STATISTIKA DESKRIPTIF

UKURAN PEMUSATAN & UKURAN PENYEBARAN DATA

MK Statistik dan Probabilitas Teknik Informatika PENS

Prof. Dr. Arna Fariza, S.Kom., M.Kom.

Topik Bahasan

- **Ukuran Pemusatan Data**
  - Mean (Rata-rata) untuk Data Tunggal & Berkelompok
  - Median (Nilai Tengah) untuk Data Tunggal & Berkelompok
  - Modus (Nilai yang Sering Muncul) untuk Data Tunggal & Berkelompok
- **Ukuran Penyebaran Data**
  - Range (Jangkauan)
  - Varians
  - Standar Deviasi
  - Koefisien Variasi
- **Aplikasi di Bidang Teknik Informatika**
  - Analisis Waktu Respons Server
  - Analisis Kinerja Sistem
  - Deteksi Anomali Data

MENGAPA UKURAN PEMUSATAN & PENYEBARAN?

Statistika Deskriptif bertujuan untuk menggambarkan karakteristik data.
Dua Aspek Penting dalam Analisis Data:

Ukuran Pemusatan	Ukuran Penyebaran
Menunjukkan nilai pusat/tengah data	Menunjukkan variasi/keragaman data
"Di mana letak pusat data?"	"Seberapa tersebar data?"
Contoh: Nilai rata-rata ujian	Contoh: Apakah nilai siswa homogen?

Kedua data memiliki mean yang sama (75), tetapi standar deviasi berbeda

UKURAN PEMUSATAN - MEAN (RATA-RATA) DATA TUNGGAL

Mean (Rata-rata)

Jumlah seluruh nilai data dibagi dengan banyaknya data.

Rumus:

$\bar{x} = (x_1 + x_2 + \ldots + x\square) / n = \Sigma x_i / n$

Contoh:

Nilai ujian 8 mahasiswa: 70, 75, 80, 85, 90, 85, 75, 80

$\bar{x} = (70+75+80+85+90+85+75+80) / 8$

$= 640 / 8$

$= 80$

Interpretasi: Rata-rata nilai ujian mahasiswa adalah 80.

Kelebihan:

- Mempertimbangkan semua nilai data
- Mudah dihitung dan dipahami

Kelemahan:

- Sensitif terhadap nilai ekstrem (outlier)

MEAN DATA TUNGGAL - CONTOH DENGAN OUTLIER

Kasus: Pengaruh Outlier terhadap Mean

Data gaji karyawan (dalam juta rupiah):

5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, **100**

Mean tanpa outlier:

$\bar{x}$ = (5+6+7+8+9+10+11+12+13+14+15+16+17+18) / 14

= 161 / 14

= **11.5**

Mean dengan outlier:

$\bar{x}$ = (5+6+7+8+9+10+11+12+13+14+15+16+17+18+100) / 15

= 261 / 15

= **17.4**

Kesimpulan:

- Mean berubah drastis dengan adanya outlier
- Data "100" membuat mean menjadi tidak representatif
- Dalam kasus ini, median mungkin lebih baik digunakan!

MEAN DATA BERKELOMPOK

Mean Data Berkelompok

Nilai rata-rata dari data yang telah disusun dalam tabel distribusi frekuensi.

Rumus: $\bar{x} = \Sigma(f_i \times x_i) / \Sigma f_i$

di mana:

- f_i = frekuensi kelas ke-i
- x_i = titik tengah kelas ke-i

Contoh: Data Nilai 20 Mahasiswa

$\bar{x} = 1620 / 20 =$ **81**

Kelas Interval	f	x_i (Titik Tengah)	$f \times x_i$
65 - 70	3	67.5	202.5
71 - 76	3	73.5	220.5
77 - 82	5	79.5	397.5
83 - 88	5	85.5	427.5
89 - 94	3	91.5	274.5
95 - 100	1	97.5	97.5
Jumlah	$\Sigma f=20$		$\Sigma fx=1620$

MEDIAN DATA TUNGGAL

Median adalah nilai yang membagi data menjadi dua bagian yang sama besar setelah data diurutkan (50% data di bawahnya, 50% di atasnya).

