



Statistika

Random Variables

Discrete Random Variables

Continuous Random Variables

Pengertian

- Random variable (variabel acak)
 - suatu fungsi yang didefinisikan pada sample space
- Jenis
 - Discrete random variables
 - Continuous random variables
- Contoh
 - jumlah hari hujan selama 1 tahun → diskrit
 - jumlah (volume) hujan selama 1 tahun → kontinu

Random Variables

■ Notasi

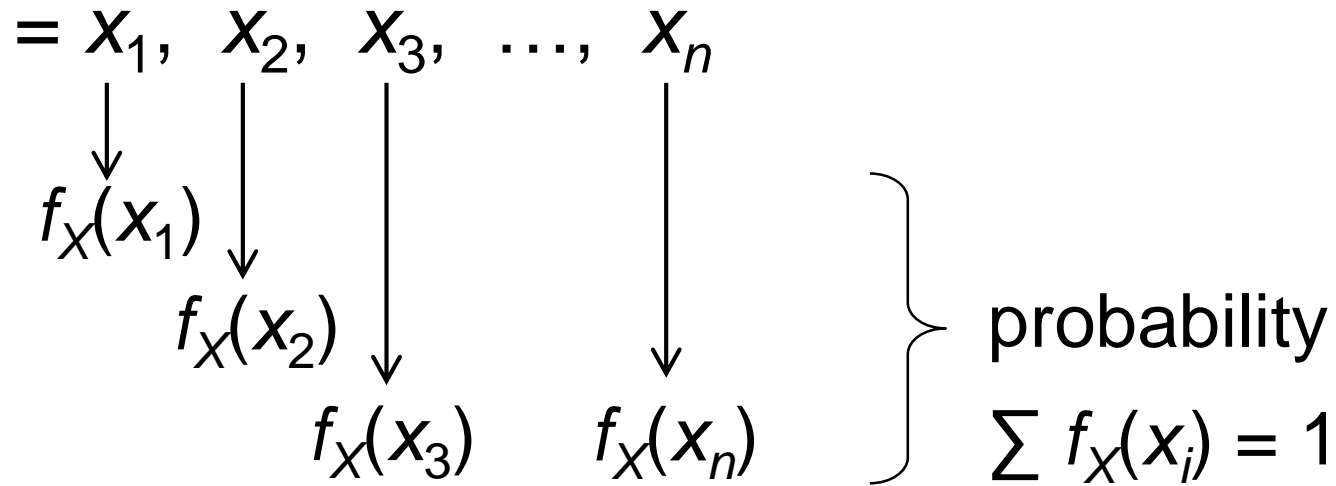
- $X \rightarrow$ variabel random
- $x \rightarrow$ nilai variabel random

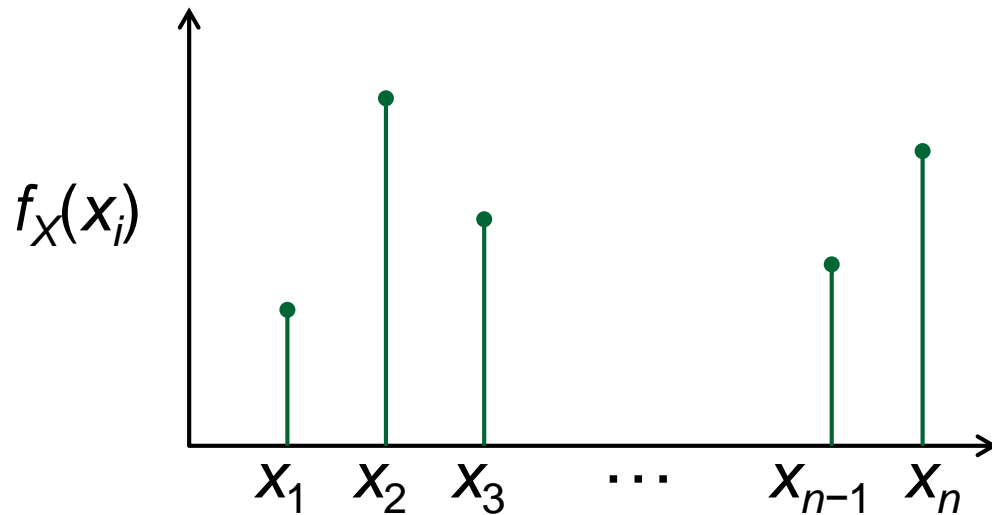
■ Fungsi

- Suatu fungsi variabel random adalah variabel random pula
- Jika X adalah variabel random, maka $Z = g(X)$ adalah juga variabel random

