UNIVERSITAS GADJAH MADA
DEPARTEMEN TEKNIK SIPIL DAN LINGKUNGAN
PRODI MAGISTER TEKNIK SIPIL

Statistika Teknik

Variabel Random

Pengertian

- Variabel random (variabel acak)
  - suatu fungsi yang didefinisikan pada sample space
- Jenis
  - Variabel random diskret
  - Variabel random kontinu
- Contoh
  - jumlah hari hujan selama 1 tahun → diskret
  - jumlah (volume) hujan selama 1 tahun → kontinu

Variabel Random

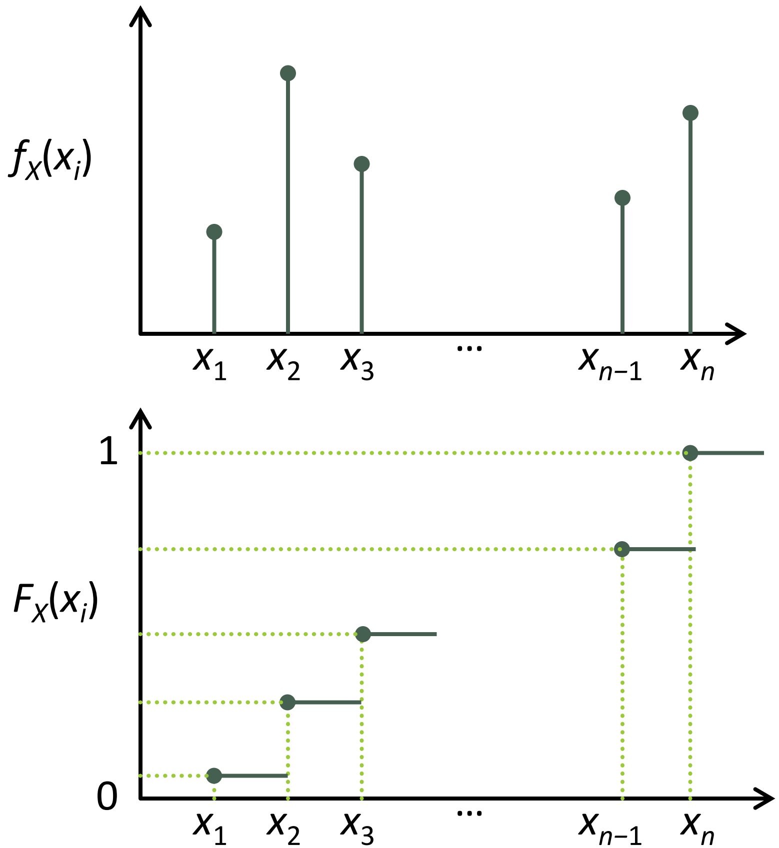
- Notasi
  - X → variabel random
  - x → nilai variabel random
- Fungsi
  - Suatu fungsi variabel random adalah variabel random pula
  - Jika X adalah variabel random, maka $Z = g(X)$ adalah juga variabel random

Univariate Probability Distributions

X = variabel random diskret

$= x_1,\ x_2,\ x_3,\ \ldots,\ x_n$

$$\begin{array}{cccc} x_1 & x_2 & x_3 & x_n \\ \downarrow & \downarrow & \downarrow & \downarrow \\ f_X(x_1) & f_X(x_2) & f_X(x_3) & f_X(x_n) \end{array} \Bigg\} \quad \begin{array}{l} \text{probabilitas} \\ \sum f_X(x_i) = 1 \end{array}$$



a discrete probability distribution

a discrete cumulative probability distribution

probabilitas

$x \leq x_i$

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

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

- Distribusi probabilitas 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, x2 , …, x_n yang masing-masing memiliki probabilitas $f_X(x_1)$, $f_X(x_2)$, …, $f_X(x_n)$ dan $\sum f_X(x_i) = 1$, maka X adalah variabel random diskret.

