji:

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$$

$H_1$  : Sekurang-kurangnya dua nilai tengah tidak sama

di mana simbol yang digunakan:

$n$  = banyaknya pengamatan tiap perlakuan (ulangan)

$p$  = banyaknya perlakuan

- ☺ Statistik uji yang digunakan adalah statistik F
- ☺ Dilakukan pembentukan Tabel Analisis Ragam, atau biasa dikenal Tabel ANOVA (*analysis of variance*), yang berisi:

a. Sumber Keragaman

Perlakuan, yaitu keragaman yang disebabkan atas perbedaan perlakuan/kondisi

Galat, yaitu keragaman yang tidak dapat dikontrol (error)

Total

b. Derajat Bebas

$$db_{\text{Perlakuan}} = p - 1$$

$$db_{\text{Total}} = np - 1$$

$$db_{\text{Galat}} = db_{\text{Total}} - db_{\text{Perlakuan}}$$

c. Jumlah Kuadrat

$$JK_{\text{Total}} = \sum_{i=1}^p \sum_{j=1}^n Y_{ij}^2 - \left( \sum_{i=1}^p \sum_{j=1}^n Y_{ij} \right)^2 / np$$

$$JK_{\text{Perlakuan}} = \left( \sum_{i=1}^p Y_{i.} \right)^2 / n - \left( \sum_{i=1}^p \sum_{j=1}^n Y_{ij} \right)^2 / np$$

$$JK_{\text{Galat}} = JK_{\text{Total}} - JK_{\text{Perlakuan}}$$

d. Kuadrat Tengah

$$KT_{\text{Perlakuan}} = JK_{\text{Perlakuan}} / db_{\text{Perlakuan}}$$

$$KT_{\text{Galat}} = JK_{\text{Galat}} / db_{\text{Galat}}$$

e.  $F_{\text{Hitung}} = KT_{\text{Perlakuan}} / KT_{\text{Galat}}$

f.  $F_{\text{Tabel}} = F_{\alpha}(db_{\text{perlakuan}}; db_{\text{galat}})$

g. Kaidah keputusan

$F_{\text{Hitung}} \geq F_{\text{Tabel}} \rightarrow$  tolak  $H_0$  (antar perlakuan berbeda nyata/signifikan)

$F_{\text{Hitung}} < F_{\text{Tabel}} \rightarrow$  terima  $H_0$  (antar perlakuan tidak berbeda nyata)

### Tabel Analisis Ragam (ANOVA)

SK	db	JK	KT	$F_{\text{Hitung}}$	$F_{\text{Tabel}}$
Perlakuan	p-1	.....	.....	.....	.....
Galat	.....	.....	.....		
Total	np-1	.....			

Untuk Ulangan Tidak Sama

Perlakuan	A	B	C	D	E
	$Y_{11}$	$Y_{21}$	$Y_{31}$	$Y_{41}$	$Y_{51}$
	$Y_{12}$	$Y_{22}$	$Y_{32}$	$Y_{42}$	$Y_{52}$
	$Y_{13}$	$Y_{23}$	$Y_{33}$	$Y_{43}$	$Y_{53}$
	$Y_{14}$		$Y_{34}$	$Y_{44}$	$Y_{54}$
	$Y_{15}$				$Y_{55}$
					$Y_{56}$
Jumlah	$Y_{1.}$	$Y_{2.}$	$Y_{3.}$	$Y_{4.}$	$Y_{5.}$

$n_i$  = banyaknya ulangan pada perlakuan ke-i

$$N = \text{total observasi} = \sum_{i=1}^p n_i$$

$$db_{\text{Perlakuan}} = p - 1$$

$$db_{\text{Total}} = N - 1$$

$$JK_{\text{Perlakuan}} = \sum_{i=1}^p \frac{Y_{i.}^2}{n_i} - \left( \sum_{i=1}^p \sum_{j=1}^n Y_{ij} \right)^2 / N$$

$$JK_{\text{Total}} = \sum_{i=1}^p \sum_{j=1}^{n_i} Y_{ij}^2 - \left( \sum_{i=1}^p \sum_{j=1}^{n_i} Y_{ij} \right)^2 / N$$

