Measure of Variation dalam Statistika

...

Video #8 dari Seri Video Belajar
Statistika Dasar
(Statistika Deskriptif)

https://www.youtube.com/IndonesiaBelajarKomputer

Apa itu Measure of Variation?

Measure of Variation dapat didefinisikan sebagai suatu pengukuran nilai yang dapat digunakan untuk merepresentasikan **keberagaman** atau **sebaran data**.

Range

Variance

Standard Deviation

Range (Jangkauan)

Range dari suatu dataset merupakan hasil perhitungan **selisih** antara **nilai tertinggi** dengan **nilai terrendah** pada dataset tersebut.

Pengukuran nilai keberagaman dengan menggunakan range memiliki kelemahan di mana **hanya menyertakan dua nilai saja** dalam proses pengukuran.

Range: contoh

Range = 47 - 38 = 10

Salary	41	38	39	45	47	41	44	41	37	42

Median = 41
Mean = 41.5

Salary	40	23	41	50	49	32	41	29	52	58

Range = 58 - 23 = 35

Variance (Variansi)

Variance dari suatu dataset merupakan hasil perhitungan **rerata simpangan tiap entri data** pada dataset **terhadap nilai mean** dari dataset tersebut.

Population Variance

$$\sigma^2 = \frac{\Sigma(x - \mu)^2}{N}$$

Sample Variance

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}$$

Variance: contoh

x	41	38	39	45	47	41	44	41	37	42	$\Sigma(x-\mu)$

$\mu = 415/10 = 41.5$

$x - \mu$	- 0.5	- 3.5	- 2.5	3.5	5.5	- 0.5	2.5	- 0.5	- 4.5	0.5	**0**
$(x-\mu)^2$	0.25	12.25	6.25	12.25	30.25	0.25	6.25	0.25	20.25	0.25	**88.5**

$$\sigma^2 = \frac{88.5}{10} \approx 8.9$$

Standard Deviation (Simpangan Baku)

Kelemahan utama dari Variance adalah nilai yang dihasilkan tidak lagi memiliki satuan yang sama dengan entri data. Kelemahan ini dapat diatasi dengan Standard Deviation.

Population Standard Deviation

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(x - \mu)^2}{N}}$$

Sample Standard Deviation

$$s = \sqrt{s^2} = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

Standard Deviation: contoh

x	41	38	39	45	47	41	44	41	37	42	$\Sigma(x-\mu)$
$x-\mu$	- 0.5	- 3.5	- 2.5	3.5	5.5	- 0.5	2.5	- 0.5	- 4.5	0.5	**0**

$(x-\mu)^2$	0.25	12.25	6.25	12.25	30.25	0.25	6.25	0.25	20.25	0.25	**88.5**

$$\sigma^2 = \frac{88.5}{10} \approx 8.9$$

$$\sigma = \sqrt{\frac{88.5}{10}} \approx 3.0$$

Review Notasi

Population	Sample
σ^2	s^2
σ	s
μ	$\bar{x}$
N	n
$x - \mu$	$x - \bar{x}$
$\Sigma(x - \mu)^2$	$\Sigma(x - \bar{x})^2$