- Frekuensi relatif

$$f_{x_i} = F_{x_i} - F_{x_{i-1}}$$

- Probabilitas

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

- Frekuensi relatif kumulatif

$$F_{x_i} = \sum_{j=1}^{i} f_{x_j}$$

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

Variabel Random Kontinu

- Probabilitas

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

n_i = jumlah data di klas ke-i

n = jumlah seluruh data

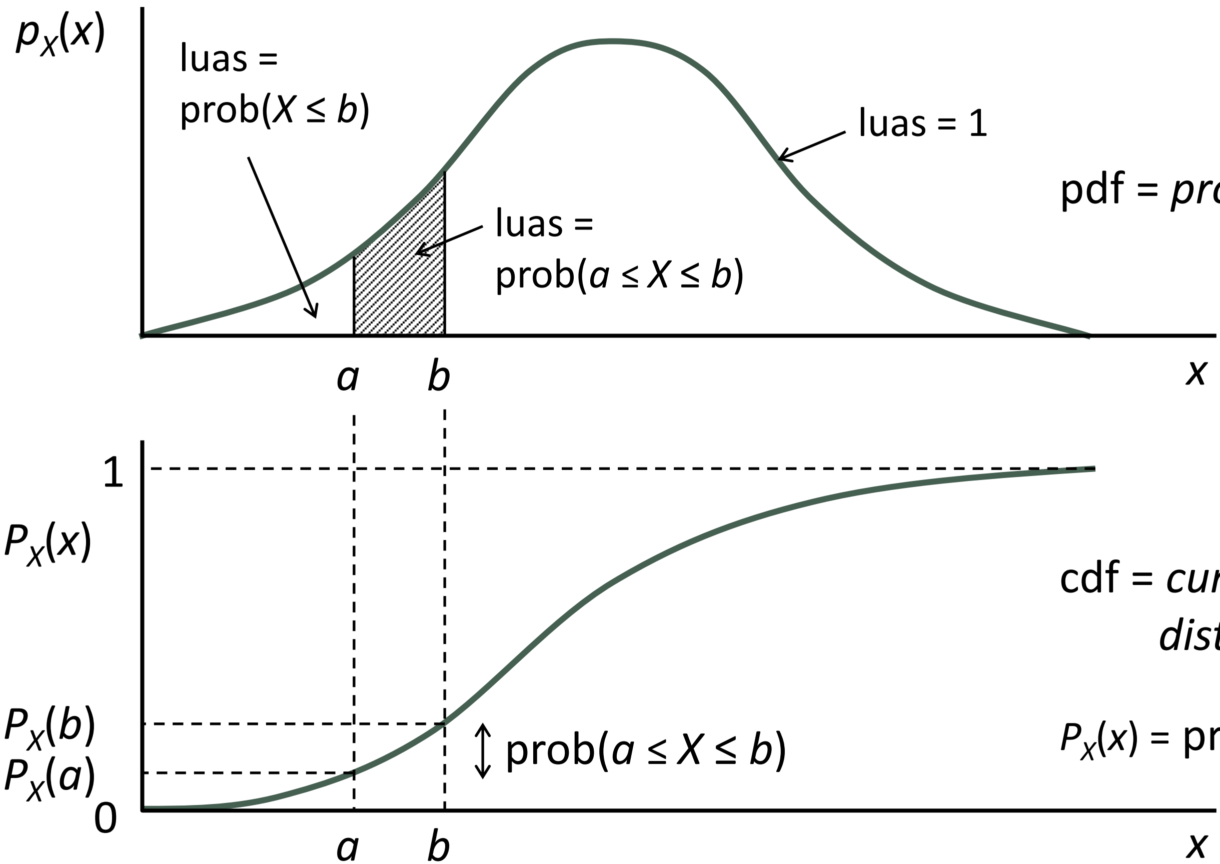
- Dengan demikian f_{xi} dapat dipandang sebagai nilai estimasi probabilitas

f_{xi} → estimasi prob (A)

histogram frekuensi → pendekatan distribusi probabilitas

frekuensi kumulatif → pendekatan distribusi probabilitas kumulatif

variabel random kontinu diperlakukan seolah-olah variabel random diskret



pdf = *probability density function*

cdf = *cumulative probability distribution function*

$P_X(x)$ = prob($X \le x$)