Contoh:

Seorang peneliti ingin menguji manakah di antara kelima pabrikan sepeda motor yang paling irit. Untuk menguji hal tersebut, dia mendapatkan data jumlah kilometer dalam satu liter yang dapat ditempuh motor-motor tersebut pada tabel di bawah ini:

Pabrikan	Yamaha	Honda	Suzuki	Kawazaki	Mochin
	42	41	55	37	29
	41	47	53	43	37
	49	42	52	39	29
	42	55	51	35	31
	50	51	54	45	27
Jumlah	224	236	265	199	153
Rerata	44,8	47,2	53,0	39,8	30,6

Ujilah apakah terdapat perbedaan konsumsi BBM untuk tiap pabrikan tersebut!

Jawab:

Hipotesis yang akan diuji:

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$$

$H_1$  : Sekurang-kurangnya dua nilai tengah tidak sama

$$n = 5 \text{ dan } p = 5$$

a. Derajat Bebas

$$db_{\text{Perlakuan}} = p - 1 = 5 - 1 = 4$$

$$db_{\text{Total}} = np - 1 = 5 \times 5 - 1 = 24$$

$$db_{\text{Galat}} = db_{\text{Total}} - db_{\text{Perlakuan}} = 24 - 4 = 20$$

b. Jumlah Kuadrat

$$\begin{aligned} JK_{\text{Total}} &= \sum_{i=1}^p \sum_{j=1}^n Y_{ij}^2 - \left( \sum_{i=1}^p \sum_{j=1}^n Y_{ij} \right)^2 / np \\ &= (42^2 + 41^2 + \dots + 27^2) - (224 + \dots + 153)^2 / (5 \times 5) \\ &= 48175,00 - 46397,16 = 1777,84 \end{aligned}$$

$$\begin{aligned} JK_{\text{Perlakuan}} &= \left( \sum_{i=1}^p Y_i^2 \right) / n - \left( \sum_{i=1}^p \sum_{j=1}^n Y_{ij} \right)^2 / np \\ &= (224^2 + 236^2 + \dots + 153^2) - (224 + \dots + 153)^2 / (5 \times 5) \\ &= 47821,40 - 46397,16 = 1424,24 \end{aligned}$$

$$JK_{\text{Galat}} = JK_{\text{Total}} - JK_{\text{Perlakuan}}$$

$$= 1777,84 - 1424,24 = 353,60$$

c. Kuadrat Tengah

$$KT_{\text{Perlakuan}} = JK_{\text{Perlakuan}}/db_{\text{Perlakuan}} = 1424,24/4 = 356,06$$

$$KT_{\text{Galat}} = KT_{\text{Galat}}/db_{\text{Galat}} = 353,60/20 = 17,68$$

d.  $F_{\text{Hitung}} = KT_{\text{Perlakuan}}/KT_{\text{Galat}} = 356,06/17,68 = 20,14$

e.  $F_{\text{Tabel}} = F_{\alpha}(db_{\text{perlakuan}}; db_{\text{galat}}) = F_{0,05}(4;20) = 2,87$

Tabel Analisis Ragam (ANOVA)

SK	db	JK	KT	$F_{\text{Hitung}}$	$F_{\text{Tabel}}$
Perlakuan	4	1424,24	356,06	20,14	2,87
Galat	20	353,60	17,68		
Total	24	1777,84			

Karena  $F_{\text{hitung}} > F_{\text{tabel}}$  maka  $H_0$  ditolak,  $H_1$  diterima, artinya terdapat perbedaan konsumsi BBM antara kelima pabrikan tersebut.

Mana yang berbeda antara satu dengan lainnya, maka dilanjutkan dengan uji pembandingan berganda.

