

# EKSPEKTASI



Achmad Basuki  
Politeknik Elektronika Negeri Surabaya  
2004

# Gambaran Permasalahan (1)

Jam	Jumlah bemo
06.00 - 06.30	5
06.30 - 07.00	9
07.00 - 07.30	7
07.30 - 08.00	7
08.00 - 08.30	5
08.30 - 09.00	4
09.00 - 09.30	2
09.30 - 10.00	4
10.00 - 10.30	3
10.30 - 11.00	2
11.00 - 11.30	4
11.30 - 12.00	1
12.00 - 12.30	3
12.30 - 13.00	6
13.00 - 13.30	5
13.30 - 14.00	4
14.00 - 14.30	7
14.30 - 15.00	3
15.00 - 15.30	3
15.30 - 16.00	6

Hasil pengukuran jumlah bemo yang lewat di depan PENS-ITS setiap setengah jam yang dilakukan mulai jam 6.00 sampai jam 18.00.

Berapa jumlah bemo yang lewat di depan PENS-ITS setiap setengah jam ?

# Gambaran Permasalahan (2)

Nilai Yudistira dan Abimanyu mahasiswa PENS-ITS semester 3 jurusan TI dalam satu semester adalah sebagai berikut:

YUDISTIRA		
Nama Mata Kuliah	SKS	Nilai
Bahasa Inggris 3	2	B
Matematika 3	2	B
Pemrograman Berbasis Obyek	2	A
Metode Numerik	2	B
Statistik dan Probabilitas	2	BC
Basis Data 1	2	B
Model dan Sistem Informasi	2	A
Rekayasa Perangkat Lunak	2	A
Prak. Pemrg. Berbasis Obyek	1	A
Prak. Metode Numerik	1	B
Prak. Basis Data 1	1	B

ABIMANYU		
Nama Mata Kuliah	SKS	Nilai
Bahasa Inggris 3	2	A
Matematika 3	2	B
Pemrograman Berbasis Obyek	2	BC
Metode Numerik	2	B
Statistik dan Probabilitas	2	A
Basis Data 1	2	B
Model dan Sistem Informasi	2	B
Rekayasa Perangkat Lunak	2	BC
Prak. Pemrograman Berbasis C	1	B
Prak. Metode Numerik	1	A
Prak. Basis Data 1	1	A

Bagaimana cara menentukan siapa yang nilainya lebih baik ?

# Ekspektasi

- Ekspektasi menyatakan suatu nilai harapan terhadap distribusi data tertentu.
- Dengan nilai ekspektasi akan diperoleh gambaran distribusi data, yang berupa besaran suatu data.
- Nilai ekspektasi dipilih berdasarkan model data, apakah data berupa data-data numerik atau katagorikal yang berhubungan dengan model statistik parametrik atau statistik non-parametrik.

# Nilai Ekspektasi

- **Ekspektasi 1:** Pusat data (*center of data*) yang dinyatakan dengan nilai rata-rata, median dan modus.
- **Ekspektasi 2:** Sebaran data (*dispersion of data*) yang dinyatakan dengan nilai standard deviasi dan varians.
- **Ekspektasi 3:** Kecenderungan yang dinyatakan dengan nilai skewness.



*Ekspektasi 1 dan 2 merupakan titik berat bahasan*

**Achmad Basuki**

# Ekspektasi 1

## Center of Data

Menyatakan nilai yang merupakan pusat atau nilai pokok dari suatu distribusi data tertentu.

- Rata-Rata (Mean)
- Nilai Tengah (Median)
- Modus

# Rata-Rata (Mean)

Model Kontinu: 
$$\bar{x} = \int_C x \cdot f(x) dx$$

Dimana  $X$  adalah kejadian yang jumlahnya tak berhingga, atau mempunyai nilai numerik.

$f(x)$  menyatakan pdf dari  $x$ .

Model Diskrit: 
$$\bar{x} = \sum_{i=1}^n x_i p(x_i)$$

$p(x_i)$  menyatakan pdf dari  $X=x_i$ .

Model diskrit ini yang banyak dibahas dalam analisa data.

