



www.esaunggul.ac.id

**Statistik (ESA153)
PERTEMUAN 1
<Dr. Marzuki Silalahi>
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER**

VISI DAN MISI UNIVERSITAS ESA UNGGUL

VISI

Menjadi perguruan tinggi kelas dunia berbasis intelektualitas, kreatifitas dan kewirausahaan yang unggul

MISI

1. Menyelenggarakan pendidikan yang bermutu dan relevan.
2. Menciptakan suasana akademik yang kondusif.
3. Memberikan pelayanan prima kepada seluruh pemangku kepentingan.

Materi Sebelum UTS

01. Pengertian dan Deskripsi Data

02. Probabilitas

03. Distribusi Probabilitas: Peubah acak diskrit

04. Distribusi Probabilitas: Peubah acak kontinu

05. Distribusi *Sampling*

06. Estimasi

07. Hipotesis

Materi Setelah UTS

08. *Analysis of Variance*

09. Regresi dan Korelasi Sederhana

10. Regresi dan Korelasi Ganda

11. Distribusi *Chi-Square* dan analisis frekuensi

12. Statistik *non-Parametrik*

13. Statistik Parametrik dengan SPSS

14. Statistik uji komparatif dan asosiatif dengan SPSS

Pengertian dan Deskripsi Data

- **Tujuan:** Mampu mengambil data, menganalisis data, dan menyimpulkan tentang data tersebut
- **Latar belakang:**
 - Memperoleh data ?
 - Analisis data ?
 - Menyimpulkan data?
secera efektif dan efisien.



(Bidang Ilmu Komputer : Data Mining)

Data mining:

- proses komputasi penemuan pola dalam himpunan data besar (dengan metode statistik dan sistem basisdata).
- mengekstraksi informasi dari sebuah kumpulan data dan mentransformasi ke dalam struktur yang dimengerti.
- Untuk data kecil: dengan Metode Baye's dan Analisis Regresi.
- Data besar dan kompleks: pemrosesan data terotomatisasi dibantu oleh penemuan-penemuan didalam Ilmu Komputer, seperti [neural networks](#), [cluster analysis](#), [genetic algorithms](#), [decision trees](#) dan [decision rules](#), dan [support vector machines](#).
- Pola-pola tersembunyi didalam Himpunan data besa terungkap, terjembatani dari statistik terapan dan inteligensi buatan ke manajemen basisdata (data disimpan dan diberi indeks) dan penemuan algoritma lebih efisien.

Pengertian

Statistika adalah sekumpulan prosedur untuk mengumpulkan, mengukur, mengklasifikasi, menghitung, menjelaskan, mensintesis, menganalisis, dan menafsirkan data yang diperoleh secara sistematis.

Statistik: Suatu kumpulan angka yang tersusun lebih dari satu angka sering dinyatakan atau disajikan dalam bentuk daftar/ tabel, diagram garis, diagram batang, diagram lingkaran, histogram, polygon frekuensi dan ogive yang menggambarkan suatu persoalan tertentu.

- **Statistika Deskriptif** : Menjelaskan atau menggambarkan berbagai karakteristik data seperti berapa rata-rata, seberapa jauh data bervariasi
- **Statistika Induktif (Inferensi)** : Membuat berbagai inferensi terhadap sekumpulan data yang berasal dari suatu sampel. Inferensi → Melakukan perkiraan, peramalan, pengambilan keputusan

Contoh :

- Data tentang penjualan mobil merek 'ABC' perbulan di suatu show room mobil di Jakarta selama tahun 1999. Dari data tersebut pertama akan dilakukan *deskripsi* terhadap data spt menghitung rata-rata penjualan, berapa standar deviasinya dan lain-lain.
- Kemudian baru dilakukan berbagai *inferensi* terhadap hasil deskripsi seperti : perkiraan penjualan mobil tersebut pada bulan Januari tahun berikutnya, perkiraan rata-rata penjualan mobil tersebut di seluruh Indonesia.