## UJI PEMBANDINGAN BERGANDA

- ☺ Setelah pengujian analisis ragam, bila didapatkan bahwa terdapat perlakuan yang berbeda nyata (signifikan) atau tolak  $H_0$ , maka perlu diuji lebih lanjut, mana yang perlakuan berbeda. Ada beberapa uji yang digunakan: BNT, BNJ, Duncan.
- ☺ Patut dan harus diingat, uji lanjut ini digunakan bila  $H_0$  dari analisis ragam ditolak. Jika  $H_0$  diterima, uji lanjut tidak perlu dilakukan.
- ☺ Uji lanjut yang biasa digunakan dan akan di bahas di sini adalah uji Beda Nyata Terkecil (BNT).
- ☺ Pada dasarnya uji lanjut adalah melakukan penotasian rata-rata perlakuan didasarkan pada Beda Nyata Terkecil (BNT)
- ☺ BNT adalah suatu nilai terkecil yang harus dipenuhi agar 2 rata-rata perlakuan dapat dikatakan sama atau berbeda.
- ☺ Prosedur yang dilakukan:
  1. Urutkan nilai rata-rata dari terbesar ke terkecil atau sebaliknya
  2. Carilah selisih antara 2 rata-rata tersebut (dari beberapa rata-rata yang ada) dengan:
    - Jika selisih  $\leq$  BNT, maka kedua rata-rata tersebut dikatakan sama sehingga dapat diberi notasi yang sama pula
    - Jika selisih  $>$  BNT, maka diberi notasi yang berbeda

$$BNT = t_{\alpha/2}(dbgalat) \sqrt{KT_{galat} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Contoh:

Dari soal sebelumnya, karena  $H_0$  ditolak, lakukan uji BNT, manakan pabrikan yang menghasilkan motor paling irit!

Jawab:

Dari hasil perhitungan sebelumnya diperoleh nilai rata-rata sebagai berikut:

Pabrikan	Yamaha	Honda	Suzuki	Kawazaki	Mochin
Rerata	44,8	47,2	53,0	39,8	30,6

Besarnya BNT adalah sebagai berikut:

$$\begin{aligned}
 BNT &= t_{\alpha/2}(dbgalat) \sqrt{KT_{galat} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \\
 &= 2,086 \sqrt{17,68 \left( \frac{1}{5} + \frac{1}{5} \right)} = 5,55
 \end{aligned}$$

Hasil Uji BNT:

Pabrikan	Rerata	Notasi
Mochin	30,6	a
Kawazaki	39,8	b
Yamaha	44,8	bc
Honda	47,2	c
Suzuki	53,0	d

Karena notasi untuk Suzuki adalah berbeda dengan lainnya, dan rerata paling tertinggi, maka pabrikan Suzuki menghasilkan motor yang paling irit, karena rata-rata pabrikan motor tersebut menghabiskan 53 km per liter.

Sedangkan pabrikan mochin menghasilkan motor yang paling boros, karena rata-rata pabrikan motor tersebut menghasilkan 30,6 km per liter, nah, coba kalian bayangkan, dengan kenaikan harga BBM sekarang ini, nggak etis kan kalo kita pake motor china, meskipun dari segi harga cukup murah, tapi biaya operasionalnya ternyata paling tinggi.

Yamaha dan Honda, menghasilkan motor yang paling irit setelah Suzuki. Mereka berdua menghabiskan banyaknya kilometer per liter yang tidak berbeda.

Pertanyaan selanjutnya, anda mau beli motor yang mana?



1. Data berikut mencantumkan 5 merek rokok yang terjual di sebuah pasar swalayan pada 8 hari yang dipilih secara acak:

Merek								
A	21	35	32	28	14	47	25	38
B	35	12	27	41	19	23	31	20
C	45	60	33	36	31	40	43	48
D	32	53	29	42	40	23	35	42
E	45	29	31	22	36	29	42	30

Lakukan analisis ragam, pada taraf nyata 0,05 dan tentukan apakah secara rata-rata di pasar swalayan ini kelima rokok di atas terjual sama banyak? Jika tidak terjual sama banyak, merek rokok apa yang paling laku di pasaran?

