STATISTIK INDUSTRI

Definisi

- Statistika: Ilmu mengumpulkan, mengolah, meringkas, menyajikan dan interpretasi data untuk dasar pengambilan keputusan
- Populasi : Seluruh objek penelitian
- Sampel : Data yang diambil dari populasi

Pengumpulan Data

- Sampel representatif dari populasi
  - Sampel random
  - Sampel sistematis
  - Sampel kelompok (cluster)
- Metode pengumpulan data : tidak dipelajari di kuliah ini
- Data menurut sumber: Primer dan Sekunder

DATA & VARIABEL

Data adalah sekumpulan datum yang berisi fakta-fakta serta gambaran suatu fenomena yang dikumpulkan, dirangkum, dianalisis dan selanjutnya diinterpretasikan.

Variabel adalah karakteristik data yang menjadi perhatian.

DATA MENURUT SKALA PENGUKURAN

a. **Nominal**, sifatnya hanya untuk membedakan antar kelompok.

   Contoh: Jenis kelamin,
   Jurusan dalam suatu sekolah tinggi
   (Manajemen, Akuntansi).

b. **Ordinal**, selain memiliki sifat nominal, juga menunjukkan peringkat.

   Contoh: Tingkat pendidikan (SD, SMP, SMA),
   Ranking

DATA MENURUT SKALA PENGUKURAN (L)

Interval, selain memiliki sifat data ordinal, juga memiliki sifat interval antar observasi dinyatakan dalam unit pengukuran yang tetap.

Contoh: Nilai Test

Rasio, selain memiliki sifat data interval, skala rasio memiliki angka 0 (nol) dan perbandingan antara dua nilai mempunyai arti.

Contoh: Temperatur

Berat badan,

JENIS DATA MENURUT SIFATNYA

1. **Kualitatif**
   - Berupa label/nama-nama yang digunakan untuk mengidentifikasikan atribut suatu elemen
   - Skala pengukuran: Nominal atau Ordinal
   - Data bisa berupa *numeric* atau *nonnumeric*

2. **Kuantitatif**
   - Mengindikasikan seberapa banyak (*how many*/diskret atau *how much*/kontinu)
   - Data selalu *numeric*
   - Skala pengukuran: Interval dan Rasio

JENIS DATA MENURUT WAKTU PENGUMPULANNYA

Cross-sectional Data

yaitu data yang dikumpulkan pada waktu tertentu yang sama atau hampir sama

Contoh: Jumlah mahasiswa STEKPI TA 2005/2006, Jumlah perusahaan *go public* tahun 2006

Time Series Data

yaitu data yang dikumpulkan selama kurun waktu/periode tertentu

Contoh: Pergerakan nilai tukar rupiah dalam 1 bulan, Produksi Padi Indonesia tahun 1997-2006

CARA PENYAJIAN DATA

1. **Tabel**
   - Tabel satu arah (*one-way table*)
   - Tabulasi silang (lebih dari satu arah (*two-way table*), dst.)
   - Tabel Distribusi Frekuensi

2. **Grafik**
   - Batang (*Bar Graph*), untuk perbandingan/pertumbuhan
   - Lingkaran (*Pie Chart*), untuk melihat perbandingan (dalam persentase/proporsi)
   - Grafik Garis (*Line Chart*), untuk melihat pertumbuhan
   - Grafik Peta, untuk melihat/menunjukkan lokasi