Contoh :

- Data tentang penjualan mobil merek 'ABC' perbulan di suatu show room mobil di Jakarta selama tahun 1999. Dari data tersebut pertama akan dilakukan *deskripsi* terhadap data spt menghitung rata-rata penjualan, berapa standar deviasinya dan lain-lain.
- Kemudian baru dilakukan berbagai *inferensi* terhadap hasil deskripsi seperti : perkiraan penjualan mobil tersebut pada bulan Januari tahun berikutnya, perkiraan rata-rata penjualan mobil tersebut di seluruh Indonesia.

Elemen Statistik

- Populasi → masalah dasar dari persoalan statistik.
Definisi : Sekumpulan data yang mengidentifikasi suatu fenomena
- Sampel : Sekumpulan data yang diambil atau diseleksi dari suatu populasi
- Statistik Inferensi : Suatu keputusan, perkiraan atau generalisasi tentang suatu populasi berdasarkan informasi yg terkandung dari suatu sampel.
- Pengukuran Realibilitas → Konsekuensi dari kemungkinan bias dalam inferensi.

Tipe Data Statistik

- I. Data Kualitatif : Data yang bukan berupa angka, ciri : tidak bisa dilakukan operasi matematika. Terbagi dua :
 - a. Nominal
Data yang paling rendah dalam level pengukuran data. Contoh : Jenis kelamin, tgl dan tempat lahir seseorang
 - b. Ordinal → ada tingkatan data. Contoh : Sangat setuju, Setuju, kurang setuju, tidak setuju

Tipe Data Statistik

II. Data Kuantitatif

Data berupa angka dalam arti sebenarnya → dapat dilakukan operasi matematika. Terbagi dua :

a. Data Interval, Contoh : Interval temperatur ruang adalah sbb :

Cukup panas jika antara 50°C - 80°C

Panas jika antara 80°C - 110°C

Sangat panas jika antara 110°C - 140°C

b. Data Rasio → tingkat pengukuran paling 'tinggi' ; bersifat angka dalam arti sesungguhnya. Beda dengan interval mempunyai titik nol dalam arti sesungguhnya.

UKURAN PEMUSATAN DAN LETAK DATA

- Pengukuran pusat data penting dilakukan karena suatu kelompok data bila diurutkan (membesar atau mengecil) maka ada kecenderungan data itu memusat pada bagian tengah.
- Jadi sesuai namanya, ukuran pemusatan data menunjukkan pusat suatu data atau pusat suatu kumpulan pengamatan yang merupakan nilai khas untuk mewakili semua data atau semua pengamatan.

Ukuran pemusatan data:

Rata-rata hitung (arithmetic mean), Median, Modus, Rata-rata ukur (geometric mean), Rata-rata harmonis (harmonic mean)

Ukuran letak data:

Kuartil, Desil, persentil

RATA-RATA HITUNG

Sering disebut dengan rata-rata

Rumus :Rata-rata hitung = (jumlah semua nilai data)/banyaknya nilai data

Jenisnya ada 2:

1. Bila $X_1, X_2, X_3, \dots, X_n$ adalah pengamatan dari sampel, maka rata-rata hitung =

$$X = (X_1 + X_2 + X_3 + \dots + X_n) / n = (\sum X_i) / n \text{ ATAU } (\sum X) / n$$

Contoh Soal 1:

Nilai ujian statistik mahasiswa dalam suatu kelas adalah 70, 75, 60, 65 dan 80. Rata-rata hitungnya: $X = (70 + 75 + 60 + 65 + 80) / 5 = 70$

2. Bila suatu data di mana masing-masing nilai data mengulang dengan frekuensi tertentu: X_1 mengulang dengan f_1

X_2 mengulang dengan f_2

X_3 mengulang dengan f_3

X_n mengulang dengan f_n maka rata-rata hitungnya adalah:

$$X = (f_1X_1 + f_2X_2 + f_3X_3 + \dots + f_n X_n) / (f_1 + f_2 + f_3 + \dots + f_n)$$

Atau:

$$X = \Sigma fX / \Sigma f$$

Di mana : ΣfX adalah jumlah hasil kali antara frekuensi f dengan nilai X
 Σf adalah jumlah frekuensi f .