2. Enam mesin yang berbeda sedang dipertimbangkan untuk digunakan dalam pembuatan tutup yang terbuat dari karet. Yang menjadi bahan pertimbangan adalah kekuatan regangan produk yang dihasilkannya. Suatu contoh acak 4 tutup dari setiap mesin diambil untuk digunakan menentukan apakah kekuatan regangan tutup karet yang dihasilkan itu bervariasi antar tiap mesin. Data yang diperoleh adalah sebagai berikut, dalam satuan kilogram per  $\text{cm}^2 \times 10^{-1}$

Mesin					
1	2	3	4	5	6
17.5	16.4	20.3	14.6	17.5	18.3
16.9	19.2	15.7	16.7	19.2	16.2
15.8	17.7	17.8	20.8	16.5	17.5
18.6	15.4	18.9	18.9	20.5	20.1

Lakukan analisis ragam pada taraf nyata 0,05 dan simpulkan apakah nilai tengah perlakuannya berbeda nyata atau tidak.

3. Tiga kelas kuliah Statistika Dasar diberikan oleh tiga dosen. Nilai akhirnya tercatat sebagai berikut:

Dosen A: 73 89 82 43 80 73 66 60 45 93 36 77

Dosen B: 88 78 48 91 51 85 74 77 31 78 62 76 96 80 56

Dosen C: 68 79 56 91 71 71 87 41 59 68 53 79 15

Apakah ada selisih yang nyata di antara nilai rata-rata yang diberikan oleh ketiga dosen tersebut? Manakah dosen yang paling baik menurut versi mahasiswa?

4. Dalam sebuah percobaan Biologi, 4 konsentrasi bahan kimia digunakan untuk merangsang pertumbuhan sejenis tanaman tertentu selama periode waktu tertentu. Data pertumbuhan berikut, dalam cm, dicatat dari tanaman yang hidup:

Konsentrasi			
1	2	3	4
8.2	7.7	6.9	6.8
8.7	8.4	5.8	7.3
9.4	8.6	7.2	6.3
9.2	8.1	6.8	6.9
	8.0	7.4	7.1
		6.1	

Apakah ada beda pertumbuhan rata-rata yang nyata yang disebabkan oleh keempat konsentrasi bahan kimia tersebut?

**LEMBAR KERJA :**

1).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

2).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

3).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

4).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

## BAB IX.

### ANALISIS REGRESI DAN KORELASI

- ☺ Tujuan utama **analisis regresi** adalah mencari ada tidaknya hubungan linier antara dua variabel:
  - Variabel bebas (X), yaitu variabel yang mempengaruhi
  - Variabel terikat (Y), yaitu variabel yang dipengaruhi
- ☺ Persamaan regresi adalah persamaan matematika yang memungkinkan kita meramalkan nilai-nilai variabel terikat (Y) dari nilai-nilai satu atau lebih variabel bebas (X)
- ☺ Setelah diketahui ada hubungan/pengaruh, maka analisis ini digunakan untuk keperluan pendugaan variabel terikat (Y) dari suatu nilai variabel bebas (X).
- ☺ Contoh:
  - Nilai stat (Y) dari skor intelegensia (X)
  - Berat badan (Y) dari tinggi badan (X)
  - Tinggi tanaman (Y) dari dosis pupuk (X)
  - Berat badan ikan (Y) dari umur (X)

- ☺ Diambil sampel berukuran n dari populasi:

$$(x_i, y_i) \text{ di mana } i = 1, 2, \dots, n$$

- ☺ Dari sampel tersebut, ingin di uji model regresi:

$$y = \alpha + \beta x$$

di duga dari data sampel, dengan pendugaan:

$$y = a + bx$$

di mana y dan x adalah data pengamatan berpasangan dari sampel, a dan b adalah koefisien regresi (parameter dalam regresi)

- ☺ **Pendugaan parameter**, dengan Metode Kuadrat Terkecil (MKT) (Ordinary Least Square (OLS)) diperoleh:

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

$$a = \frac{\sum_{i=1}^n Y_i}{n} - b \frac{\sum_{i=1}^n X_i}{n}$$

☺ Ada dua koefisien regresi:

- a → dinamakan intersep, titik perpotongan garis dengan sumbu Y (dalam interpretasi, a merupakan nilai konstan dari Y jika variabel X bernilai 0)
- b → dinamakan slope, mengukur kemiringan dari garis regresi.
  - ✓ Bila b positif, (hubungan searah) jika variabel X bertambah besar maka variabel Y bertambah besar pula, sebaliknya jika variabel X bertambah kecil maka variabel Y bertambah kecil pula
  - ✓ Bila b negatif, (hubungan berlawanan arah), jika variabel X bertambah kecil maka variabel Y malah bertambah kecil, demikian sebaliknya

Interpretasi: setiap kenaikan satu unit (satuan) dari X, maka nilai Y akan bertambah/berkurang sebesar b.

☺ Contoh: sebuah penelitian ingin menguji apakah suhu (°C) (X) mempengaruhi banyaknya gula yang terbentuk (Y):

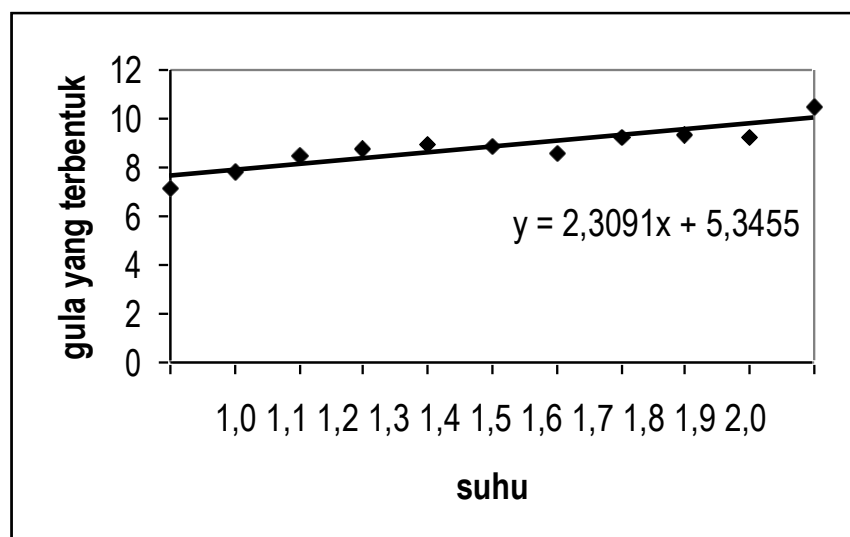
X	1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
Y	7,1	7,8	8,5	8,8	9,0	8,9	8,6	9,2	9,3	9,2	10,5

$$b = \frac{11(147,89) - (16,5)(96,9)}{11(25,85) - (16,5)^2} = 2,309$$

$$a = (96,9/11) - 2,309(16,5/11) = 5,345$$

dengan demikian, garis regresinya adalah:

$$y = 5,345 + 2,309x$$



interpretasi:

- Jika pada suhu 0°C, maka banyaknya gula yang terbentuk adalah 5,345
- Setiap kenaikan 1°C, maka banyaknya gula yang terbentuk akan naik sebesar 2,309

☺ **Pengujian parameter**, ingin diuji apakah koefisien regresi yang terbentuk berarti atau tidak. Terdapat 2 macam uji, yaitu uji simultan dan uji parsial.

☺ **Uji Simultan**, hipotesis yang akan di uji:

$$H_0 : \alpha = \beta = 0$$

$H_1$  : salah satu (baik  $\alpha$ , atau  $\beta$ ) tidak sama dengan nol

Jika  $H_0$  di terima, artinya persamaan regresi tersebut tidak mengandung apa-apa, atau dengan kata lain, tidak ada intersep dan tidak ada slope

Jika  $H_0$  di tolak, artinya persamaan regresi tersebut mengandung suatu arti, apakah  $y = a$ , atau  $y = bx$ , atau  $y = a + bx$

Statistik uji menggunakan F, dapat dideteksi menggunakan tabel analisis ragam:

SK	db	JK	KT	F
Regresi	dbR=1	JKR= $b^2S_{xx}$	KTR=JKR/dbR	KTR/KTG
Galat	dbG=n - 2	JKG= $S_{yy}-b^2S_{xx}$	KTG=JKG/dbG	
Total	dbT=n - 1	JKT= $S_{yy}$		

$$\text{di mana } S_{xx} = \sum_{i=1}^n X_i^2 - \left( \sum_{i=1}^n X_i \right)^2 / n$$

$$S_{yy} = \sum_{i=1}^n Y_i^2 - \left( \sum_{i=1}^n Y_i \right)^2 / n$$

dibandingkan dengan  $F_{\alpha(db1=dbR; db2=dbG)} = F_{\alpha(1; n-2)}$

Jika  $F > F_{\alpha(1; n-2)}$  maka  $H_0$  ditolak, sedangkan jika  $F < F_{\alpha(1; n-2)}$  maka  $H_0$  diterima

☺ **Uji Parsial**, untuk intersep, hipotesis yang diuji:

$$H_0 : \alpha = 0$$

$$H_1 : \alpha \neq 0$$

Jika  $H_0$  di terima, artinya persamaan regresi tersebut tidak melewati salip sumbu, atau jika  $X=0$ , maka  $Y=0$

Jika  $H_0$  di tolak, artinya persamaan regresi tersebut mengandung suatu nilai intersep tertentu

Statistik uji t

$$t = \frac{a}{\sqrt{\frac{KTG \sum_{i=1}^n X_i^2}{nS_{xx}}}}$$

dibandingkan dengan  $t_{\alpha/2}(\text{db galat}) = t_{\alpha/2}(n-2)$

$H_0$  ditolak bila  $t < -t_{\alpha/2}(n-2)$  atau  $t > t_{\alpha/2}(n-2)$  dan  $H_0$  diterima bila

$$-t_{\alpha/2}(n-2) < t < t_{\alpha/2}(n-2)$$

☺ **Uji Parsial**, untuk slope, hipotesis yang diuji:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Jika  $H_0$  di terima, artinya variabel bebas (X) tidak mempengaruhi variabel terikat (Y)

Jika  $H_0$  di tolak, artinya variabel bebas (X) mempengaruhi variabel terikat (Y)

Statistik uji t

$$t = \frac{b}{\sqrt{\frac{KTG}{S_{xx}}}}$$

dibandingkan dengan  $t_{\alpha/2}(db \text{ galat}) = t_{\alpha/2}(n-2)$

$H_0$  ditolak bila  $t < -t_{\alpha/2}(n-2)$  atau  $t > t_{\alpha/2}(n-2)$  dan  $H_0$  diterima bila

$$-t_{\alpha/2}(n-2) < t < t_{\alpha/2}(n-2)$$

☺ Contoh untuk permasalahan sebelumnya:

➤ Uji simultan, diperoleh  $F = 31,722$  (hitung sendiri)

SK	db	JK	KT	F
Regresi	1	5,865	5,865	31,722
Galat	9	1,664	0,185	
Total	10	7,529		

Berdasarkan tabel F, diperoleh  $F_{0,05(1,9)} = 5,12$

Karena nilai  $F > F_{0,05(1,9)}$  maka  $H_0$  ditolak, artinya persamaan regresi tersebut mengandung intersep, atau slope, atau kedua-duanya

➤ Uji parsial untuk intersep,  $t = 8,505$ . Berdasarkan tabel t, diperoleh  $t_{0,025(9)} = 2,262$ . Karena  $t > t_{0,025(9)}$  maka  $H_0$  ditolak, artinya persamaan regresi tersebut mengandung intersep (atau tidak melewati salip sumbu)

➤ Uji parsial untuk slope,  $t = 5,632$ . Berdasarkan tabel t, diperoleh  $t_{0,025(9)} = 2,262$ . Karena  $t > t_{0,025(9)}$  maka  $H_0$  ditolak, artinya adalah benar bahwa suhu mempengaruhi banyaknya gula yang terbentuk.