# Rata-Rata (*Cont...*)

Diketahui data jumlah bemo yang lewat di depan PENS-ITS setiap 30 menit

Jam	Jumlah bemo
06.00 - 06.30	5
06.30 - 07.00	9
07.00 - 07.30	7
07.30 - 08.00	7
08.00 - 08.30	5
08.30 - 09.00	4
09.00 - 09.30	2
09.30 - 10.00	4
10.00 - 10.30	3
10.30 - 11.00	2
11.00 - 11.30	4
11.30 - 12.00	1
12.00 - 12.30	3
12.30 - 13.00	6
13.00 - 13.30	5
13.30 - 14.00	4
14.00 - 14.30	7
14.30 - 15.00	3
15.00 - 15.30	3
15.30 - 16.00	6

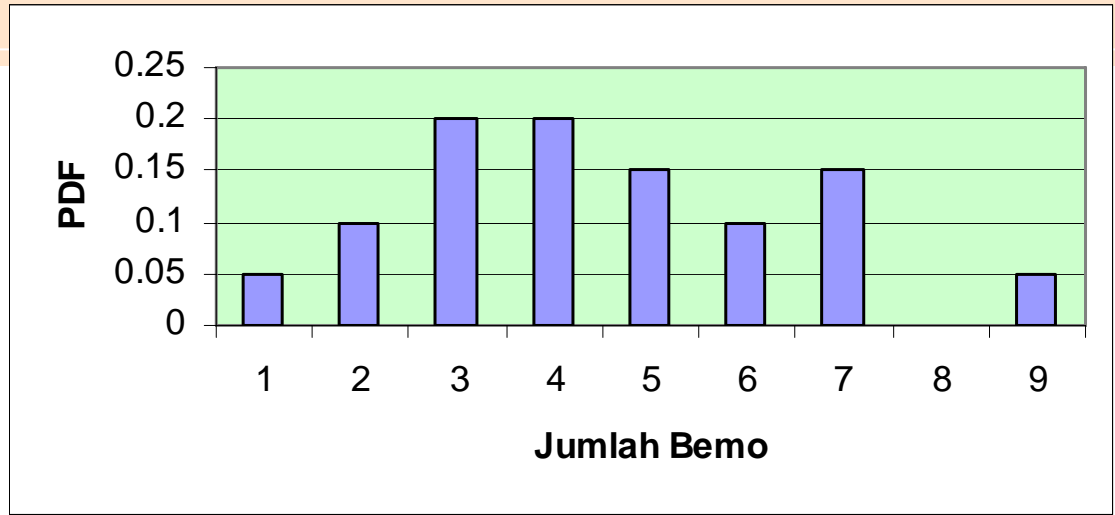
Rata-rata jumlah bemo yang lewat di depan PENS-ITS dalam setiap 30 menit adalah:

$$\begin{aligned}\bar{x} &= \frac{1}{N} \sum_{i=1}^N x_i \\ &= \frac{1}{20} (90) = 4.5\end{aligned}$$

Teknik perhitungan merupakan teknik perhitungan yang biasa digunakan dalam menghitung rata-rata langsung dari data, tetapi teknik ini tidak melibatkan pdf sehingga tidak ada informasi lain selain rata-rata ☹️

# Rata-Rata (*Cont...*)

Jumlah bemo	Histogram	PDF
1	1	0.05
2	2	0.1
3	4	0.2
4	4	0.2
5	3	0.15
6	2	0.1
7	3	0.15
8	0	0
9	1	0.05



Perhitungan rata-rata menggunakan:

$$\bar{x} = \sum_{i=1}^n x_i f(x_i)$$

Jumlah bemo	Histogram	PDF	J X P
1	1	0.05	0.05
2	2	0.1	0.2
3	4	0.2	0.6
4	4	0.2	0.8
5	3	0.15	0.75
6	2	0.1	0.6
7	3	0.15	1.05
8	0	0	0
9	1	0.05	0.45
<b>Total</b>	<b>20</b>		<b>4.5</b>

# Rata-Rata (*Cont...*)

(1)

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

(2)

$$= \sum_{i=1}^N x_i \left( \frac{1}{N} \right)$$



$$\bar{x} = \sum_{i=1}^n x_i f(x_i)$$

Ini berarti pada rumus (1) menyatakan bahwa  $f(x) = 1/N$ , atau dengan kata lain peluang setiap kejadian  $X$  adalah sama

Banyak kasus dalam kehidupan yang setiap kejadian mempunyai peluang yang berbeda, apakah ini bisa diabaikan dan dianggap bahwa semua kejadian mempunyai peluang yang sama seperti halnya pelemparan coin dan dadu ☹️ ?

# Median (Nilai Tengah)

Median adalah  $x$  dimana:  $CDF(x) = 1/2$

Median banyak digunakan untuk keperluan mencari pusat data dalam data-data hasil kuantisasi.