Variance

Standard Deviation

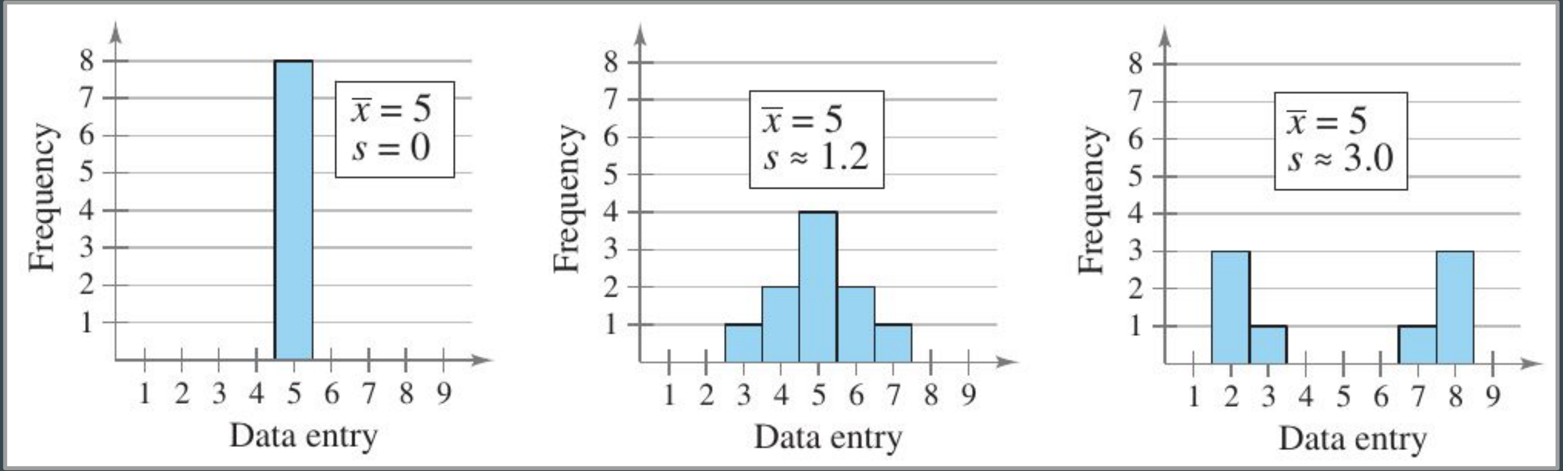
Mean

Jumlah Entri

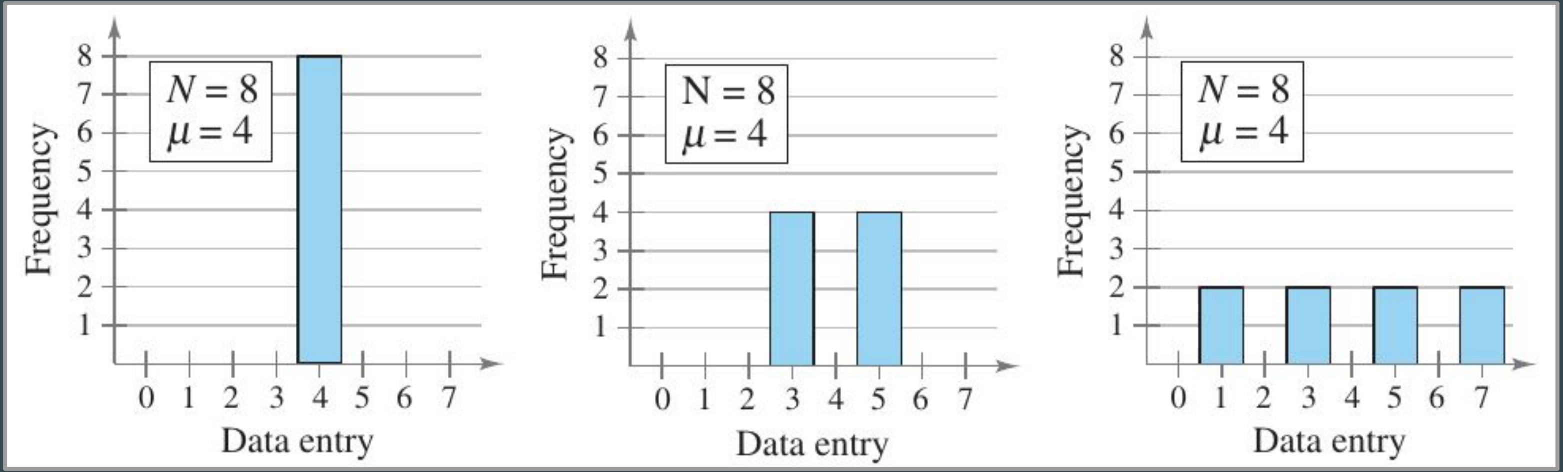
Deviation

Sum of Square

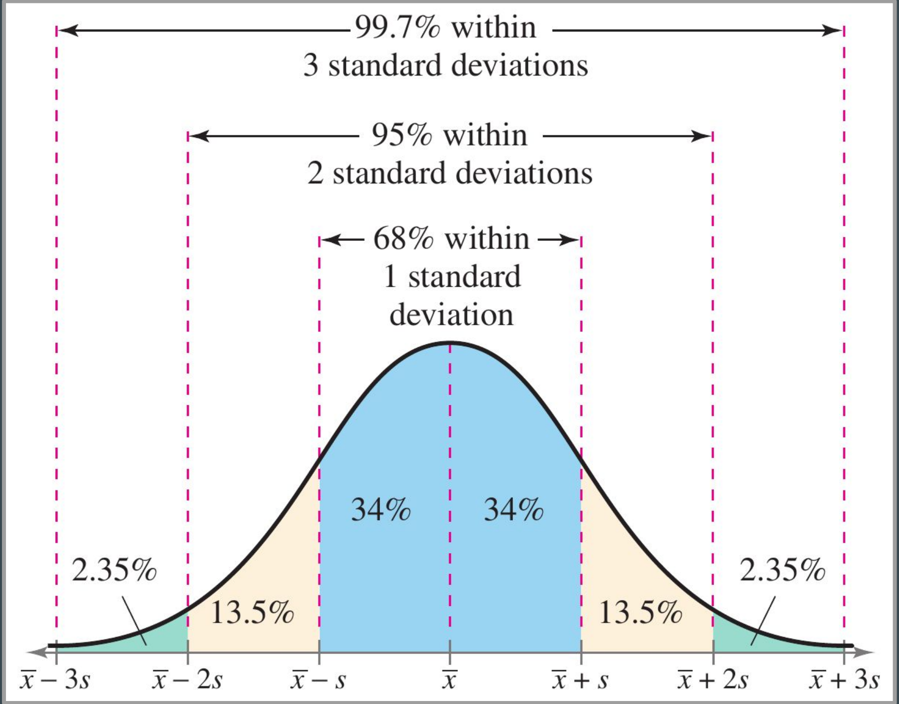
Standard Deviation dan Bentuk Distribusi [1/2]



Standard Deviation dan Bentuk Distribusi [2/2]

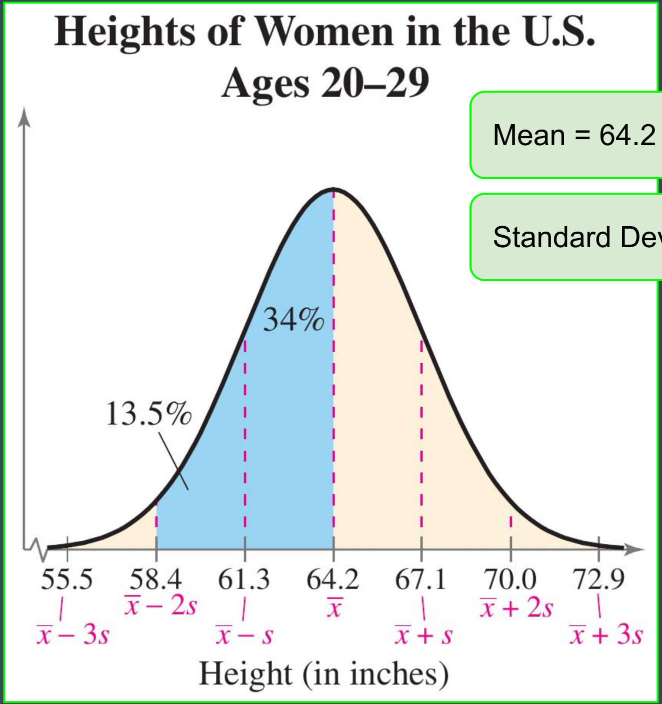


Empirical Rule



- Data yang kita temui di lapangan, umumnya memiliki bentuk distribusi yang mendekati bentuk distribusi simetris (bell shaped).