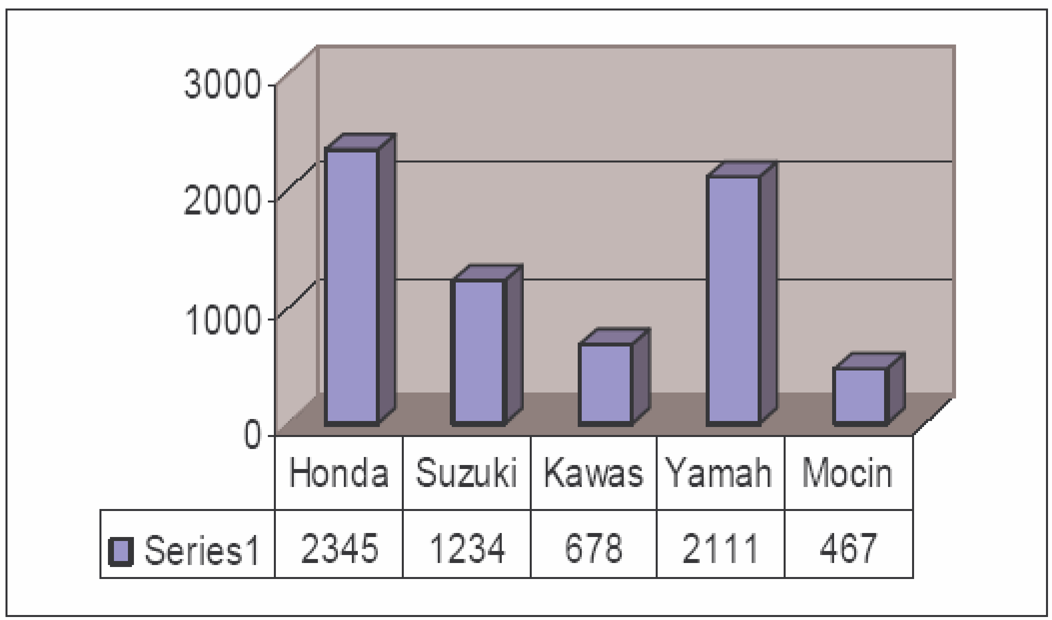
Penyajian Data

- Tabel

Kendaraan	Jumlah
Honda	2345
Suzuki	1234
Kawasaki	2111
Yamaha	678
Mocin	467
Total	6835

Penyajian Data

- Grafik



MANFAAT
TABEL DAN GRAFIK

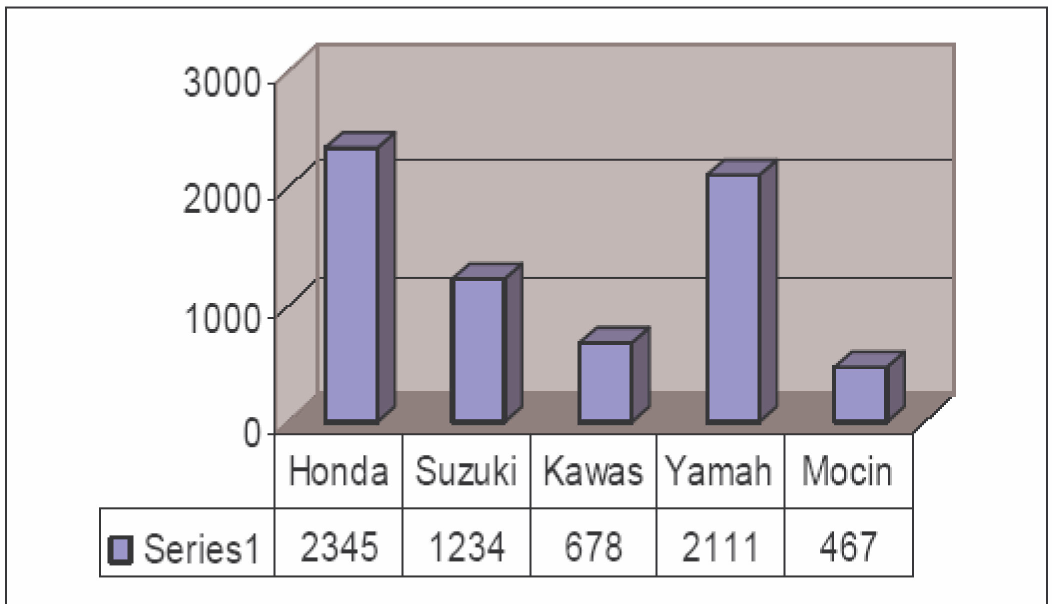
- Meringkas/rekapitulasi data, baik data kualitatis maupun kuantitatif
  - Data kualitatif berupa distribusi Frekuensi, frekuensi relatif, persen distribusi frekuensi, grafik batang, grafik lingkaran.
  - Data kuantitatif berupa distribusi frekuensi, relatif frekuensi dan persen distribusi frekuensi, diagram/plot titik, histogram, distribusi kumulatif, ogive.
- Dapat digunakan untuk melakukan eksplorasi data
- Membuat tabulasi silang dan diagram sebaran data