Model kontinu:  $x = median \Rightarrow \int_{-\infty}^x f(u) du = \frac{1}{2}$

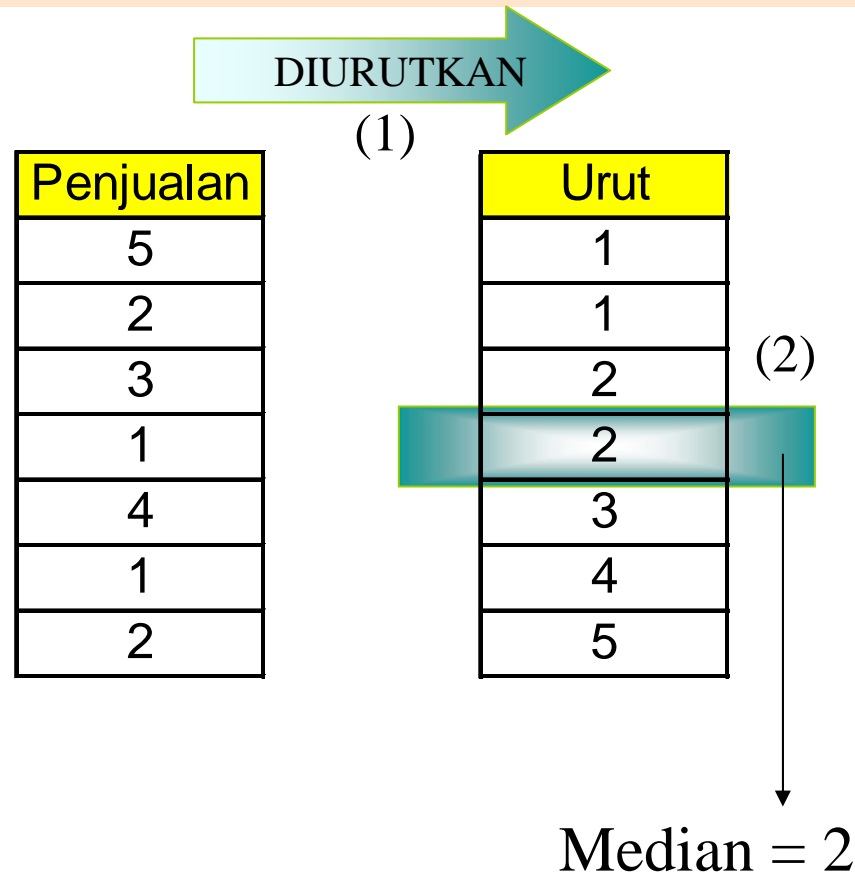
Model diskrit:  $x = median \Rightarrow \sum_{i=0}^x p(x) = \frac{1}{2}$



Achmad Basuki → Median = 4

# Median (*cont...*)

Data penjualan PC dalam 7 hari terakhir adalah:



## Algoritma:

- (1) *Data diurutkan*
- (2) *Median diambil pada data di posisi tengah*

# Median (*cont...*)

## Algoritma Menghitung Median:

- (1) *Data diurutkan*
- (2) *Median diambil pada data di posisi tengah*

- Pengurutan data membuat proses perhitungan menjadi lambat untuk data-data yang berukuran besar. Untuk data berukuran  $n$ , metode Bubble Sort diperlukan waktu  $n(n-1)$  kali proses, dan quick sort membutuhkan waktu  $n \cdot \ln(n)$
- Pengambilan data pada posisi tengah membutuhkan pengecekan apakah jumlah data ( $n$ ) genap atau ganjil.

## Kesimpulan:

Algoritma ini lambat, diperlukan algoritma yang lebih cepat, yang tidak melibatkan proses pengurutan data.



# Median (*cont...*)

Data penjualan PC dalam 7 hari terakhir adalah:

Penjualan
5
2
3
1
4
1
2



Penjualan	Histogram	PDF	CDF
1	2	0.29	0.29
2	2	0.29	0.57
3	1	0.14	0.71
4	1	0.14	0.86
5	1	0.14	1.00

(2)

Median = 2

## Algoritma:

(1) *Hitung CDF*

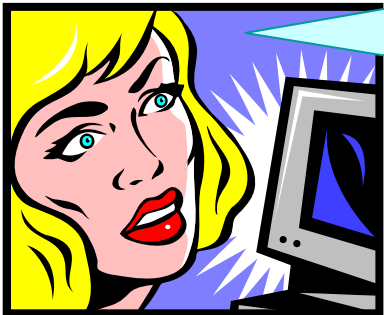
(2) *Median diambil pada data yang nilainya mengandung nilai  $\frac{1}{2}$ .*

# Median (*cont...*)

## Algoritma Menghitung Median Yang Disarankan :

- (1) *Hitung CDF*
- (2) *Median diambil pada data yang nilainya mengandung nilai  $\frac{1}{2}$ .*

- Perhitungan CDF membutuhkan waktu  $n$  untuk  $n$  buah data.
- Pengambilan data pada posisi tengah tidak membutuhkan pengecekan apakah jumlah data ( $n$ ) genap atau ganjil.