☺ Koefisien determinasi ( $R^2$ ), yaitu proporsi keragaman (nilai terletak antara 0 dan 1) total nilai-nilai variabel Y yang dijelaskan oleh nilai-nilai X dari hubungan linier tersebut:

$$R^2 = JKR/JKT$$

Artinya jika  $R^2$  (bisa juga ditampilkan dalam %) semakin dekat dengan 1 (atau 100%), maka model regresi tersebut cukup baik untuk digunakan, sedangkan bila  $R^2$  semakin dekat dengan 0 (atau 0%), maka model regresi tersebut tidak cukup baik untuk digunakan.

☺ Dari contoh di atas diperoleh  $R^2 = 5,865/7,529 = 0,779$  atau 77,9%. Berarti persamaan regresi yang terbentuk  $y = 5,345 + 2,309x$  adalah sudah cukup baik, karena variabel tidak bebas (y, yaitu banyaknya gula yang terbentuk) dipengaruhi oleh suhu sebesar 77,9%.

☺ Jika nilai  $R^2$  sudah cukup baik, bisa dilakukan peramalan suatu nilai X terhadap nilai Y. Misalkan saja, ingin diuji berapa banyaknya gula yang terbentuk pada suhu 1,75°C?

$$x = 1,75, \text{ maka } y = 5,345 + 2,309(1,75) = 9,386$$

Artinya pada saat suhu 1,75°C, maka banyaknya gula yang terbentuk adalah 9,386



☺ Ada satu lagi analisis yang menguji hubungan antar variabel, yaitu **analisis korelasi**.

☺ Perbedaannya dengan regresi, dua variabel di analisis korelasi tidak membedakan mana variabel bebas, mana variabel terikat. Dan pada analisis korelasi, kita tidak bisa meramalkan nilai sebagaimana pada analisis regresi. Jadi pada notasi analisis korelasi di sini, variabel X bukan berarti variabel bebas, dan variabel Y bukan berarti variabel terikat.

☺ Koefisien korelasi  $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$

di mana  $S_{xy} = \sum_{i=1}^n X_i Y_i - \left( \sum_{i=1}^n X_i \right) \left( \sum_{i=1}^n Y_i \right) / n$

dan  $S_{xx}$  dengan  $S_{yy}$  udah didefinisikan sebelumnya.

☺ **Uji koefisien korelasi**, hipotesis yang akan di uji:

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

Nilai korelasi terletak antara -1 sampai dengan 1.

- Jika nilai korelasi dekat dengan 0 (korelasi rendah), maka tidak ada korelasi (hubungan) antara nilai X dan nilai Y
- Jika nilai korelasi dekat dengan 1 (korelasi rendah), maka ada korelasi positif (hubungan searah) antara nilai X dan nilai Y
- Jika nilai korelasi dekat dengan -1, maka ada korelasi negatif (hubungan berlawanan arah) antara nilai X dan nilai Y

Statistik uji digunakan  $Z = \frac{\sqrt{n-3}}{2} \ln \left( \frac{(1+r)}{(1-r)} \right)$

di bandingkan dengan  $Z_{\alpha/2}$ .

$H_0$  diterima jika  $-Z_{\alpha/2} < Z < Z_{\alpha/2}$

$H_0$  ditolak jika  $Z < -Z_{\alpha/2}$  atau  $Z > Z_{\alpha/2}$

- ☺ Contoh soal yang di atas, diperoleh  $r = 0,883$ .
- ☺ Setelah di uji hipotesis diperoleh  $Z = 3,929$  dan  $Z_{0,025} = 1,96$ . Karena  $Z > Z_{0,05}$  maka  $H_0$  di tolak, berarti terdapat korelasi positif antara suhu dan banyaknya gula yang terbentuk.

**LEMBAR KERJA :**

1).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

2).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**LEMBAR KERJA :**

3).....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

4).....  
.....  
.....  
.....  
.....  
.....  
.....

.....

.....

.....

.....

.....