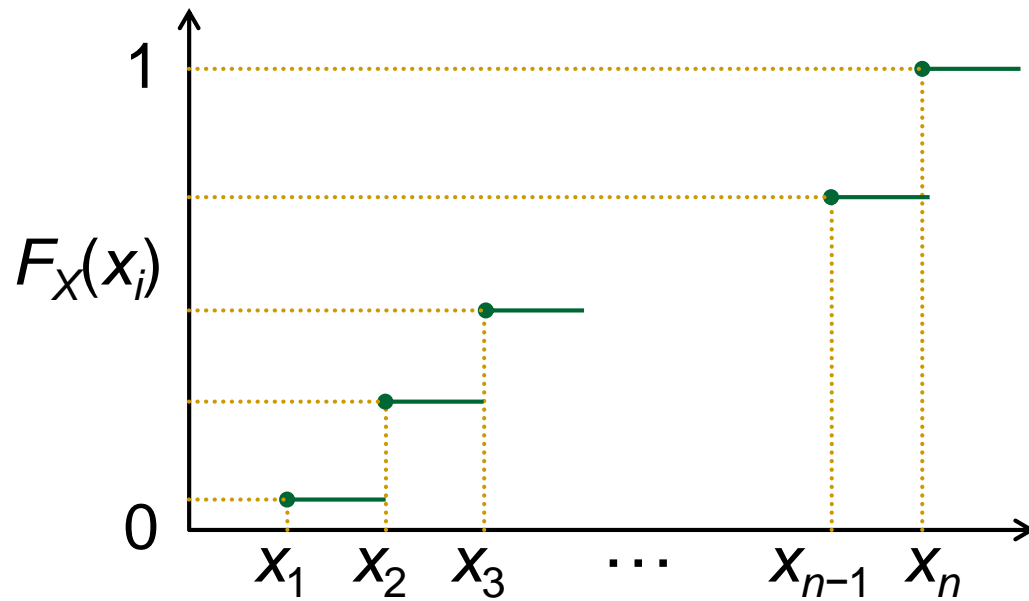
Univariate Probability Distributions

$X =$ discrete random variables





a discrete probability distribution

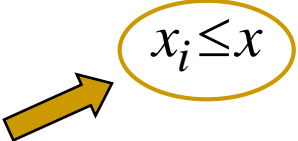


a discrete cumulative probability distribution



probability
 $x \leq x_i$

- Cumulative probability distribution
suatu variabel random X untuk $X = x$

$$F_X(x) = \sum_{x_i \leq x} f_X(x_i)$$


- Probability distribution
suatu variabel random X untuk $X = x$

$$f_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$

Univariate random variables

Jika X dapat bernilai x_1, x_2, \dots, x_n yang masing-masing memiliki probability $f_X(x_1), f_X(x_2), \dots, f_X(x_n)$ dan $\sum f_X(x_i) = 1$, maka X adalah variabel random diskrit.

■ Frequency

$$f_{x_i} = F_{x_i} - F_{x_{i-1}} \quad \rightarrow \text{frekuensi relatif}$$

$$F_{x_i} = \sum_{j=1}^i f_{x_j} \quad \rightarrow \text{frekuensi relatif kumulatif}$$

■ Probability

$$f_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$

$$F_X(x) = \sum_{x_i \leq x} f_X(x_i)$$

Continuous Random Variables

■ Probability

$$\text{prob}(A) = \frac{n_i}{n} = f_{x_i}$$

n_i = jumlah data di klas ke- i

n = jumlah seluruh data

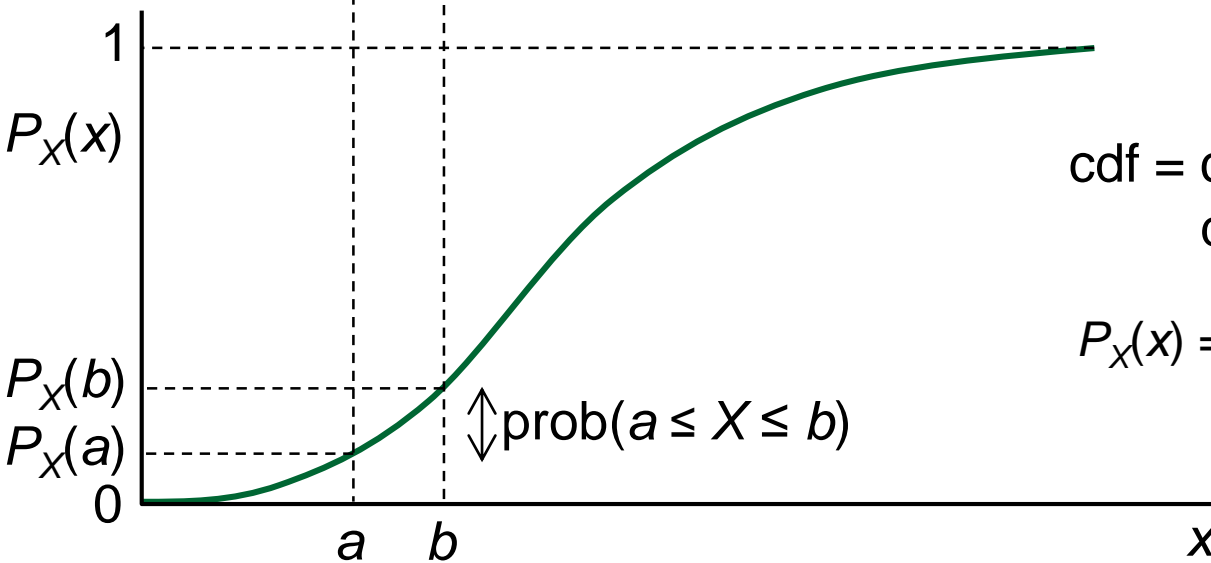
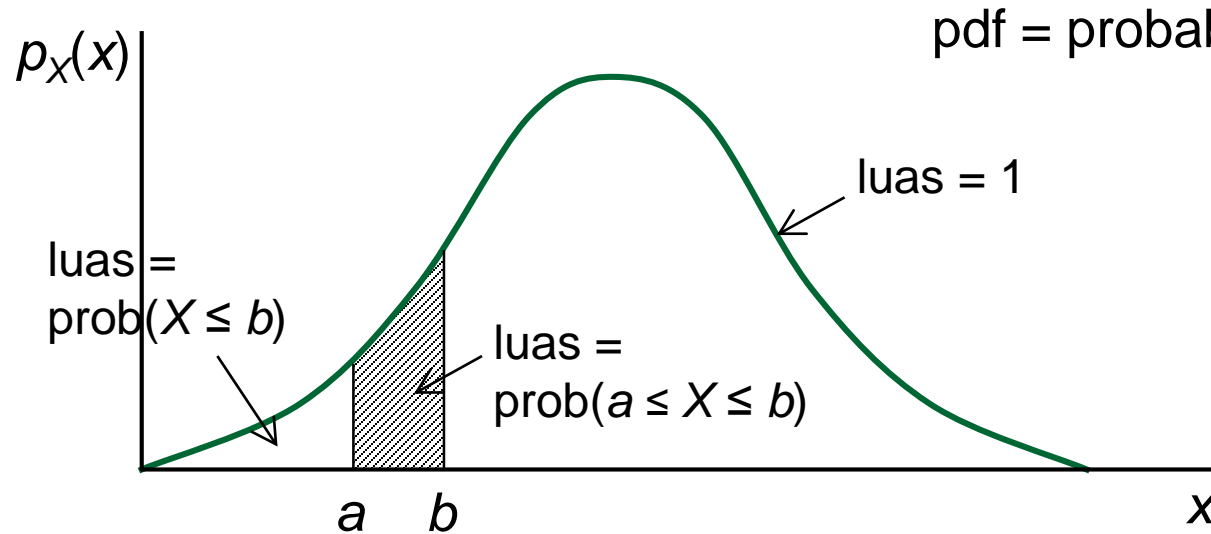
Dengan demikian, f_{x_i} dapat dipandang sebagai nilai estimasi probability.