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

$P_X(x)$ = cumulative probability distribution function

$$P_X(x) = \mathrm{prob}(X \le x) \Rightarrow dP_X(x) = p_X(x)\,dx \Rightarrow P_X(x) = \int_{-\infty}^{x} p_X(t)\,dt$$

Beberapa Sifat Probabilitas

(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) $\mathrm{prob}(a \leq x \leq b) = \int_a^b p_X(t)dt = P_X(b) - P_X(a)$

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

$$\mathrm{prob}(a \leq x \leq b) = \mathrm{prob}(a < x \leq b) = \mathrm{prob}(a \leq x < b) = \mathrm{prob}(a < x < b)$$

Kala Ulang

$$\text{prob}(x \geq b) = \text{prob}(x > b)$$

- Jadi dalam definisi kala ulang
  1. suatu kejadian yang menyamai atau melampaui suatu nilai tertentu
  2. suatu kejadian yang melampaui suatu nilai tertentu

Kedua definisi, 1 dan 2, adalah sama mengingat probabilitas suatu kejadian (*event*) variabel random kontinu menyamai suatu nilai tertentu (konstanta) adalah nol

Contoh #1

ST Contoh *univariate probability distribution*

Contoh #2

- Diketahui suatu variabel random X memiliki fungsi kerapatan probabilitas (pdf) sbb.

$$p_X(x) = \begin{cases} x/2 & \text{untuk } 0 < x < 2 \\ 0 & \text{untuk nilai } x \text{ yang lain} \end{cases}$$

  - Gambarlah pdf tersebut
  - Tunjukkan bahwa prob(0 < X < 2) = 1
  - Hitunglah prob(X < 1.5)= PX(1.5)
  - Hitunglah prob(0.5 < X < 1.5)

Contoh #3

- Pengolahan data *annual series* curah hujan harian maksimum, H mm, di suatu stasiun ARR (*Automatic Rainfall Recorder*) menunjukkan bahwa sebaran probabilitas suatu besaran curah hujan, $p_H(h)$, dapat dinyatakan dengan suatu fungsi (pdf) sbb.

$$p_H(h) = \begin{cases} \dfrac{1}{75} & \text{untuk } 0 < h < 50 \\ \dfrac{1}{3750}(100 - h) & \text{untuk } 50 < h < 100 \\ 0 & \text{untuk nilai } h \text{ yang lain} \end{cases}$$

  - Gambarlah pdf tsb.
  - Carilah fungsi cdf berdasarkan pdf tsb.
  - Hitunglah prob(40 mm < H < 60 mm)

Bivariate Distributions

- Pada bahasan sebelumnya, variabel random adalah variabel tunggal (*univariate distribution*)
- Pada bahasan berikut ini, variabel random terdiri dari dua variabel (*bivariate distributions*)
  - Apabila kita ingin mempelajari perilaku dua atau lebih variabel random, maka kita perlu menghitung probabilitas gabungan atau probabilitas bersama (*joint probabilities*)
  - Probabilitas gabungan → pdf gabungan, *joint probability density function*

Bivariate Distributions

- *Joint probability density function*

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

$$P_{X,Y}(x,y) = \mathrm{prob}(X < x \text{ dan } Y < y)$$ → cdf

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

Bivariate Distributions

- Beberapa persamaan

$P_{X,Y}(x,\infty)$ ⟹ cdf variabel random *X* saja (*univariate*)

$P_{X,Y}(\infty,y)$ ⟹ cdf variabel random *Y* saja (*univariate*)

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

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

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

Distribusi Marginal

- Dua variabel random *X* dan *Y*
  - Ingin diketahui perilaku *X* tanpa mempertimbangkan variabel *Y*
  - *Marginal density*

