Bab 3 Kinerja Rata-rata dan Variabilitas

Pada kehidupan sehari-hari sering kali yang menjadi ukuran kinerja adalah nilai rata-rata dari beberapa hasil pengukuran. Pada saat itu kita mempercayai bahwa nilai rata-rata memiliki makna. Metriks sebagai variabel ukur juga sering dimaknai melalui nilai rata-ratanya. Misalnya dalam sistem komputer dikenal kecepatan rata-rata jaringan komputer, beban rata-rata. Nilai rata-rata (mean) sendiri merupakan besaran statistik yang diambil dari suatu pengukuran langsung. Besaran statistik yang lain adalah nilai tengah (median), terbesar (mode atau maksimum) dan nilai minimum. Sehingga metriks pun dikenal metriks berdasarkan nilai tengah, terbesar dan minimum.

Misalkan dari beban kerja suatu server dalam seminggu dinyatakan dengan jumlah layanan sebagai berikut: 40, 40, 70, 20, 100, 60, 30 layanan. Tentukan nilai mean, median dan modenya
Jawab:

$$mean \quad = \frac{40+40+70+20+100+60+30}{7} = \frac{360}{7} = 51.4 \quad layanan$$

untuk mencari median, nilai layanan perhari ini perlu diurutkan dari nilai terendah sampai tertinggi
jadi $median \quad = \quad nilai \quad tengah \quad X \quad = \quad 40$ layanan
dan mode-nya adalah nilai terbesar yaitu 100 layanan.
Secara umum persamaan mean, median dan mode dapat dituliskan dengan persamaan berikut:

$$mean = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3.1}$$

$$median = \left\{ \begin{matrix} x_{N/2} & untuk \quad N \quad genap \\ x_{(N-1)/2 \quad + \quad 1} & untuk \quad N \quad ganjil \end{matrix} \right\} \tag{3.2}$$

$$mode = max \quad (x) \tag{3.3}$$

3.1. Analisa nilai peluang

Suatu nilai ukur pada statistik merupakan nilai yang diperoleh dengan tanpa saling mempengaruhi atau dalam bahasa statistik disebut kejadian yang bebas (independen). Maksud dari bebas adalah setiap nilai ukur dapat terjadi tanpa dipengaruhi kejadian sebelumnya dan kejadiannya atau besar nilainya terjadi bebas tanpa adanya aturan (acak atau *random*).

Berdasarkan jumlah data dalam statistik suatu metriks dapat dianalisa berdasarkan peluang kejadiannya dan jumlah (kumulatif) peluang.
a, Peluang suatu nilai metriks dinyatakan dengan f(x) yang disebut fungsi kerapatan peluang (*Probability Density Function*)
b. Kumulatif suatu nilai metriks dinyatakan dengan F(x) yaitu fungsi distribusi kumulatif (*Cumulative Distribution Function*)

Sebagai contoh jika kinerja diasumsikan seperti nilai pada dadu yang masing-masing nilai memiliki peluang kemunculan 1/6 maka dapat dikatakan peluang kejadian nilai dadu =1 akan sama dengan nilai dadu =2 atau ditulis f(1) = f(2).
Sedangkan nilai kemunculan nilai dadu 1 sampai 3 adalah f(1)+f(2)+f(3) = F(3). Contoh ini secara statistik ditulis sebagai berikut:

$$P\,(1 \le x \le 3) = \sum_{x=1}^{3} f(x) = F\,(3)$$

Jika kita ingin mencari peluang kejadian nilai dadu yang muncul 2 sampai 6 maka dapat dihitung sebagai berikut:

$$P\left(2 \le x \le 6\right) = \sum_{x=2}^{6} f_x = F\left(6\right) - F\,(2) = 1 - \frac{2}{6} = \frac{4}{6}$$

Ini adalah contoh dari kejadian yang nilai peluang tiap kejadiannya sama yaitu setiap kejadian pada dadu nilai peluangnya 1/6. Dilihat dari nilainya ini merupakan peluang untuk kejadian diskrit.