$f_{x_i} \rightarrow$ estimasi $\text{prob}(A)$

histogram frekuensi \rightarrow pendekatan distribusi probability

frekuensi kumulatif \rightarrow pendekatan distribusi probability kumulatif

continuous random
variable treated as
though it were
discrete



cdf = cumulative probability distribution function

$P_X(x) = \text{prob}(X \leq x)$

$p_X(x)$ = probability density function of a continuous random variable

$P_X(x)$ = cumulative probability distribution function

$$P_X(x) = \text{prob}(X \leq x)$$

$$dP_X(x) = p_X(x) dx$$

$$P_X(x) = \int_{-\infty}^x p_X(t) dt$$

Beberapa sifat probability

$$1) p_X(x) \geq 0, \forall x$$

$$2) \int_{-\infty}^{+\infty} p_X(x) dx = 1$$

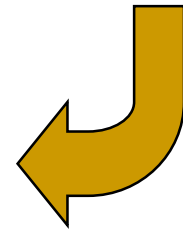
$$3) P_X(-\infty) = 0$$

$$4) P_X(+\infty) = 1$$

$$5) \text{prob}(a \leq X \leq b) = \int_a^b p_X(t) dt \\ = P_X(b) - P_X(a)$$

$$6) \text{prob}(X = c) = \int_c^c p_X(t) dt \\ = P_X(c) - P_X(c) = 0$$

$$\text{prob}(a \leq X \leq b) = \text{prob}(a < X \leq b) = \\ \text{prob}(a \leq X < b) = \text{prob}(a < X < b)$$



Kala Ulang

$$\text{prob}(X \geq a) = \text{prob}(X > a)$$

- Jadi dalam definisi kala ulang
 - a. suatu kejadian yang menyamai atau melampaui suatu nilai tertentu
 - b. suatu kejadian yang melampaui suatu nilai tertentu



Kedua definisi, **a** dan **b**, adalah sama mengingat probability suatu kejadian (event) menyamai suatu nilai tertentu adalah nol

Contoh

ST Contoh univariate probability distribution

Bivariate Distributions

(1)

- Apabila diinginkan untuk mengetahui perilaku dua atau lebih random variables → joint probabilities
- Joint probability density function

$$p_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} P_{X,Y}(x, y)$$

$$P_{X,Y}(x, y) = \text{prob}(X < x \text{ and } Y < y)$$
$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{X,Y}(s, t) dt ds$$

Bivariate Distributions

(2)

- Beberapa persamaan

$p_{X,Y}(x, \infty)$ is a cumulative univariate probability function of X only

$P_{X,Y}(\infty, y)$ is a cumulative univariate probability function of Y only

$$p_{X,Y}(x, y) \geq 0$$

$$P_{X,Y}(\infty, \infty) = 1$$

$$P_{X,Y}(-\infty, y) = P_{X,Y}(x, -\infty) = 0$$

Marginal Distributions

(1)

- Two random variables, X and Y
 - Ingin diketahui perilaku X tanpa mempertimbangkan nilai variabel Y
 - Marginal density:

$$p_{X,Y}(x, y) \rightarrow p_X(x)$$

$$p_X(x) = \int_{-\infty}^{+\infty} p_{X,Y}(x, t) dt$$

Marginal Distributions

(2)