Langkah-langkah:

1. Urutkan data dari terkecil ke terbesar
2. Tentukan posisi median

Rumus:

Data Ganjil	Data Genap
Me = x_{(n+1)/2}	Me = (x_{n/2} + x_{n/2 + 1}) / 2

Contoh 1: Data Ganjil

Data: 65, 70, 75, 80, 85, 90, 95

n = 7

Me = x_{(7+1)/2} = x_4 = **80**

Contoh 2: Data Genap

Data: 65, 70, 75, 80, 85, 90

n = 6

Me = (x_3 + x_4) / 2 = (75 + 80) / 2 = **77.5**

Kelebihan:

- Tidak terpengaruh oleh outlier
- Cocok untuk data miring (skewed)

MEDIAN DATA BERKELOMPOK

Median Data Berkelompok

Median untuk data yang telah disusun dalam tabel distribusi frekuensi.

Rumus:

$$Me = Tb + ((n/2 - F) / f\square) \times p$$

di mana:

- Tb = Tepi bawah kelas median
- n = jumlah total data
- F = frekuensi kumulatif sebelum kelas median
- f□ = frekuensi kelas median
- p = panjang kelas interval

Langkah-langkah:

1. Tentukan letak kelas median: n/2
2. Identifikasi kelas median (kelas di mana nilai n/2 berada)
3. Gunakan rumus di atas

MEDIAN DATA BERKELOMPOK

Data Nilai 20 Mahasiswa (Tabel Distribusi Frekuensi)

Langkah:

1. Letak median = n/2 = 20/2 = 10
2. Kelas median: 77 - 82 (frekuensi kumulatif 11 > 10)
3. Tb = 77 - 0.5 = 76.5
4. F = 6 (frek. kumulatif sebelum kelas median)
5. f□ = 5
6. p = 6

Me = 76.5 + ((10 - 6) / 5) × 6
= 76.5 + (4/5) × 6
= 76.5 + 4.8
= **81.3**

Kelas	f	f kumulatif
65 - 70	3	3
71 - 76	3	6
77 - 82	5	11
83 - 88	5	16
89 - 94	3	19
95 - 100	1	20
Total	20	

MODUS DATA TUNGGAL

Modus adalah nilai yang paling sering muncul dalam suatu data.

Karakteristik Modus:

- Bisa lebih dari satu (multimodus)
- Dapat digunakan untuk data kualitatif dan kuantitatif

Kelebihan:

- Sederhana dan mudah dipahami
- Tidak terpengaruh outlier

Kelemahan:

- Tidak selalu ada (kadang tidak ada modus)
- Kurang stabil (perubahan kecil data dapat mengubah modus)

Data	Modus	Keterangan
5, 7, 8, 8, 10, 11, 13, 13, 13, 15	13	Muncul 3 kali (unimodus)
5, 7, 8, 8, 10, 11, 13, 13, 15	8 dan 13	Masing-masing muncul 2 kali (bimodus)
5, 7, 8, 10, 11, 13, 15	Tidak ada	Semua nilai muncul 1 kali

MODUS DATA BERKELOMPOK

Modus Data Berkelompok

Modus untuk data yang telah disusun dalam tabel distribusi frekuensi.