GRAFIK BATANG (*BAR GRAPH*)

- Bermanfaat untuk merepresentasikan data kuantitatif maupun kualitatif yang telah dirangkum dalam frekuensi, frekuensi relatif, atau persen distribusi frekuensi.
- Cara:
  - Pada sumbu horisontal diberi label yang menunjukkan kelas/kelompok.
  - Frekuensi, frekuensi relatif, maupun persen frekuensi dinyatakan dalam sumbu vertikal yang dinyatakan dengan menggunakan gambar berbentuk batang dengan lebar yang sama/tetap.

Penyajian Data

- Grafik



Grafik dari data…

Mapel	Rata-rata
Matematika	8,5
Bhs Indonesia	7,2
Bhs Inggris	9,1
I P A	4,8
I P S	6,3

GRAFIK LINGKARAN (*PIE CHART*)

- Digunakan untuk mempresentasikan distribusi frekuensi relatif dari data kualitatif maupaun data kuantitatif yagn telah dikelompokkan.
- Cara:
  - Gambar sebuah lingkaran, kemudian gunakan frekuensi relatif untuk membagi daerah pada lingkaran menjadi sektor-sektor yang luasnya sesuai dengan frekuensi relatif tiap kelas/kelompok.
  - Contoh, bila total lingkaran adalah 360^o maka suatu kelas dengan frekuensi relatif 0,25 akan membutuhkan daerah seluas $(0,25)(360) = 90^o$ dari total luas lingkaran.

Freq Pemakaian Kendaraan Roda Dua di Kalangan Mahasiswa

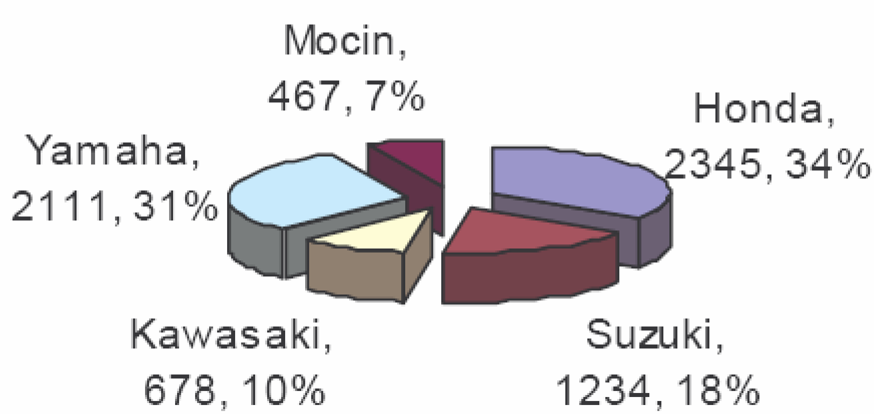


Diagram Lingkaran dari data…

Mapel	Rata-rata
Matematika	8,5
Bhs Indonesia	7,2
Bhs Inggris	9,1
I P A	4,8
I P S	6,3

OGIVE

- Merupakan grafik dari distribusi frekuensi kumulatif.
- Nilai data disajikan pada garis horisontal (sumbu-x).
- Pada sumbu vertikal dapat disajikan:
  - Frekuensi kumulatif, atau
  - Frekuensi relatif kumulatif, atau
  - Persen frekuensi kumulatif
- Frekuensi yang digunakan (salah satu diatas)masing-masing kelas digambarkan sebagai titik.
- Setiap titik dihubungkan oleh garis lurus.