Dalam bentuk kontinu persamaan (3.2) dapat digeneralisasi menjadi:

$$P\left(x_1 \le x \le x_2\right) = \lim_{\triangle x -> 0} \sum_{x_1}^{x_2} f(x) = \int_{x_1}^{x_2} f(x)dx = F\left(x_2\right) - F\left(x_1\right) \tag{3.4}$$

Bagaimana jika peluangnya untuk tiap kejadian tidak sama, maka perhitungannya pada dasarnya sama saja. Misal suatu pengukuran pengiriman data untuk sejumlah kilo byte data adalah sebagai berikut:

Jumlah Data yang dikirim (kb)	Data yang berhasil terkirim dari 10x uji coba
10	4
20	6
30	0
40	10

a. Tentukan peluang dari masing-masing pengiriman data
b. Berapakah peluang pengiriman data berhasil untuk 20 - 30 kb
c. Berapakah peluang pengiriman data berhasil untuk 0 - 30 kb

Jawab:
Peluang dan kumulatif peluangnya dituliskan pada kolom ke -3 dan ke-4 sebagai berikut

Jumlah Data yang dikirim (kb)	**Data yang berhasil terkirim dari 10x uji coba**	**Peluang berhasil f(x)**	**Kumulatif Peluang berhasil F(x)**
a	b	$c = \frac{b}{10} . \frac{1}{4}$	$\sum c$
10	4	0,1	0,1
20	6	0,15	0,25
30	0	0	0,25
40	10	0,25	0,5

a. Karena percobaan diulang 10 kali pengujian dan masing-masing grup data yang diuji memiliki peluang yang sama yaitu 1/4 maka peluang berhasil dinyatakan dengan nilai c di kolom ke tiga
b. Peluang pengiriman data berhasil untuk 20-30 kb adalah: F(30) - F(20) = 0
c. Peluang pengiriman data berhasil untuk 0 - 30 kb : F(30) = 0,25

3.2. Nilai rata-rata

Persamaan median pada persamaan 3.1 adalah persamaan untuk nilai rata-rata aritmetik (*Arithmetic mean*). Jenis nilai rata-rata ini memiliki sifat bahwa metriks yang dirata-ratakan adalah pengukuran langsung bukan hasil perhitungan dan memiliki orde angka yang sama. Pengukuran langsung berarti metriks bukan hasil perhitungan dari dua variabel atau lebih, sedangkan orde yang sama artinya jika nilai ukur berorde puluhan(10^1) maka tidak ada satu atau lebih nilai yang memiliki dua tingkat order yaitu ribuan (10^3).

Untuk kasus metriks yang merupakan nilai perhitungan dikenal nilai rata-rata harmonik (*Harmonic mean*), $\bar{x}_H$

$$\bar{x}_H = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}} \tag{3.5}$$

sedangkan

untuk kasus metriks yang berbeda orde nikenal nilai rata-rata geometrik (*Geometric Mean)*, $\bar{x}_G$

$$\bar{x}_G = \left(\prod_{i=1}^{N} x_i \right)^{1/N} \tag{3.6}$$

Nilai rata-rata geometrik sering digunakan untuk melakukan perbandingan urutan terbaik (*Rangking)* pada uji benchmarking.

Contoh nilai rata-rata harmonik adalah saat nilai rata-rata digunakan untuk mengukur MFLOPS, MIPS, Kbps. Karena nilai ini merupakan perhitungan dari besaran ukur dibagi terhadap waktu.

JIka telah diperoleh nilai rata-rata aritmetik dari beberapa pengukuran, maka tidak boleh lagi dicari nilai rata-ratanya dari pengukuran tersebut menggunakan nilai rata-rata aritmetik juga. Misal nilai rata-rata kecepatan download dalam satu minggu adalah untuk masing-masing hari sebagai berikut 40, 85, 30, 50, 80, 20, 32 kbps maka tidak boleh dicari rata-rata perhari dengan menjumlahkan angka tersebut dan membaginya dengan tujuh. Tetapi untuk mencari rata-rata per harinya harus menggunakan nilai rata-rata harmonik karena nilai masing-masingnya per hari adalah sudah menjadi nilai yang dihitung.