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

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

Distribusi Marginal

- Dua variabel random, *X* dan *Y*
  - *Cumulative marginal distribution*

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

$$\begin{aligned} P_X(x) &= P_X(x,\infty) = \mathrm{prob}(X \leq x \wedge Y \leq \infty) \\ &= \mathrm{prob}(X \leq x) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{+\infty} p_X(s,t)\,dt\,ds \\ &= \int_{-\infty}^{x} p_X(s)\,ds \end{aligned}$$

Distribusi Marginal

- Untuk variabel *Y*
  - *Marginal density*

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

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

  - *Cumulative marginal distribution*

$$P_Y(y) = P_Y(\infty,y) = \mathrm{prob}(Y \leq \infty \wedge Y \leq y) = \mathrm{prob}(Y \leq y) = \int_{-\infty}^{y} p_Y(t)dt$$

Distribusi Bersyarat (*Conditional Probabilities*)

- Dua variabel random, *X* dan *Y*
  - Ingin diketahui perilaku *X* yang bergantung pada *Y*
    - Distribusi *X* jika $Y = y_0$
    - Distribusi *Y* jika $x_1 \leq X \leq x_2$

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

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

$$p_{X|Y}(x|y = y_0\,) = \frac{p_{X,Y}(x,y_0)}{p_Y(y_0)}$$

yang lebih sering dituliskan sebagai berikut:

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

Independence

- Dua variabel random, *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 kerapatan marginal kedua variabel adalah:

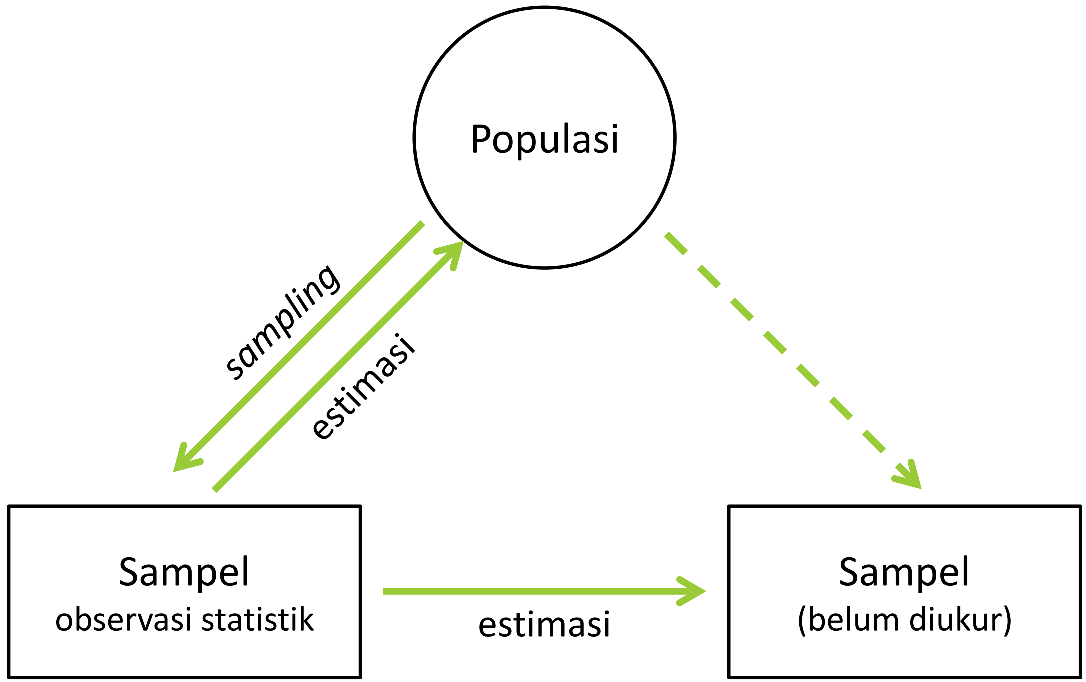
    $$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$$

Contoh

ST Contoh *bivariate probability distribution*

Variabel Random

Sifat Variabel Random


Populasi
sampling
estimasi
Sampel
observasi statistik
estimasi
Sampel
(belum diukur)