- Two random variables, X and Y
 - Cumulative marginal distribution:

$$P_{X,Y}(x, y) \rightarrow P_X(x)$$

$$\begin{aligned} P_X(x) &= P_X(x, \infty) = \text{prob}(X \leq x \text{ dan } Y \leq \infty) \\ &= \text{prob}(X \leq x) \end{aligned}$$

$$= \int_{-\infty}^x \int_{-\infty}^{+\infty} p_{X,Y}(s, t) dt ds$$

$$= \int_{-\infty}^x p_X(s) ds$$

Marginal Distributions

(3)

- Untuk variabel Y

- Marginal density

$$p_{X,Y}(x, y) \rightarrow p_Y(y)$$

$$p_Y(y) = \int_{-\infty}^{+\infty} p_{X,Y}(s, y) ds$$

- Cumulative marginal distribution:

$$P_{Y \leq y} = P_Y(\infty, y) = \text{prob}(X \leq \infty \text{ dan } Y \leq y) = \text{prob}(Y \leq y)$$

$$= \int_{-\infty}^y p_Y(t) dt$$

Conditional Distributions

- Two random variables, X and Y
 - Ingin diketahui perilaku X yang bergantung pada Y
 - Distribusi X jika $Y = y_0$
 - Distribusi Y jika $x_1 \leq X \leq x_2$

$$p_{X,Y}(x_i | y \text{ di dalam } S) = \frac{\int_S p_{X,Y}(x, t) dt}{\int_S p_Y(t) dt}$$

$$\text{prob}(x \text{ di dalam } R | y \text{ di dalam } S) = \int_R p_{X|Y}(x | y \text{ di dalam } S) dx$$

$$p_{X|Y}(x | y = y_0) = \frac{p_{X,Y}(x, y_0)}{p_Y(y_0)} \text{ yang lebih sering dituliskan}$$

$$p_{X|Y}(x | y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

Independence

- Two random variables, X dan Y

- Kedua variabel independence jika

$p_{X|Y}(x|y)$ bukan fungsi y

$$p_{X|Y}(x|y) = p_X(x)$$

- Joint probabilities: perkalian density marginal kedua variabel adalah

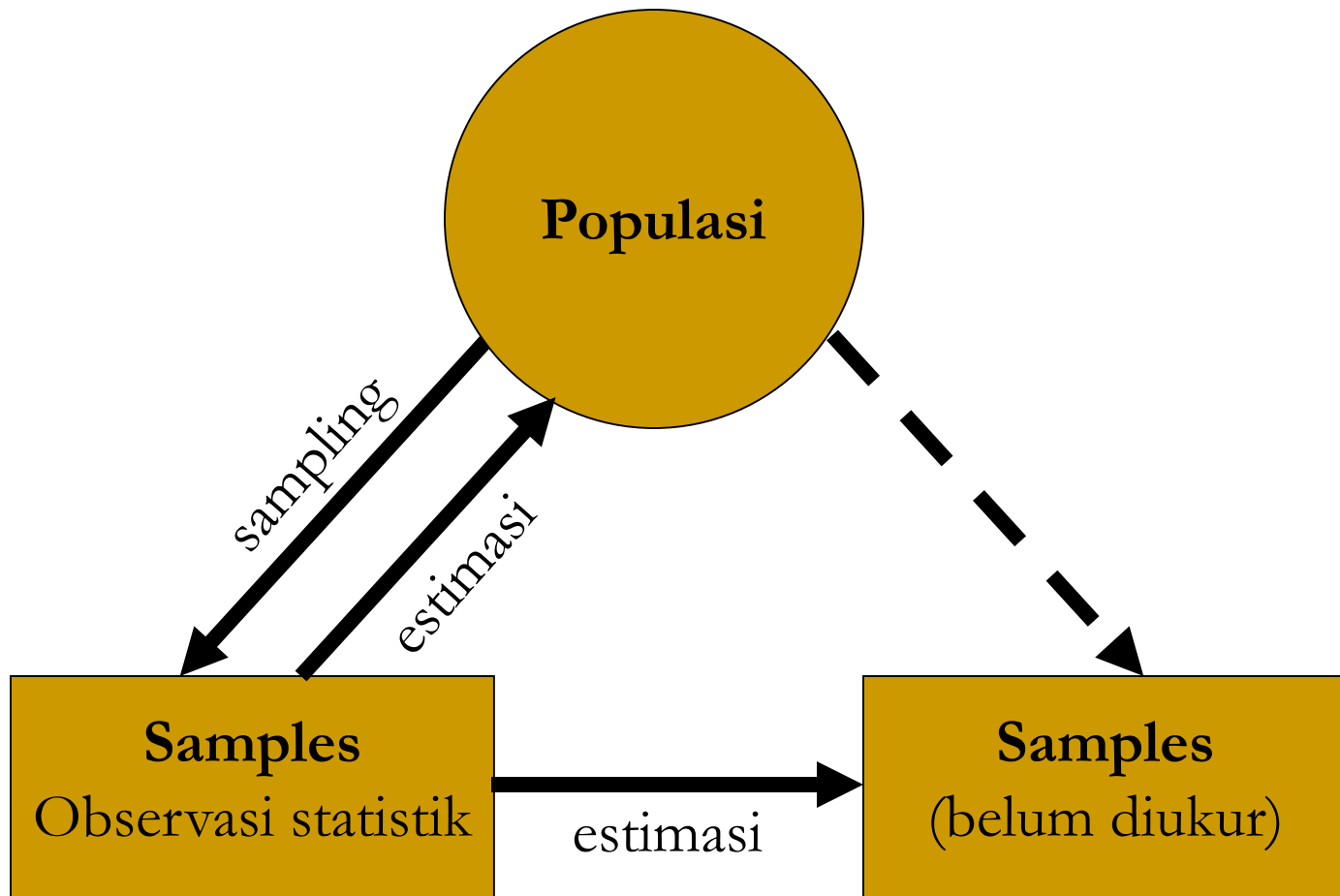
$$p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$$