Contoh : Carilah nilai rata-rata harmonik dari MFLOPS (Million Floating Point Operastion per Second) pada tabel 3.1.

Tabel 3.1. Waktu Eksekusi dan Jumlah intruksi floating point pada 5 kali pengukuran

Pengukuran (i)	***T_i(s)***	***F (10^9 FLOP)***	***M_i (MFLOPS)***
1	321	130	405
2	436	160	367
3	284	115	405
4	601	252	419
5	482	187	388
Waktu Total	*2124*	*844*	
$\bar{T}_A$	425		

$$\overline{MFLOPS}_H = \frac{N}{\sum_{i=1}^{N} \frac{1}{MFLOPS_i}} = \frac{5}{\frac{1}{405} + \frac{1}{367} + \frac{1}{405} + \frac{1}{419} + \frac{1}{388}} = 396 \text{ MFLOPS}$$

Contoh: Pada tabel 2 terdapat hasil pengukuran waktu eksekusi pada tiga sistem komputer menggunakan 5 program penguji. Pada tabel terlihat ada satu program uji yang menghasilkan waktu eksekusi sangat lama dan memiliki orde yang sangat besar dibandingkan dengan orde nilai yang lain.

Tabel 3.2. Pengukuran waktu eksekusi terhadap 3 sistem komputer

Program	S_1	S_2	S_3
1	417	244	134
2	83	70	70
3	66	153	135
4	39.449	33.527	66.000
5	772	368	369

Carilah a) waktu eksekusi rata-rata dari masing-masing sistem komputer ; b) tentukan rangking terbaik yaitu yang memiliki nilai rata-rata terkecil
Jawab:
a. Waktu Ekseskusi geometrik dari masing-masing sistem komputer adalah

untuk sistem 1: $\left(417x83x66x39449x772\right)^{1/5}$ = 587
untuk sistem 2: $(244x70x153x33527x368)^{1/5}$ = 503
untuk sistem 3: $(134x70x135x66000x369)^{1/5}$ = 499

b. rangking terbaik adalah sistem komputer ke-3

	S_1	S_2	S_3
waktu eksekusi rata-rata geometrik	587	503	499
Rangking	3	2	1

3.3. Nilai Penyebaran

Cara lain untuk merepresentasikan metriks diperoleh dari nilai penyebaran suatu data ukur kinerja. Nilai penyebara bisa berupa rentang, standar deviasi, varian. Contoh bahasa sehari-hari dari nilai penyebaran ini misalnya beberapa alat dikatakan memiliki kinerja yang baik jika beban kerjanya memiliki rentang yang besar, beberapa alat dikatakan baik jika nilai kesalahannya berada pada standar deviasi yang kecil, suatu kualitas sistem jaringan (QoS) dikatakan konsisten jika memiliki varian delay (jitter) yang kecil.

Nilai standar deviasi dan varian dinyatakan dengan persamaan berikut:

$$\sigma_x = \sqrt{var(x)} = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N-1}} \tag{3.7}$$

Beberapa kinerja sering menuliskan hasil ukur menggunakan format $x_{median} \pm \sigma_x$ misalnya kinerja dari bandwidth pengiriman data adalah 50 ± 2.5 kbps yang artinya rentang dari bandwidth pengiriman data adalah 47.5 - 52.5 kbps. Terkadang pada beberapa perhitungan kinerja nilai median sama dengan nilai rata-ratanya yaitu pada saat sistem memiliki data berbentuk distribusi

Gaussian (normal) sehingga ditulis $\bar{x} \pm \sigma_x$. Suatu sistem yang memiliki nilai standar deviasi terlalu besar berarti sistem ini memiliki kinerja dengan retang yang besar di luar nilai rata-ratanya, bagi sistem yang kualitas rentang yang besar berarti sistem yang kualitasnya tidak konsisten, tetapi bagi sistem beban, rentang yang besar berarti sistem yang tangguh (robust) menghadapi beban yang diberikan. Untuk itu ditambahkan metriks baru berupa pembagian antara nilai standar deviasi dengan nilai rata-rata yang disebut sebagai koefisien varian (Coefficient of Variance),

$$\text{COV} = \frac{\sigma_x}{\bar{x}} \tag{3.8}$$

koefiesien varian menunjukkan besarnya penyebaran atau rentang data, jika COV > 0.1 berarti variasi dari metriks tersebut sangat besar sedangkan untuk $COV \leq 0.1$ berarti metriks rentangnya kecil dan memiliki konsistensi dan kemiripan.