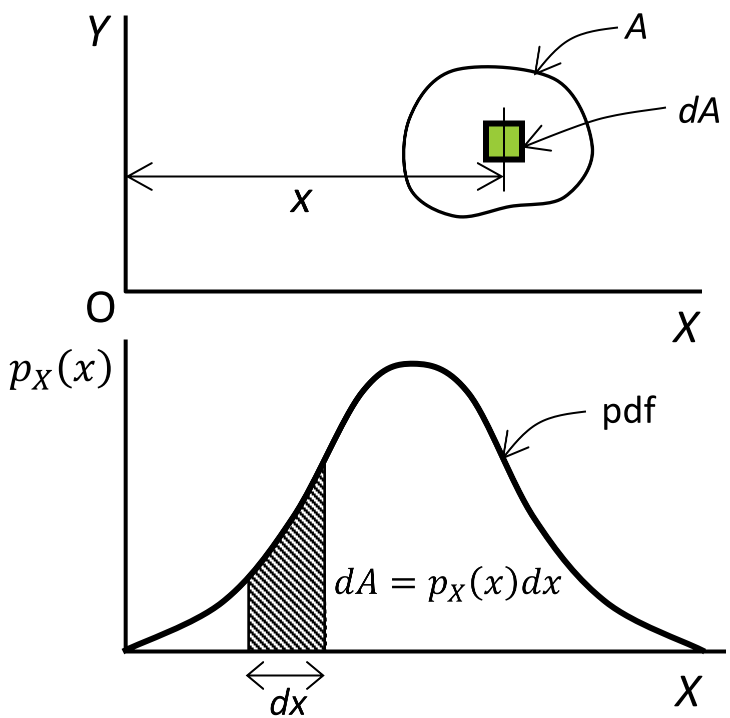
Statistika

- Langkah
  - Pengambilan sampel
  - Observasi (analisis) terhadap sampel
  - Penyimpulan tentang perilaku sampel
  - Estimasi tentang perilaku populasi berdasarkan butir (3)
  - Estimasi tentang sampel lain (yang belum diambil/diukur) berdasarkan butir (3)

Statistika

- Contoh
  - Data debit suatu sungai selama 50 tahun 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



Momen pertama terhadap O

$$d\mu_1' = xdA$$

$$\mu_1' = \int_A xdA$$

Untuk suatu variabel random

$$\mu_1' = \int_A xdA$$

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

Moment and Expectation: Univariate Distributions

- Secara umum berlaku bahwa momen ke-*i* terhadap O adalah:
  - distribusi random kontinu $$\mu_i' = \int_{-\infty}^{+\infty} x^i p_X(x) dx$$
  - distribusi random diskret $$\mu_i' = \sum_j x_j^i f_X(x_j)$$
- Momen sentral ke-*i*: momen ke-*i* terhadap *mean* (nilai rerata)

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

Moment and Expectation: Univariate Distributions

- Nilai ekspektasi suatu variabel random *X*

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x p_X(x) dx \qquad X \text{ kontinu}$$

$$\mathrm{E}(X) = \sum_{j}^{+\infty} x_j f_X(x_j) \qquad X \text{ diskret}$$

- Dengan demikian

$$\mathrm{E}(X) = \mu_i'$$

$$\mathrm{E}\left[(X-\mu)^i\right] = \mu_i$$

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

- Nilai rerata (*average*)
  - rerata (*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

- Contoh
  - Jumlah hari hujan selama 11 bulan terakhir adalah sbb.
    21, 21, 21, 20, 18, 16, 12, 12, 6, 2, 1
    - rerata = 14 =AVERAGE(...)
    - *mode* = 21 =MODE(...)
    - median = 16 =MEDIAN(...)