Rumus:

$Mo = Tb + (d_1 / (d_1 + d_2)) \times p$

di mana:

- Tb = Tepi bawah kelas modus
- d_1 = selisih frekuensi kelas modus dengan kelas sebelumnya
- d_2 = selisih frekuensi kelas modus dengan kelas sesudahnya
- p = panjang kelas interval

Langkah-langkah:

1. Identifikasi kelas modus (kelas dengan frekuensi tertinggi)
2. Gunakan rumus di atas

MODUS DATA BERKELOMPOK

Data Nilai 20 Mahasiswa (Tabel Distribusi Frekuensi)

Kelas Interval	f
65 - 70	3
71 - 76	3
77 - 82	5 ← Kelas modus (f=5)
83 - 88	5 ← Bisa juga ini (multimodus)
89 - 94	3
95 - 100	1

Kasus 1: Menggunakan kelas 77-82

- Tb = 77 - 0.5 = 76.5
- $d_1 = 5 - 3 = 2$
- $d_2 = 5 - 5 = 0$
- p = 6

Mo = 76.5 + (2/(2+0)) × 6
= 76.5 + 6
= **82.5**

Kasus 2: Menggunakan kelas 83-88

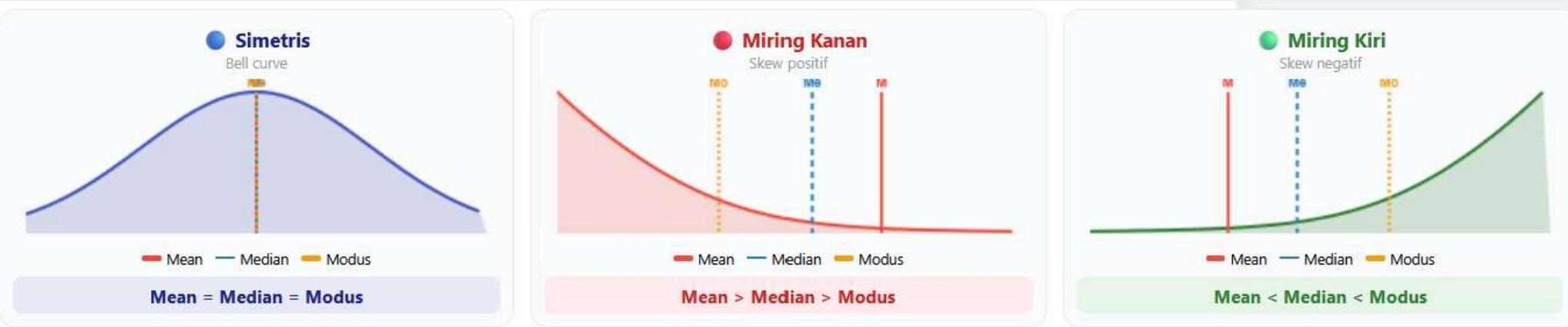
- Tb = 83 - 0.5 = 82.5
- $d_1 = 5 - 5 = 0$
- $d_2 = 5 - 3 = 2$
- p = 6

Mo = 82.5 + (0/(0+2)) × 6
= **82.5**

Catatan: Karena ada 2 kelas dengan frekuensi sama, data ini **bimodus** dengan nilai modus sekitar 82.5.

PERBANDINGAN MEAN, MEDIAN, MODUS

Visualisasi Perbandingan



Kapan Menggunakan:

Ukuran	Kelebihan	Kapan Digunakan
Mean	Mempertimbangkan semua data	Data simetris, tidak ada outlier
Median	Tidak terpengaruh outlier	Data miring, ada outlier
Modus	Mudah, untuk data kategorik	Data kualitatif, distribusi multimodal

RANGE (JANGKAUAN)

Range (Jangkauan)

Selisih antara nilai maksimum dan nilai minimum dalam data.

Rumus: Range = Nilai Maksimum - Nilai Minimum

Contoh 1 - Data Tunggal:

Data: 65, 70, 75, 80, 85, 90, 95

Range = 95 - 65 = **30**

Contoh 2 - Data Berkelompok:

Range = Nilai Maksimum Kelas Terakhir - Nilai Minimum Kelas Pertama

Data Nilai 20 Mahasiswa:

Kelas pertama: 65-70, Kelas terakhir: 95-100

Range = 100 - 65 = **35**

Kelebihan:

- Sangat mudah dihitung dan dipahami

Kelemahan:

- Sangat sensitif terhadap outlier
- Hanya mempertimbangkan dua nilai ekstrem
- Tidak memberikan informasi tentang penyebaran data di antara keduanya

VARIANS DATA TUNGGAL

Varians (σ^2 atau s^2)

Ukuran penyebaran data yang menunjukkan rata-rata kuadrat jarak setiap data terhadap mean.