Contoh soal 2

mengulang dengan frekuensi tertentu:

Pada ujian Bahasa Inggris:3 mahasiswa mendapat nilai 605

mahasiswa mendapat nilai 654 mahasiswa mendapat nilai 801

mahasiswa mendapat nilai 502 mahasiswa mendapat nilai 95maka rata-rata hitungnya adalah:

$$X = ((3 \times 60) + (5 \times 65) + (4 \times 80) + (1 \times 50) + (2 \times 95)) / (3 + 5 + 4 + 1 + 2) \\ = 1.065 / 15 = 71$$

Nilai (X)	Frekuensi (f)	fX
60	3	180
65	5	325
80	4	320
50	1	50
95	2	190
Jumlah	15	1065

$$X = \sum fX / \sum f = 1.065 / 15 = 71$$

Pedoman Umum membuat Tabel Distribusi Frekuensi

Langkah pertama dalam membuat tabel distribusi frekuensi adalah menentukan kelas interval.

Terdapat 3 pedoman yang dapat diikuti:

- a) Berdasarkan Pengalaman, berdasarkan pengalaman jumlah kelas interval yang digunakan dalam menyusun tabel distribusi frekuensi berkisar antara 6 sd 15 kelas.
- b) Ditentukan dengan membaca grafik
- c) Ditentukan dengan rumus Sturges

Rumus Sturges :

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

Dimana :

k = Jumlah Kelas Interval

n = Jumlah data observasi

MEDIAN

Median adalah nilai tengah dari kelompok data yang telah diurutkan (membesar atau mengecil).

Jika jumlah data ganjil, median = nilai paling tengah

Jika jumlah data genap, median = rata-rata dari dua nilai tengah.

Dengan kata lain, median terletak pada nilai ke: $(n/2)$

Contoh soal 3

Median dari data: 3, 4, 4, 5, 6, 8, 8, 9, 10 adalah nilai ke-5, yaitu sama dengan 6, sebab banyaknya data $n = 9$ adalah ganjil.

Ditulis Med = 6

Untuk data berkelompok yang dinyatakan dalam bentuk tabel distribusi frekuensi:

$$\text{Med} = L_0 + c \left(\frac{(n/2) - F}{f} \right)$$

Dimana

Med = median

L_0 = batas bawah kelas median

c = lebar kelas

n = banyaknya data

F = jumlah frekuensi semua kelas sebelum kelas yang mengandung median

f = frekuensi kelas median

HUBUNGAN EMPIRIS ANTARA NILAI RATA-RATA HITUNG, MEDIAN DAN MODUS

Ada tiga hubungan empiris antara nilai rata-rata hitung, median dan modus yang ditentukan oleh kesimetrian kurva distribusi data yang bersangkutan:

- jika nilai rata-rata hitung, median dan modus saling berdekatan, maka kurva adalah mendekati simetri
- jika nilai modus lebih kecil dari median, dan median lebih kecil dari rata-rata hitung, maka distribusi data akan miring ke kanan
- jika nilai rata-rata hitung lebih kecil dari median, dan median lebih kecil dari modus, maka distribusi data akan miring ke kiri.

Untuk distribusi data tidak simetri, yaitu miring kanan atau kiri, hubungan empiris antara nilai rata-rata hitung, median dan modus adalah:

Rata-rata Hitung – Modus = 3 x (Rata-rata Hitung – Median)
atau: $X - \text{Mod} = 3 (X - \text{Med})$ Contoh

Contoh soal 5

Suatu kelompok data diketahui mempunyai distribusi tidak simetri dengan rata-rata hitung (X) = 75,9 dan median (Med) = 77,2.