### Kesimpulan:

Algoritma ini cepat karena tidak melalui proses pengurutan data

# Modus

Modus menyatakan kejadian  $X$  yang sering muncul, atau dengan kata lain modus adalah nilai  $x$  dimana  $f(x)$  maksimum.

$$x \text{ modus} \Rightarrow f(x) \text{ maksimum}$$

Modus juga merupakan nilai center of data untuk model data hasil kuantisasi

## Modus (*cont...*)

Diketahui data nilai test pemrograman yang diperoleh 10 orang mahasiswa adalah sebagai berikut:

**A B A A B B B C B C**

Nilai	Histogram	PDF
A	3	0.3
B	5	0.5
C	2	0.2

→ F(B) Maksimum  
Nilai=B adalah modus

# Varians

Varians adalah nilai sebaran data sekitar rata-rata

Model Kontinu: 
$$\text{var}(x) = \int_X (x - \bar{x})^2 f(x) dx$$

$f(x)$  menyatakan pdf dari  $x$ .

Model Diskrit: 
$$\text{var}(x) = \sum_{i=1}^N (x_i - \bar{x})^2 p(x_i)$$

# Varians (*cont...*)

Rumus varians yang juga banyak digunakan dalam perhitungan analisa data adalah:

$$\text{var}(x) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \quad \Rightarrow \quad \text{Untuk data yang tidak bias}$$

$$\text{var}(x) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad \Rightarrow \quad \text{Untuk data yang bias}$$

*Rumus ini digunakan untuk data-data sample.*

## Varians (*cont...*)

Nilai sebaran data standard dinyatakan dengan nilai standard deviasi, dimana standard deviasi adalah akar dua dari varians.

$$\textit{std}(x) = \sqrt{\text{var}(x)}$$

*Nilai standard deviasi ini yang digunakan sebagai ukuran standard sebaran data sekitar rata-rata.*

# Varians (*cont...*)

Diketahui data-data penjualan PC dalam 1 minggu terakhir sebagai berikut:

Hari	Penjualan
1	4
2	4
3	2
4	5
5	4
6	3
7	3

Berapa penjualan PC dalam minggu ini?

Untuk menyatakan nilai secara umum dapat digunakan nilai rata-rata dan standard deviasi sebagai berikut

Penjualan	Histogram	PDF	$x.p(x)$	$(x-m)^2.p(x)$
2	1	0.14	0.29	0.35
3	2	0.29	0.86	0.09
4	3	0.43	1.71	0.08
5	1	0.14	0.71	0.29
Total			3.57	0.82

$$\text{Rata-rata} = 3.57 \rightarrow \text{rata-rata} = 4$$

$$\text{Varians} = 0.82 \rightarrow \text{std} = 0.9$$

Penjualan berada dalam kisaran:

$3.57 - 0.9$  sampai dengan  $3.57 + 0.9$

Atau berada dalam kisaran  $2.67 - 4.47$

Dengan rata-rata penjualan 4

# Varians (*cont...*)

Diketahui data jumlah bemo yang lewat di depan PENS-ITS setiap 30 menit

Jam	Jumlah bemo
06.00 - 06.30	5
06.30 - 07.00	9
07.00 - 07.30	7
07.30 - 08.00	7
08.00 - 08.30	5
08.30 - 09.00	4
09.00 - 09.30	2
09.30 - 10.00	4
10.00 - 10.30	3
10.30 - 11.00	2
11.00 - 11.30	4
11.30 - 12.00	1
12.00 - 12.30	3
12.30 - 13.00	6
13.00 - 13.30	5
13.30 - 14.00	4
14.00 - 14.30	7
14.30 - 15.00	3
15.00 - 15.30	3
15.30 - 16.00	6

Dengan menggunakan nilai ekspektasi diperoleh:

$$\text{Rata-rata} = 4.5$$

$$\text{Varians} = 4.16$$

Maka standard deviasi = 2.04

Jumlah bemo yang lewat di depan PENS berada dalam kisaran  $(4.5-2.04)=2.46$  sampai dengan  $(4.5+2.04)=6.54$  dan rata-rata = 5