Rumus (Populasi): $\sigma^2 = \Sigma(x_i - \mu)^2 / N$

Rumus (Sampel): $s^2 = \Sigma(x_i - \bar{x})^2 / (n - 1)$

Perbedaan Populasi vs Sampel:

Populasi	Sampel
$\sigma^2 = \Sigma(x_i - \mu)^2 / N$	$s^2 = \Sigma(x_i - \bar{x})^2 / (n-1)$
Menggunakan μ (mean populasi)	Menggunakan $\bar{x}$ (mean sampel)
Dibagi N	Dibagi (n-1)

Catatan: Pembagian dengan (n-1) disebut **degrees of freedom** (derajat kebebasan).

VARIANS DATA TUNGGAL

Data Nilai 8 Mahasiswa: 70, 75, 80, 85, 90, 85, 75, 80
Langkah 1: Hitung mean $\bar{x} = 640/8 = 80$
Langkah 2: Hitung selisih kuadrat

No	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	70	-10	100
2	75	-5	25
3	80	0	0
4	85	5	25
5	90	10	100
6	85	5	25
7	75	-5	25
8	80	0	0

$\Sigma(x_i - \bar{x})^2 =$ **300**
Varians Sampel:
$s^2 = 300 / (8-1) = 300/7 =$ **42.86**
Interpretasi: Rata-rata kuadrat jarak data dari mean adalah 42.86.

STANDAR DEVIASI DATA TUNGGAL

Standar Deviasi (σ atau s)

Akar kuadrat dari varians. Mengembalikan ukuran ke satuan yang sama dengan data.

Rumus (Populasi): $\sigma = \sqrt{\sigma^2} = \sqrt{(\Sigma(x_i - \mu)^2 / N)}$

Rumus (Sampel): $s = \sqrt{s^2} = \sqrt{(\Sigma(x_i - \bar{x})^2 / (n-1))}$

Contoh (dari data sebelumnya):

$s^2 = 42.86$

$s = \sqrt{42.86} =$ **6.55**

Interpretasi:

- Data menyebar sekitar **6.55** dari mean (80)
- Semakin kecil standar deviasi, semakin homogen data

VARIANS & STANDAR DEVIASI DATA BERKELOMPOK

Varians Data Berkelompok

Sama seperti data tunggal, tetapi menggunakan titik tengah kelas (x_i) dan frekuensi (f_i).

Rumus (Sampel): $s^2 = \Sigma f_i(x_i - \bar{x})^2 / (n - 1)$

Standar Deviasi Data Berkelompok: $s = \sqrt{s^2}$

Langkah-langkah:

1. Hitung mean data berkelompok ($\bar{x}$)
2. Hitung selisih setiap titik tengah dengan mean
3. Kuadratkan dan kalikan dengan frekuensi
4. Jumlahkan dan bagi dengan (n-1)
5. Akar kuadratkan untuk mendapatkan standar deviasi

VARIANS & STANDAR DEVIASI DATA BERKELOMPOK

Data Nilai 20 Mahasiswa

Kelas	f	x_i	$x_i - \bar{x}$	$(x_i-\bar{x})^2$	$f\times(x_i-\bar{x})^2$
65-70	3	67.5	-13.5	182.25	546.75
71-76	3	73.5	-7.5	56.25	168.75
77-82	5	79.5	-1.5	2.25	11.25
83-88	5	85.5	4.5	20.25	101.25
89-94	3	91.5	10.5	110.25	330.75
95-100	1	97.5	16.5	272.25	272.25
Jumlah	20				Σ=1431

$\bar{x}$ = 81 (dari perhitungan sebelumnya)
s^2 = 1431 / (20-1) = 1431/19 = **75.32**
s = $\sqrt{75.32}$ = **8.68**