Tentukan Modus dari data tersebut!

Jawab:

$$\begin{aligned} X - \text{Mod} &= 3 (X - \text{Med}) \\ \text{Mod} &= 3 \text{Med} - 2 X \\ &= 3 (77,2) - 2 (75,9) \\ &= 79,8 \end{aligned}$$

Kuartil : Sekumpulan data dibagi menjadi empat bagian yang sama banyak, sesudah disusun menurut urutan nilainya, maka bilangan pembagiannya disebut *kuartil*.

Rumus mencari kuartil ke i (Q_i): **Letak** $Q_i = \frac{i(n+1)}{4}$, dengan $i = 1, 2, 3$.

Kuartil data berkelompok (tabel Distribusi frekuensi): $Q_i = Tb + \left(\frac{\frac{i}{4}n - F}{f} \right) C$

keterangan: Tb = Tepi bawah kuartil ke- i .

F = Jumlah frekuensi sebelum frekuensi kuartil ke- i .

f = Frekuensi kuartil ke- i . $i = 1, 2, 3$

n = Jumlah seluruh frekuensi.

C = panjang interval kelas.

Rumus Jangkauan kuartil dan simpangan kuartil: $JQ = Q_3 - Q_1$ dan $Qd = \frac{1}{2}(Q_3 - Q_1)$

keterangan: JQ = Simpangan kuartil.

Qd = Jangkauan semi inter kuartil atau simpangan kuartil.

Q1 = Kuartil ke-1 (Kuartil bawah).

Q3 = Kuartil ke-3 (Kuartil atas).

Desil: Tiap pembagi dari kumpulan data (dibagi menjadi sepuluh bagian yang sama)

Letak desil ke i : ~~Letak~~ $D_i = \frac{i(n+1)}{10}$, dengan $i = 1-9$

Data berkelompok:

$$D_i = Tb + \left(\frac{\frac{i}{10}n - F}{f} \right) C$$

keterangan:

Tb = Tepi bawah desil ke- i .

F = Jumlah frekuensi sebelum frekuensi kuartil ke- i .

f = Frekuensi kuartil ke- i dengan $i = 1-9$.

n = Jumlah seluruh frekuensi.

C = panjang interval kelas.

Persentil: Pembagi sekumpulan data yang dibagi menjadi seratus bagian yang sama, disimbolkan menjadi P1, P2, P99

Letak persentil ke i: *Letak* $P_i = \frac{i(n+1)}{100}$, dengan i = 1-99.

Persentil data berkelompok:
$$P_i = Tb + \left(\frac{\frac{i}{100}n - F}{f} \right) C$$

keterangan:

Tb = Tepi bawah persentil ke-i.

F = Jumlah frekuensi sebelum frekuensi kuartil ke-i.

f = Frekuensi kuartil ke-i. i = 1, 2, 3, ..., 99

n = Jumlah seluruh frekuensi.

C = panjang interval kelas

TEKNIK PENGUMPULAN DATA

Teknik (metode) pengumpulan data Merupakan serangkaian prosedur dan instrumen yang digunakan peneliti untuk memperoleh data yang reliabel (akurat, relevan, cukup, dan tepat waktu)

Beberapa teknik (metode) pengumpulan data yang umum digunakan adalah :

- Observasi
- Interview
- Kuesioner

Outlier

Dalam proses collecting data, peneliti (researcher) sering menemukan nilai pengamatan yang bervariasi (beragam).

Keberagaman data ini, di satu sisi sangat dibutuhkan dalam analisa statistika, namun di sisi yang lain keberagaman data menyebabkan adanya nilai pengamatan yang berbeda dengan nilai pengamatan lainnya. Dengan kata lain terdapat beberapa data yang berbeda dengan pola keseluruhan data. Penyebabnya mungkin terdapat kesalahan pada pengamatan, pencatatan, maupun kesalahan yang lain.

Data yang berbeda ini disebut sebagai outlier atau data pencilan.