MSExcel

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

Measure of Central Tendency

- 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 rerata
    - *mode*, modus
    - median

Measure of Central Tendency

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

$$\mu_X = \mathrm{E}(X) = \mu_1'$$

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

Measure of Central Tendency

- Simbol dan rumus
  - Rerata (variabel random diskret)

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Nilai rerata sampel
n = jumlah anggota sampel

estimasi nilai rrata populasi

$$\mu_X = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Nilai rerata populasi
n = jumlah anggota populasi

parameter: berdasarkan seluruh anggota populasi

besaran statistik: hanya berdasarkan sebagian anggota populasi

Measure of Central Tendency

- Beberapa sifat nilai rerata

$$C\bar{X} = \frac{1}{n}\sum_{i=1}^{n} Cx_i$$

$$C + \bar{X} = \frac{1}{n}\sum_{i=1}^{n}(C + x_i)$$

C = konstanta

Measure of Central Tendency

- Nilai rerata
  - *Arithmetic mean* $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$ =AVERAGE(…)
  - *Geometric mean* $$\bar{X} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$ =GEOMEAN(…)
  - *Harmonic mean* $$\bar{X} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$ =HARMEAN(…)
  - *Weighted mean* $$\bar{X} = \frac{\sum_{i=1}^{n} w x_i}{\sum_{i=1}^{n} w}$$

Measure of Central Tendency

- Nilai rerata
  - *Root mean square*

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

Measure of Central Tendency

- Median
  - Variabel random kontinu

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

  - Variabel random diskret

$\mu_{md} = x_p$ dalam hal ini x_p ditentukan dari $\sum_{i=1}^{p} f_X(x_i) = 0.5$

Measure of Central Tendency

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

    μ_{mo} = mode populasi dihitung sedemikian sehingga:

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

  - Variabel random diskret

    mode adalah suatu nilai x sedemikian sehingga: $\max_{1 \le i \le n} f_X(x_i)$

Measure of Variability

- 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

- Simbol dan rumus
  - *Variance* (varians, ragam) variabel random **kontinu** merupakan momen kedua terhadap nilai rerata

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

Measure of Variability

- Simbol dan rumus
  - *Variance* (variansi, ragam) variabel random **diskret**

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

variansi populasi

=VAR.P(…)

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1}$$

variansi sampel

=VAR.S(…)

estimasi nilai varians populasi

Measure of Variability

- 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

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1}$$

Cobalah
Saudara uraikan

$$s^2 = \frac{\sum_{i=1}^{n} {x_i}^2 - n\bar{X}^2}{n-1}$$

Measure of Variability

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1} = \frac{\sum_{i=1}^{n}\left({x_i}^2 - 2x_i\bar{X} + \bar{X}^2\right)}{n-1}$$

$$= \frac{\sum_{i=1}^{n} {x_i}^2 - 2\bar{X}\sum_{i=1}^{n} x_i + n\bar{X}^2}{n-1} = \frac{\sum_{i=1}^{n} {x_i}^2 - 2\frac{\sum_{i=1}^{n} x_i}{n}\sum_{i=1}^{n} x_i + n\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} {x_i}^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{\sum_{i=1}^{n} {x_i}^2 - n\bar{X}^2}{n-1}$$

Measure of Variability

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

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$

simpangan baku populasi

=STDEV.P(…)

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1}}$$

simpangan baku sampel

=STDEV.S(…)