Contoh

ST Contoh bivariate probability distribution

Random Variables

Properties of Random Variables



Statistik

■ Langkah

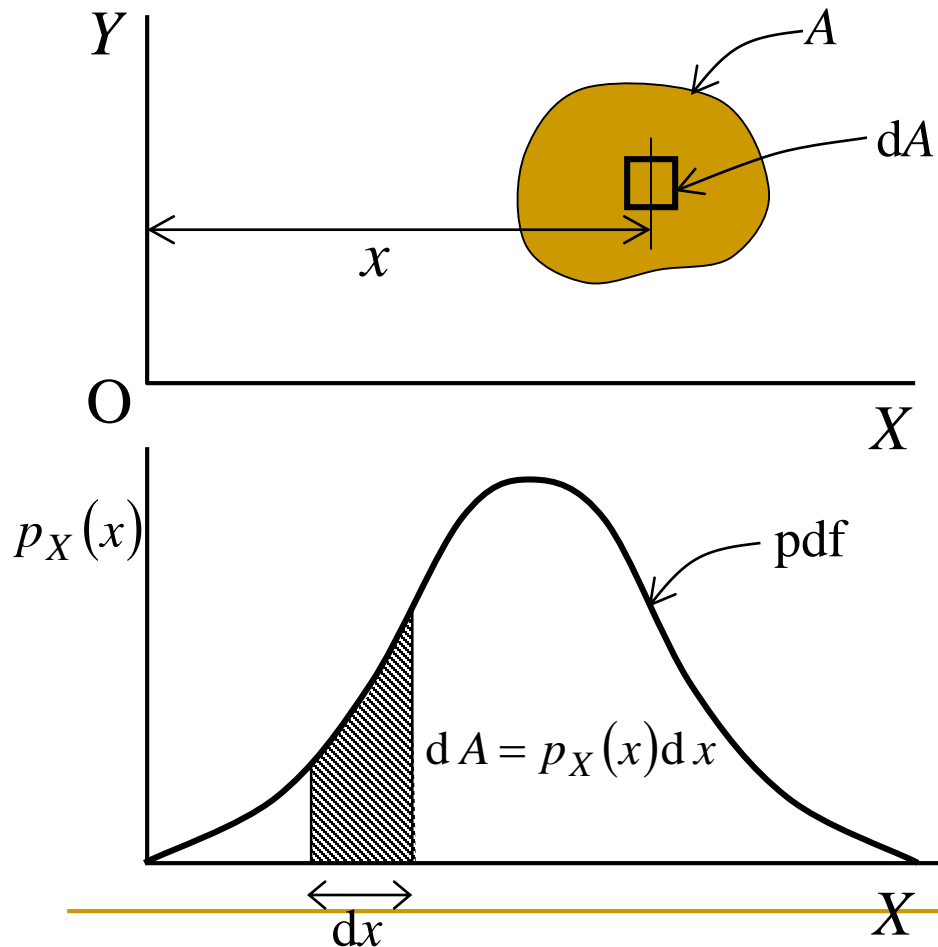
1. Pengambilan sampel
2. Observasi (analisis) terhadap sampel
3. Penyimpulan tentang perilaku sampel
4. Estimasi tentang perilaku populasi berdasarkan butir (3)
5. Estimasi tentang sampel lain (yang belum diambil) berdasarkan butir (3)

Statistik

■ Contoh

- Data debit suatu sungai selama 50 yang telah lalu dipakai sebagai dasar untuk melakukan estimasi debit sungai tersebut selama periode tak terbatas (estimasi tentang perilaku populasi).
- Informasi tersebut dapat pula dipakai untuk estimasi debit sungai tersebut selama periode tertentu pada masa yang akan datang (sampel yang belum diambil).