KOEFISIEN VARIASI (CV)

Koefisien Variasi (CV)
Ukuran penyebaran relatif yang menyatakan standar deviasi sebagai persentase dari mean.
Rumus: $CV = (s / \bar{x}) \times 100\%$
Kegunaan:

- Membandingkan penyebaran dua data dengan satuan berbeda
- Membandingkan penyebaran dua data dengan mean berbeda

Contoh:
Data Gaji Karyawan:

- Mean = Rp 5.000.000
- Standar Deviasi = Rp 1.000.000
- CV = (1.000.000 / 5.000.000) × 100% = **20%**

Interpretasi: Standar deviasi gaji adalah 20% dari rata-rata gaji.
Kriteria CV:

- CV ≤ 10%: Data sangat homogen
- 10% < CV ≤ 30%: Data cukup homogen
- CV > 30%: Data heterogen/bervariasi

PERBANDINGAN CV

Kasus: Membandingkan Dua Variabel Berbeda

Variabel	Mean	Standar Deviasi	CV
Gaji (Rp)	5.000.000	1.000.000	20%
Masa Kerja (tahun)	5	1.5	30%

Analisis:

- Meskipun standar deviasi gaji (1.000.000) lebih besar dari masa kerja (1.5), **CV gaji (20%) < CV masa kerja (30%)**
- Artinya: **Masa kerja lebih bervariasi** dibandingkan gaji
- Pengambilan keputusan harus mempertimbangkan CV, bukan standar deviasi absolut!

Aplikasi di IT:

- Membandingkan variasi waktu respons server dengan traffic berbeda
- Membandingkan variasi performa di berbagai lingkungan pengujian

RINGKASAN FORMULA

Ukuran	Data Tunggal	Data Berkelompok
Mean	$\bar{x} = \Sigma x_i / n$	$\bar{x} = \Sigma fx / n$
Median	n ganjil: x_{(n+1)/2}	Me = Tb + ((n/2-F)/f□)×p
Modus	Nilai paling sering	Mo = Tb + $(d_1/(d_1+d_2))$×p
Range	Max - Min	Max - Min
Varians	$s^2 = \Sigma(x_i-\bar{x})^2/(n-1)$	$s^2 = \Sigma f(x-\bar{x})^2/(n-1)$
SD	$s = \sqrt{s^2}$	$s = \sqrt{s^2}$
CV	$(s/\bar{x}) \times 100\%$	$(s/\bar{x}) \times 100\%$

APLIKASI DI BIDANG TEKNIK INFORMATIKA

1. Analisis Waktu Respons Server

Data Waktu Respons (ms): 120, 125, 130, 128, 122, 300, 127, 126
Mean = 147.25 ms
Median = 127 ms ← Lebih representatif (outlier 300)
SD = 61.9 ms
CV = 42% ← Waktu respons cukup bervariasi

2. Analisis Kinerja Sistem

Data Jumlah Request per Menit:
Server A: 95, 100, 105, 98, 102 → CV ≈ 3.5% (stabil)
Server B: 80, 120, 90, 130, 110 → CV ≈ 19.7% (fluktuatif)
Server A lebih stabil dan dapat diandalkan!

3. Deteksi Anomali (Outlier Detection)

Data di luar **mean ± 2×SD** dapat dianggap sebagai anomali.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt


# Data nilai 20 mahasiswa
data = [65, 67, 70, 72, 74, 75, 78, 78, 79, 80,
        82, 83, 85, 85, 88, 88, 90, 91, 92, 95]


# 1. UKURAN PEMUSATAN
print("=== UKURAN PEMUSATAN ===")
print(f"Mean: {np.mean(data):.2f}")
print(f"Median: {np.median(data):.2f}")