Populasi vs Sampel

- Populasi : keseluruhan pengamatan
- Sampel = Contoh = sample : himpunan bagian populasi
- Ukuran Populasi = N = banyak anggota populasi
- Ukuran Sampel = n = banyak anggota sampel
- Parameter : nilai yang menyatakan ciri populasi
- Statistik (Statistic) : nilai yang menyatakan ciri sampel

Tabel 1. Notasi Parameter Populasi dan Statistik Sampel

Ciri	Parameter	Statistik
Rata-rata	$\mu = \text{myu}$	\bar{x}
Standar Deviasi, Simpangan Baku	$\sigma = \text{sigma}$	s
Ragam, Variance	σ^2	s^2
proporsi	π	p atau \hat{p}

Kovarian dan Korelasi

Untuk memahami korelasi linier antara dua variabel, terdapat dua elemen yang harus kita tinjau, mengukur hubungan diantara dua variabel (kovarian) dan proses standarisasi.

Kovarian

- Salah satu ukuran kekuatan hubungan linear antara dua variabel acak kontinu adalah dengan menentukan seberapa banyak kedua variabel tersebut co-vary, yaitu bervariasi bersama-sama.
- Jika salah satu variabel meningkat (atau menurun) sebagai akibat peningkatan (atau penurunan) variabel pasangannya, maka dua variabel tersebut dinamakan covary.
- Namun jika satu variabel tidak berubah dengan meningkatnya (atau penurunan) variabel lain, maka variabel tersebut tidak covary.

Kovariansi

Ukuran numerik untuk variansi bersama dua variabel random

$$\begin{aligned}\text{Kov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E(XY) - \mu_X \mu_Y\end{aligned}$$

Korelasi

Kovariansi dibagi dengan standar deviasi X dan standar deviasi Y

$$\text{Kor}(X, Y) = \frac{\text{Kov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

Statistik untuk mengukur berapa banyak kedua variabel covary dalam sampel pengamatan adalah kovarian.

$$\text{Covarian} = S_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

Selain mengukur besarnya kekuatan hubungan di antara dua variabel, kovarian juga menentukan arah hubungan dari kedua variabel tersebut.

1. Apabila nilainya positif, berarti bahwa apabila nilai x berada di atas nilai rata-ratanya, maka nilai y juga berada di atas nilai rata-rata y, dan sebaliknya (Searah).
2. Nilai kovarian negatif menunjukkan bahwa apabila nilai x berada di atas nilai rata-ratanya sedangkan nilai y berada di bawah nilai rata-ratanya (berlawanan arah).
3. Terakhir, apabila nilai kovarian mendekati nol, menandakan bahwa kedua variabel tersebut tidak saling berhubungan.

Standarisasi

Salah satu keterbatasan kovarian sebagai ukuran kekuatan hubungan linier adalah arah/besarnya gradien yang tergantung pada satuan dari kedua variabel tersebut. Misalnya, kovarian antara serapan N (%) dan Hasil Padi (ton) akan jauh lebih besar apabila satuan % (1/100) kita konversi ke ppm (1/sejuta).

Agar nilai kovarian tidak tergantung kepada unit dari masing-masing variabel, maka kita harus membakukannya terlebih dahulu yaitu dengan cara membagi nilai kovarians tersebut dengan nilai standar deviasi dari kedua variabel tersebut sehingga nilainya akan terletak antara -1 dan +1.

Ukuran statistik tersebut dikenal dengan Pearson product moment correlation yang mengukur kekuatan hubungan linier (garis lurus) dari kedua variabel tersebut. Koefisien korelasi linear kadang-kadang disebut sebagai koefisien korelasi pearson untuk menghormati Karl Pearson (1857-1936), yang pertama kali mengembangkan ukuran statistik ini.