estimasi nilai deviasi standar populasi

Measure of Variablitiy

- *Coefficient of variation*

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

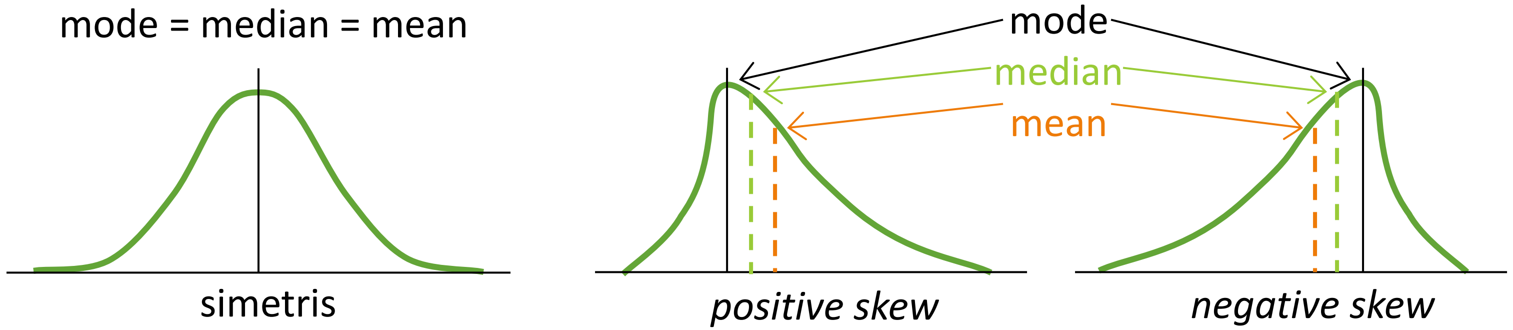
- Catatan

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

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

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

Simetri


mode = median = mean
simetris
mode
median
mean
positive skew
negative skew

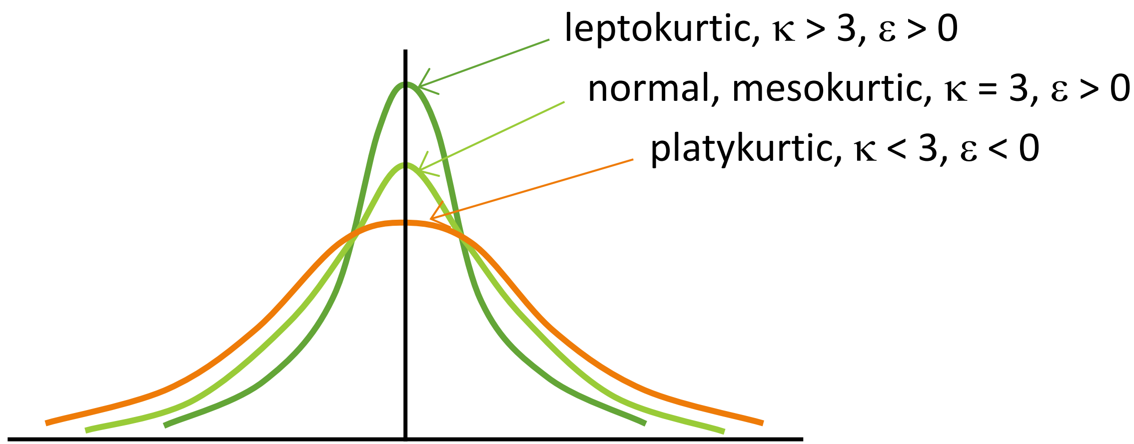
Simetri

- *Skewness* populasi $= \dfrac{\mu - \mu_{mo}}{\sigma}$
- *Skewness* sampel $= \dfrac{\bar{X} - X_{mo}}{s}$
- *Skewness coefficient* $c_s = \dfrac{n^2}{(n-1)(n-2)} \dfrac{M_3}{{s_X}^3}$

Dalam persamaan di atas:

n	jumlah sampel
M_3	momen ke-3 (sampel)
s_X	simpangan baku (sampel)

Peakedness



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

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

$\varepsilon = \kappa - 3$ (*excess kurtosis*)

Sample Moments

Lihat catatan: ST Sample moments

Some Measures of An Individual in A Population

- *z score*

$$z_X = \frac{x - \mu}{\sigma}$$

untuk menunjukkan posisi suatu score dalam populasi

- *Percentile rank*

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

untuk populasi besar

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

Some Measures of An Individual in A Population

- Beberapa fungsi dalam MSExcel
  - =RANK.AVG(...), =RANK.EQ(...)
    - posisi suatu nilai (angka) pada suatu urutan angka
  - =PERCENTILE.EXC(...), =PERCENTILE.INC(...)
    - nilai percentile dalam suatu kisaran angka
  - =PERCENTRANK.EXC(...), PERCENTRANK.INC(...)
    - posisi suatu nilai (angka) dalam suatu urutan angka, dalam persen

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

perhatikan perbedaannya dengan PR_X

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

Terima kasih