# Modus (menggunakan pandas)
mode_result = pd.Series(data).mode()
print(f"Modus: {mode_result.tolist()}")


# 2. UKURAN PENYEBARAN
print("\n=== UKURAN PENYEBARAN ===")
print(f"Range: {np.max(data) - np.min(data)}")
print(f"Varians: {np.var(data, ddof=1):.2f}")  # ddof=1 untuk sampel
print(f"Standar Deviasi: {np.std(data, ddof=1):.2f}")
print(f"Koefisien Variasi: {(np.std(data, ddof=1)/np.mean(data))*100:.2f}%")


# 3. VISUALISASI
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 5))


# Histogram
ax1.hist(data, bins=6, color='#3498db', edgecolor='black', alpha=0.7)
ax1.axvline(np.mean(data), color='red', linewidth=2, label=f'Mean = {np.mean(data):.1f}')
ax1.axvline(np.median(data), color='green', linewidth=2, label=f'Median = {np.median(data):.1f}')
ax1.set_title('Histogram Nilai Mahasiswa')
ax1.legend()


# Boxplot
ax2.boxplot(data, vert=False)
ax2.set_title('Boxplot Nilai Mahasiswa')


plt.tight_layout()
plt.show()
```

Ringkasan

Ukuran Pemusatan Data:

1. **Mean** → Rata-rata, sensitif terhadap outlier
2. **Median** → Nilai tengah, tidak sensitif terhadap outlier
3. **Modus** → Nilai paling sering muncul, untuk data kategorik

Ukuran Penyebaran Data:

4. **Range** → Selisih max-min, sangat sederhana
5. **Varians** → Rata-rata kuadrat jarak dari mean
6. **Standar Deviasi** → Akar varians, dalam satuan data
7. **Koefisien Variasi** → SD relatif terhadap mean (dalam %)

MOTIVASI

- Memahami ukuran pemusatan dan penyebaran adalah **fondasi analisis data**
- Digunakan dalam **monitoring sistem, deteksi anomali, dan optimasi performa**
- Sangat berguna untuk **Data Science, Machine Learning, dan DevOps**

LATIHAN SOAL

Soal 1 (Data Tunggal)
Data waktu booting (detik) dari 10 komputer: 12, 15, 14, 13, 16, 18, 14, 15, 13, 17
Hitunglah:
a) Mean, Median, Modus
b) Range, Varians, Standar Deviasi, Koefisien Variasi

Soal 2 (Data Berkelompok)
Tabel distribusi frekuensi waktu loading aplikasi (detik):

Interval Waktu	f
2.0 - 2.9	5
3.0 - 3.9	8
4.0 - 4.9	12
5.0 - 5.9	10
6.0 - 6.9	5

Hitunglah:
a) Mean, Median, Modus
b) Varians, Standar Deviasi, Koefisien Variasi

LATIHAN SOAL

Soal 3 (Aplikasi IT)
Anda adalah seorang DevOps Engineer yang menganalisis waktu respons dari 2 server:
| Server A (ms) | 45 | 50 | 48 | 52 | 47 | 51 | 49 | 53 | 46 | 49 |
| Server B (ms) | 30 | 70 | 40 | 60 | 50 | 80 | 35 | 65 | 45 | 75 |
Server mana yang lebih stabil (konsisten)? Gunakan CV sebagai indikator!

Soal 4
Dataset: Data waktu loading aplikasi (detik) dari 50 pengguna:

Interval Waktu (detik)	Jumlah Pengguna
1.0 - 1.9	5
2.0 - 2.9	10
3.0 - 3.9	15
4.0 - 4.9	12
5.0 - 5.9	5
6.0 - 6.9	3

- Hitunglah **Mean, Median, dan Modus** dari data tersebut!
- Hitunglah **Range, Varians, Standar Deviasi, dan Koefisien Variasi**!
- Buatlah visualisasi (histogram + boxplot) menggunakan Excel/Python!
- Berikan **interpretasi** dari hasil perhitungan Anda!
- Jika target loading aplikasi adalah < 3 detik, berapa persentase pengguna yang mengalami loading di atas target