Pada sistem kualitas manufaktur yang disebut manajemen six sigma, nilai standar deviasinya adalah 1/6 dari nilai meannya berarti kualitasnya dijaga semirip mungkin dengan variasi yang kecil.

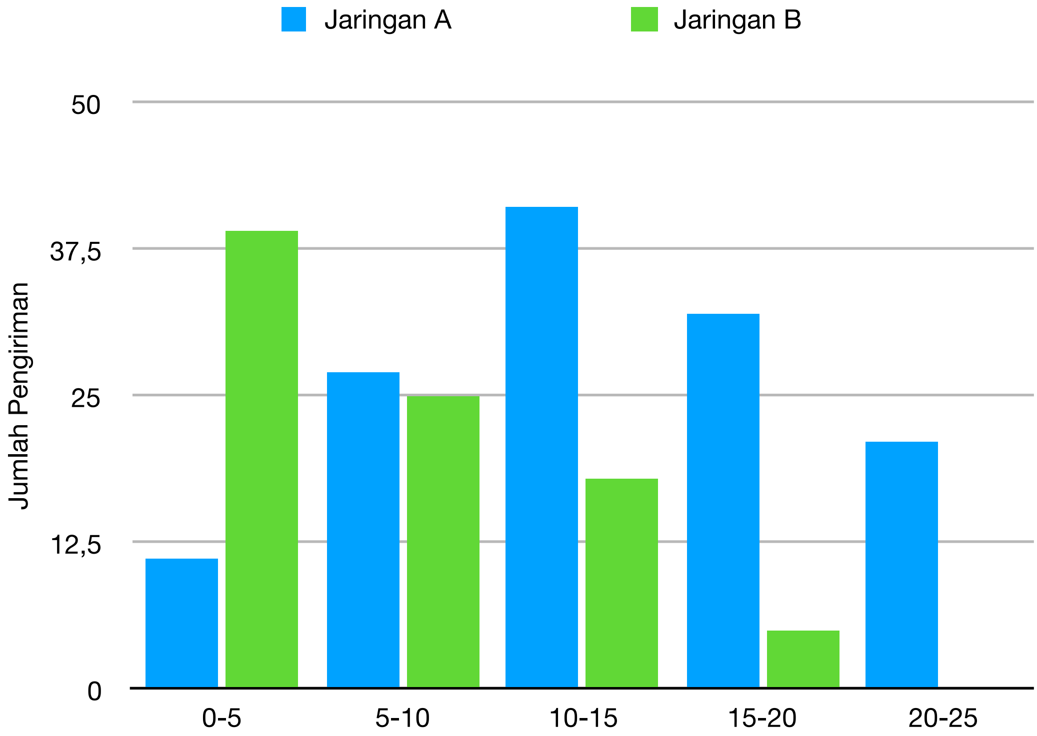
3.4. Histogram dan Percentile

Untuk menyatakan data ukur bisa dilakukan dalam bentuk interval data yang memilik rentang yang sama yang kemudian metriksnya dinyatakan dalam bentuk histogram. Pembagian selang interval bisa berdasarkan intuisi maupun perhitungan empiris dari beberapa penelitian. Berikut contoh jumlah pengiriman data pada dua jaringan, A dan B.