- Empirical Rule dapat diterapkan pada bentuk distribusi simetris (bell shaped).

Empirical Rule: contoh



Mean = 64.2

Standard Deviation = 2.9

Berapa estimasi persentase wanita US (usia 20-29 tahun) dengan tinggi badan antara 58.4-64.2 inch?

Jawaban:
13.5% + 34% = 47.5%

Chebychev's Theorem

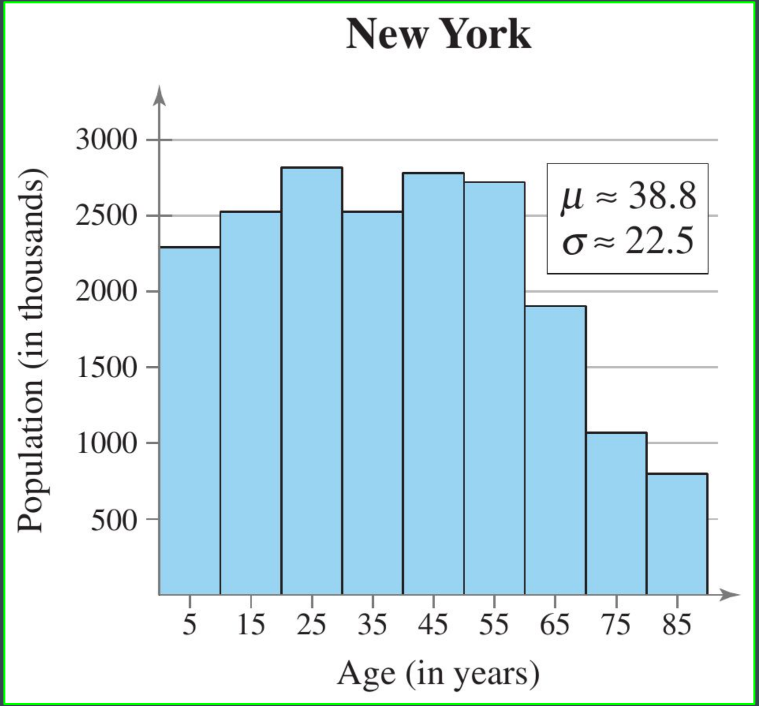
Chebychev's Theorem: proporsi minimum dari dataset yang berada pada K standard deviation diformulasikan dengan $1 - \frac{1}{k^2}$