Standar Deviasi variabel X dan Y:

$$S_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \text{ dan } S_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}}$$

Korelasi: $Korelasi = r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$

Nilai kovarian distandarkan dengan membagi nilai kovarian tersebut dengan nilai standar deviasi kedua variabel.

$$Korelasi = r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \cdot \sqrt{\sum(y_i - \bar{y})^2}}$$

atau

$$Korelasi = r_{xy} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \cdot \sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}}$$

atau

$$Korelasi = r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \cdot \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

Koefisien Korelasi

Koefisien korelasi mengukur kekuatan dan arah hubungan linier dari dua variabel. Harus diingat bahwa nilai koefisien korelasi yang kecil (tidak signifikan) bukan berarti kedua variabel tersebut tidak saling berhubungan. Mungkin saja dua variabel mempunyai keeratan hubungan yang kuat namun nilai koefisien korelasinya mendekati nol, misalnya pada kasus hubungan non linier. Dengan demikian, koefisien korelasi hanya mengukur kekuatan hubungan linier dan tidak pada hubungan non linier.

SOAL:

1. mengulang dengan frekuensi tertentu: Hitung nilai rata-rata data berikut:

x	8	6	4	5	6	9
f	2	3	4	3	2	1

2. Misalnya modal (dalam jutaan rupiah) dari 40 perusahaan disajikan pada tabel distribusi frekuensi berikut:

Hitung nilai rata-rata hitungannya!

Modal	Nilai Tengah	Frekuensi	fX
112-120	116	4	
121-129	125	5	
130-138	134	8	
139-138	143	12	
148-156	152	5	
157-165	161	4	
166-174	170	2	
		$\Sigma f=40$	$\Sigma fX=$

3. Susun Tabel Distribusi Frekuensi dari data kuliah statistika berikut:

27	79	69	40	51	88	55	48	36	61
53	44	93	51	65	42	58	55	69	63
70	48	61	55	60	25	47	78	61	54
57	76	73	62	36	67	40	51	59	68
27	46	62	43	54	83	59	13	72	57
82	45	54	52	71	53	82	69	60	35
41	65	62	75	60	42	55	34	49	45
49	64	40	61	73	44	59	46	71	86
43	69	54	31	36	51	75	44	66	53
80	71	53	56	91	60	41	29	56	57
35	54	43	39	56	27	62	44	85	61
59	89	60	51	71	53	58	26	77	68
62	57	48	69	76	52	49	45	54	41
33	61	80	57	42	45	59	44	68	73
55	70	39	59	69	51	85	46	55	67

4. Berapa median untuk himpunan bilangan: 5, 5, 7, 9, 11, 12, 15, 18?
5. Jika data upah harian 8 karyawan (dalam ribuan) adalah sbb: 20, 80, 75, 60, 50, 85, 45, 90, Berapa median-nya?
6. Tentukan median data modal 40 perusahaan pada tabel dibawah:

Modal	Nilai Tengah
112-120	116
121-129	125
130-138	134
139-138	143
148-156	152
157-165	161
166-174	170

Kesimpulan:

1. Statistika adalah keseluruhan prosedur untuk mendapatkan data, menganalisis data dan menyimpulkan data.
2. Statistik dibagi menjadi dua bagian utama yaitu statistik deskriptif dan statistik inferensi

Referensi:

1. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye, 2007, *Probability and Statistics for Engineers and Scientists*, 8th edition, Pearson Prentice Hall.
2. Sharma, Subhash, 1996, *Applied Multivariate Techniques*, John Wiley & Son, Inc., USA.
3. Johson & Wichern, 2007, *Applied multivariate statistical analysis*, Upper Saddle River: Pearson Prentice Hall.
4. J. Supranto, M.A. ,2001, *Statistika Teori dan Aplikasi*, Erlangga, Jakarta.
5. Douglas C. Montgomery, George C. Runger, 2003, *Applied Statistic and Probability for Engineer*, third edition, John Wiley and Son Inc.
6. Singgih Santoso, 2014, *Panduan Lengkap SPSSversi 20*, Alex Media Komputindo.