Tabel 3.3 Jumlah pengiriman data berdasarkan rentang

Ukuran Pesan (kbyte)	Jaringan A	Jaringan B
$0 < x_i \leq 5$	11	39
$5 < x_i \leq 10$	27	25
$10 < x_i \leq 15$	41	18
$15 < x_i \leq 20$	32	5
$20 < x_i \leq 25$	21	19
$25 < x_i \leq 30$	12	42
$30 < x_i \leq 35$	4	0

Tabel ini dapat dinyatakan dengan bentuk histogram (Gambar 3.1)



Gambar 3.1. Histogram data dari tabel 3.1

Sebenarnya secara teori untuk selang interval yang sangat kecil maka histogram akan berubah menjadi fungsi kerapatan peluang, f(x).

Percentile adalah metriks khusus yang nilainya berdasarkan persentase dari total kumulatif data (*Data Qumulative), Q.* Karena perhitungan distribusi kumulatif tidak dimulai dari nol maka perhitungan percentile juga tidak dari nol tapi dimulai dari 5%-95% dari total kumulatif data, Q
Beberapa istilah :

a. Rentang Percentile : 5% sampai 95% dari Q,
b. Quartile yaitu peningkatan percentile untuk tiap 0,25 atau 25%, dengan nilai maksimal Q3. MIsal Q_2 berarti 50% Q
c. Decile yaitu peningkatan percentile untuk tiap 0,1 atau 10%, dengan nilai maksimal D9. Misal D_2 = 20% Q
d. Semi intequartile untuk $(Q_3\text{-}Q_1)/2$

Contoh untuk jaringan A pada tabel 3.3 dapat dihitung kumulatifnya, seperti yang diperlihatkan pada tabel 3.4 :

Tabel 3.4. Fungsi distribusi kumulatif untuk Jaringan A

Ukuran Pesan (kbyte)	**Jumlah yang dikirimkan**	**Kumulatif dari jumlah yang dikirimkan**
$0 < x_i \leq 5$	11	11
$5 < x_i \leq 10$	27	38
$10 < x_i \leq 15$	41	79
$15 < x_i \leq 20$	32	111
$20 < x_i \leq 25$	21	132
$25 < x_i \leq 30$	12	144
$30 < x_i \leq 35$	4	148
	N=148	

Pertanyaaan:
Jika batas *fault tolerance* dinyatakan dengan quartile ke-1, rentang beban normal dinyatakan dengan percentile ke-6 serta rentang beban optimum dinyatakan dengan semi interquartile. Tentukan :
a) batas *fault tolerance*
b) titik optimum
c) rentang beban kerja

Jawab:

a. Batas *fault tolerance* = Q_1

$\frac{N}{4} = \frac{148}{4} = 37$ berarti berada di rentang $5 < x_i \leq 10$ atau tepatnya dapat dihitung dengan interpolasi

L= Batas bawah interval = 5; CF= Cumulative frekuensi di interval bawah = 11
F = frekuensi 27, i = lebar interval = 5. Maka

$$Q_1 = L + \frac{\frac{N}{4} - CF}{f} x \quad i$$

$$= 5 + \frac{37 - 11}{27} x \quad 5 = 9{,}8$$

b. Nilai optimumnya = D_6

Untuk D_6 berarti berada pada kumulatif : 0,6 x 148 = 88,8. Berarti untuk D_6 berada di rentang 15 < x_i ≤ 20 atau tepatnya dapat dihitung dengan interpolasi:

L = 15; CF = 79; f= 32

Maka:

$$D_6 = L + \frac{88{,}8 - CF}{f} x \quad i$$

$$= 15 + \frac{88{,}8 - 79}{32} x \quad 5 = 16{,}5$$

c. Rentang beban kerja = semi interquartile = (Q_3 - Q_1)/2

Untuk Q3 :

$\frac{3N}{4} = \frac{3x148}{4} = 111$ berarti berada pada di rentang 15 < x_i ≤ 20 atau tepatnya dapat adalah 20 karena kumulatif 111 adalah ujung dari interval 15 < x_i ≤ 20

Maka nilai semi interquartilenya adalah :

$$\frac{Q_3 - Q_1}{2} = \frac{20 - 9{,}8}{2} = 5{,}1$$