Moment and Expectation: Univariate Distributions



Moment pertama terhadap O

$$d\mu'_1 = x \, dA$$

$$\mu'_1 = \int_A x \, dA$$

Untuk suatu random variable

$$\mu'_1 = \int_A x \, dA$$

$$= \int_{-\infty}^{+\infty} x \, p_X(x) \, dx$$

Moment and Expectation: Univariate Distributions

- Secara umum berlaku bahwa momen ke- i terhadap O adalah
 - continuous random variables

$$\mu'_i = \int_{-\infty}^{+\infty} x^i p_X(x) dx$$

- discrete random variables

$$\mu'_i = \sum_j x_j^i f_X(x_j)$$

Moment and Expectation: Univariate Distributions

- Momen sentral ke- i : momen ke- i terhadap mean (nilai rata-rata)

$$\mu_i = \int_{-\infty}^{+\infty} (x - \mu)^i p_X(x) dx$$

Moment and Expectation: Univariate Distributions

Nilai expektasi suatu random variable X

$$E(X) = \int_{-\infty}^{+\infty} x p_X(x) dx \quad X \text{ continuous}$$

$$E(X) = \sum_j x_j f_X(x_j) \quad X \text{ discrete}$$

Dengan demikian:

$$E(X) = \mu'_i$$

$$E[(X - \mu)^i] = \mu_i$$

Statistical Measures

- Common statistical measures
 - Measure of central tendency
 - Mean
 - Mode
 - Median
 - Measure of variability
 - Range
 - Variance
 - Standard deviation
 - Measure of an individual in a population
 - z score
 - Percentile rank

Measure of Central Tendency (1)

- Nilai rata-rata (*average*)
 - rata-rata (*mean*)
 - mode → score yang paling sering muncul
 - median → score yang berada di tengah dari suatu rangkaian score urut (dari nilai kecil ke besar atau sebaliknya)

Measure of Central Tendency (2)

■ Contoh

- Jumlah hari hujan selama 11 bulan terakhir adalah sbb.

21, 21, 21, 20, 18, 16, 12, 12, 6, 2, 1

- rata-rata = 14 =AVERAGE(...)
- mode = 21 =MODE(...)
- median = 16 =MEDIAN(...)

MSEexcel

- Dari ketiga ukuran statistik tersebut, manakah yang paling baik menceritakan tentang pola jumlah hari hujan dalam 11 bulan tersebut?

Measure of Central Tendency (3)

■ Contoh

- Carilah contoh sejenis, yang berhubungan dengan pengelolaan sumberdaya air; misal:
 - perilaku penduduk dalam pemakaian air (waktu, volume, debit, dsb.)
 - data klimatologi (temperatur udara, kelembaban udara, lama penyinaran matahari, dsb.)
- Diskusikan
 - nilai rata-rata
 - mode
 - median

Measure of Central Tendency (4)

■ Contoh

- Cari dan diskusikan contoh-contoh yang berhubungan dengan bencana alam
 - debit dan tinggi muka air banjir sungai
 - lama genangan banjir di suatu kawasan
 - banjir lahar, *debris flow*
 - tanah longsor

Measure of Central Tendency (5)

- Simbol dan rumus
 - Nilai rata-rata (variabel random kontinu)

$$\begin{aligned}\mu_X &= E(X) = \mu'_1 \\ &= \int_{-\infty}^{+\infty} x \cdot p_X(x) dx\end{aligned}$$

Measure of Central Tendency (6)

- Simbol dan rumus

- Rata-rata (variabel random diskrit)

$$\bar{X} = \frac{1}{n} \sum X$$

Nilai rata-rata sampel
 n = jumlah anggota sampel

estimasi
nilai rata-
rata
populasi

$$\mu_X = \frac{1}{n} \sum X$$

Nilai rata-rata populasi
 n = jumlah anggota populasi

→ **parameter**: berdasarkan seluruh anggota populasi

→ **besaran statistik**: hanya berdasarkan sebagian anggota populasi

Measure of Central Tendency (6)

- Beberapa sifat nilai rata-rata

$$C \bar{X} = \frac{1}{n} \sum C X$$

$$C + \bar{X} = \frac{1}{n} \sum (C + X)$$

C = konstanta

Measure of Central Tendency (7)

■ Nilai rata-rata

□ Arithmetic mean

$$\bar{X} = \frac{1}{n} \sum X$$

=AVERAGE(...)

□ Geometric mean

$$\bar{X} = \left(\prod X \right)^{1/n}$$

=GEOMEAN(...)

□ Harmonic mean

$$\bar{X} = \frac{n}{\sum \frac{1}{X}}$$

=HARMEAN(...)

□ Weighted mean

$$\bar{X}_w = \frac{\sum w X}{\sum w}$$

Measure of Central Tendency (8)

- Nilai rata-rata
 - Root mean square

$$\bar{X}_{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

Measure of Central Tendency (9)

■ Median

- Variabel random kontinu

$$\mu_{md} \Rightarrow \int_{-\infty}^{\mu_{md}} p_X(x) dx = 0.5$$

- Variabel random diskrit

$\mu_{md} = x_p$ dalam hal ini x_p ditentukan dari

$$\sum_{i=1}^p f_X(x_i) = 0.5$$

Measure of Central Tendency (10)

- Mode, nilai yang paling sering muncul/terjadi
 - Variabel random kontinu

μ_{mo} = mode populasi, dihitung sedemikian hingga

$$\frac{d p_X(x)}{d x} = 0 \text{ dan } \frac{d^2 p_X(x)}{d x^2} < 0$$

- Variabel random diskrit

mode adalah suatu nilai X sedemikian hingga

$$\max_{i=1}^n f_X(x_i)$$

Measure of Variability (1)

■ Keragaman

- Variability, scatter, spread
 - menunjukkan apakah angka dalam distribusi saling berdekatan atau berjauhan
- Range → beda antara nilai tertinggi dan terendah dalam distribusi
 - mungkin biasa digunakan dalam permasalahan sehari-hari
- Standard deviation (simpangan baku)
 - biasa dipakai dalam permasalahan “teknis”