$k = 2$ → $1 - \frac{1}{2^2} = \frac{3}{4}$ → 75%

$k = 3$ → $1 - \frac{1}{3^2} = \frac{8}{9}$ → 88.9%

- Empirical Rule hanya berlaku untuk symetric distribution (bell shaped distibution).
- Sedangkan Chebychev's Theorem dapat diterapkan untuk semua bentuk distribusi.

Chebychev's Theorem: contoh



Berapakah proporsi data yang berada dalam rentang dua standard deviation dari mean? Sertakan interpretasi!

Kalkulasi:

- $38.8 - 2(22.5) = -6.2$
- $38.8 + 2(22.5) = 83.8$

Interpretasi:

Setidaknya 75% dari populasi penduduk di New York berusia antara 0 sampai 83.8 tahun

Standard Deviation for Grouped Data (Frequency Distribution)

$$\text{Sample standard deviation} = s = \sqrt{\frac{\Sigma(x - \bar{x})^2 f}{n - 1}}$$

$$n = \Sigma f$$

Standard Deviation for Grouped Data: contoh

Number of children in 50 households				
1	3	1	1	1
1	2	2	1	0
1	1	0	0	0
1	5	0	3	6
3	0	3	1	1
1	1	6	0	1
3	6	6	1	2
2	3	0	1	1
4	1	1	2	2
0	3	0	2	4

x	f	xf
0	10	0
1	19	19
2	7	14
3	7	21
4	2	8
5	1	5
6	4	24
	$\Sigma = 50$	$\Sigma = 91$

$x - \bar{x}$	$(x - \bar{x})^2$	$(x - \bar{x})^2 f$
−1.82	3.3124	33.1240
−0.82	0.6724	12.7756
0.18	0.0324	0.2268
1.18	1.3924	9.7468
2.18	4.7524	9.5048
3.18	10.1124	10.1124
4.18	17.4724	69.8896
		$\Sigma = 145.38$

$$\bar{x} = \frac{\Sigma xf}{n} = \frac{91}{50} = 1.82 \approx 1.8$$

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2 f}{n - 1}} = \sqrt{\frac{145.38}{49}} \approx 1.7$$

Coefficient of Variation

- **Standard Deviation** dapat digunakan untuk membandingkan keberagaman/sebaran data antar dataset yang memiliki **satuan pengukuran yang sama** dengan **nilai mean yang mirip**.
- Sedangkan untuk dataset yang memiliki **satuan pengukuran yang berbeda** atau **nilai mean yang jauh berbeda**, maka kita mesti menggunakan **Coefficient of Variation.**

Population CV

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Sample CV

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Coefficient of Variation: contoh

Heights and Weights of a Basketball Team

Heights	Weights
72	180
74	168
68	225
76	201
74	189
69	192
72	197
79	162
70	174
69	171
77	185
73	210
inches	pounds

Heights:

$\mu \approx 72.8$ inches

$\sigma \approx 3.3$ inches

$$\begin{aligned} CV_{\text{height}} &= \frac{\sigma}{\mu} \cdot 100\% \\ &= \frac{3.3}{72.8} \cdot 100\% \\ &\approx 4.5\%. \end{aligned}$$

Weights:

$\mu \approx 187.8$ pounds

$\sigma \approx 17.7$ pounds

$$\begin{aligned} CV_{\text{weight}} &= \frac{\sigma}{\mu} \cdot 100\% \\ &= \frac{17.7}{187.8} \cdot 100\% \\ &\approx 9.4\%. \end{aligned}$$

Indonesia Belajar

https://www.youtube.com/IndonesiaBelajarKomputer

Banyak Belajar Biar Bisa Bantu Banyak Orang