Measure of Variability (2)

- Simbol dan rumus

- Variance (ragam) variabel random kontinu merupakan momen kedua terhadap nilai rata-rata

$$\begin{aligned}\text{var}(X) &= \sigma^2 = \mu_2 \\ &= E[(X - \mu)^2] \\ &= E(X^2) - E^2(X)\end{aligned}$$

Measure of Variability (3)

- Simbol dan rumus
 - Variance (ragam) variabel random diskrit

$$\sigma^2 = \frac{\sum (X - \mu)^2}{n} \quad \text{variance populasi}$$

$$s^2 = \frac{\sum (X - \bar{X})^2}{n-1} \quad \text{variance sampel} \quad \boxed{= \text{VAR}(\dots)}$$



estimasi nilai
variance
populasi

Measure of Variability

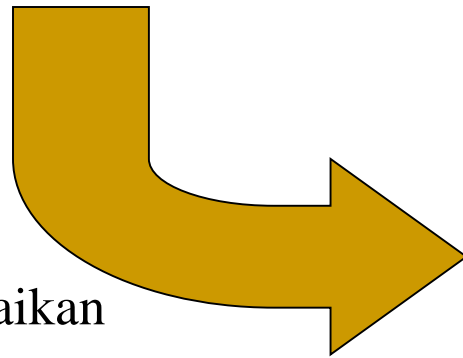
(4)

- Kenapa penyebut $n - 1$
 - menghasilkan nilai yang lebih besar daripada dibagi dengan n ; ini untuk mengompensasi kecenderungan variabilitas sampel yang lebih kecil daripada variabilitas populasi
 - dari sisi praktis, hal ini juga menunjukkan variabilitas dari sampel beranggota 1 adalah tidak ada (tidak ada variabilitas dari 1 score)

Measure of Variability

(5)

$$s^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$$



Cobalah
Saudara uraikan

$$s^2 = \frac{\sum X^2 - n \bar{X}^2}{n - 1}$$

Measure of Variability

(6)

$$\begin{aligned} s^2 &= \frac{\sum (X - \bar{X})^2}{n-1} = \frac{\sum (X^2 - 2X\bar{X} + \bar{X}^2)}{n-1} \\ &= \frac{\sum X^2 - 2\bar{X} \sum X + n\bar{X}^2}{n-1} = \frac{\sum X^2 - 2 \frac{\sum X}{n} \sum X + n \left(\frac{\sum X}{n} \right)^2}{n-1} \\ &= \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1} = \frac{\sum X^2 - n\bar{X}^2}{n-1} \end{aligned}$$

Measure of Variability

(7)

- Simbol dan rumus
 - Standard deviation (deviasi standar, simpangan baku)

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{n}}$$

deviasi standar populasi

=STDEV.S(...)

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

deviasi standar sampel

estimasi nilai
deviasi standar
populasi

Measure of Variability (8)

- Coefficient of variation

$$c_v = \frac{s}{\bar{X}}$$

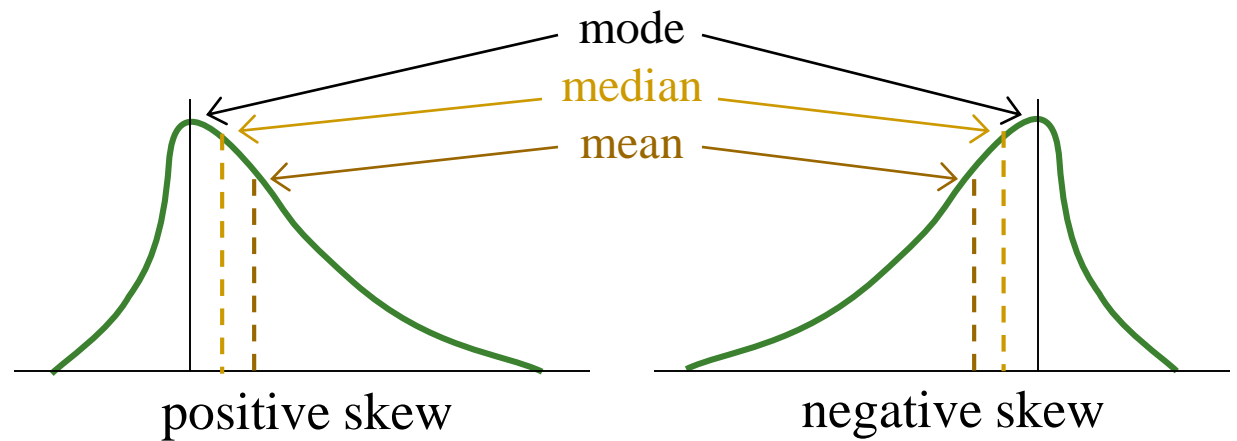
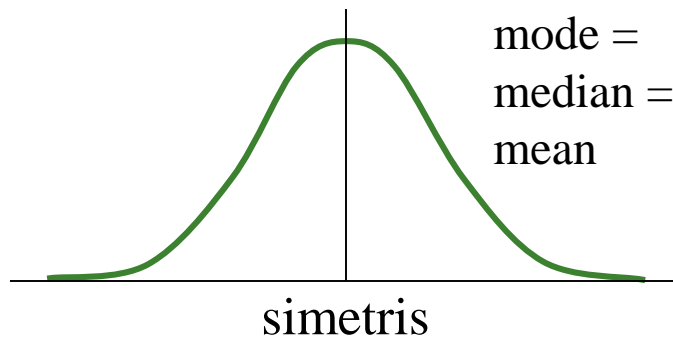
- catatan

$$\text{var}(c) = 0$$

$$\text{var}(cX) = c^2 \text{var}(X)$$

$$\text{var}(a + bX) = b^2 \text{var}(X)$$

Simetri



Simetri

$$\text{skewness populasi} = \frac{\mu - \mu_{mo}}{\sigma}$$

$$\text{skewness sampel} = \frac{\bar{X} - X_{mo}}{s}$$

$$\text{skewness coefficient, } c_s = \frac{n^2}{(n-1)(n-2)} \frac{M_3}{s_X^3}$$

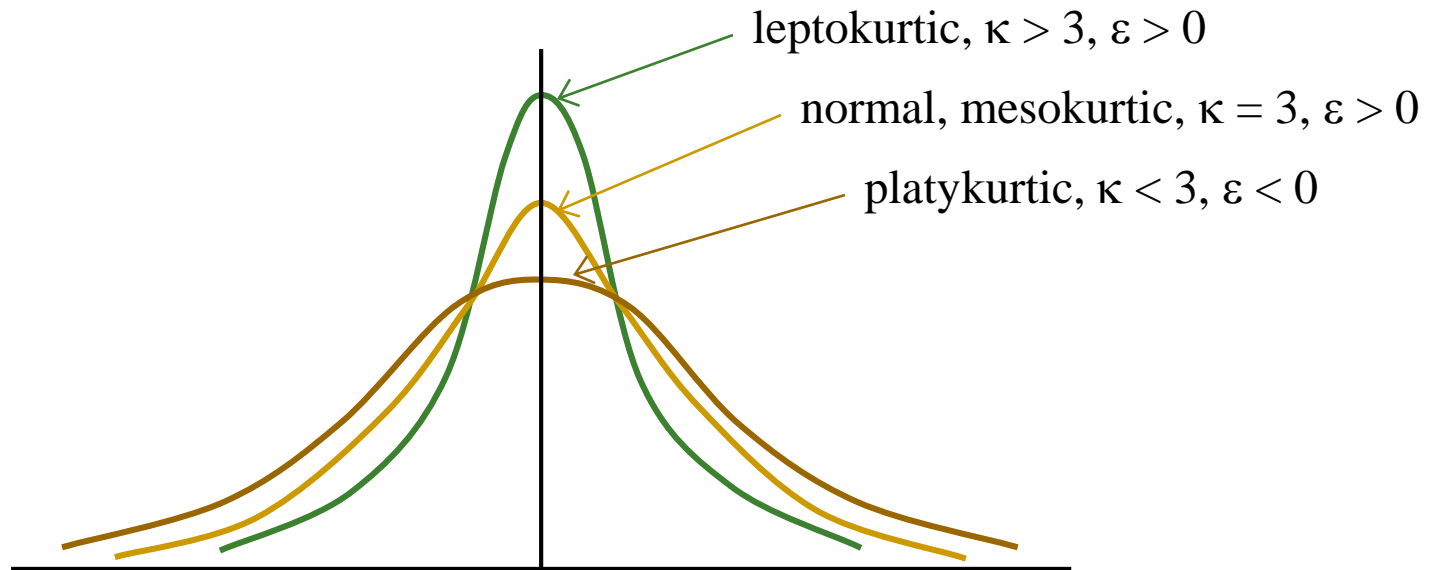
dalam persamaan di atas:

n = jumlah sampel

M_3 = momen ke-3 (sampel)

s_X = simpangan baku (sampel)

Peakedness



kurtosis: $\kappa = \frac{\mu_4}{\mu_2^2}$ (populasi)

$k = \frac{M_4}{s_X^4}$ (sampel)

$\varepsilon = \kappa - 3$

Sample Moments

Lihat catatan: ST Sample moments

Some Measures of An Individual in A Population

(1)

■ z scores

$$z_X = \frac{X - \mu}{\sigma} \quad z_X = \frac{X - \bar{X}}{s}$$

■ Percentile rank

untuk menunjukkan posisi suatu score dalam populasi

$$PR_X = \frac{B + \frac{1}{2}E}{n} (100)$$

B = jumlah score yang bernilai di bawah X
 E = jumlah score yang bernilai sama dengan X
 n = jumlah score seluruhnya

untuk populasi besar

Some Measures of An Individual in A Population (2)

■ Beberapa fungsi di dalam MS Excel

□ =RANK(...)

- posisi suatu nilai (angka) pada suatu urutan angka

□ =PERCENTILE(...)

- nilai percentile dalam suatu kisaran angka

□ =PERCENTRANK(...)

- posisi suatu nilai (angka) dalam suatu urutan angka, dalam persen

$$= \frac{B}{(B + A)} (100)$$

B = jumlah score yang bernilai lebih kecil daripada X
 A = jumlah score yang bernilai lebih besar daripada X

perhatikan perbedaannya dengan PR_X

Sekian