OGIVE

Contoh: Bengkel Hudson Auto

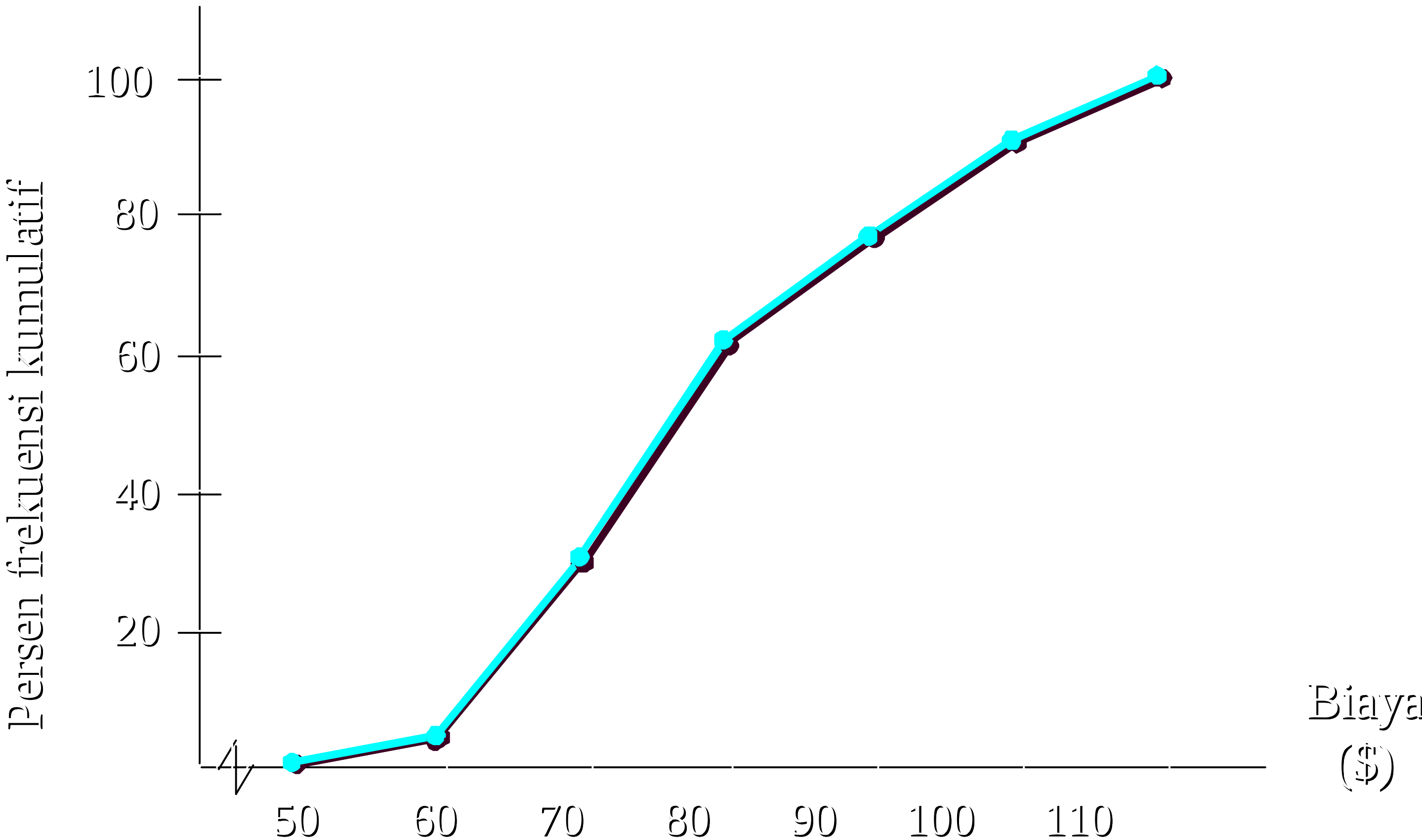


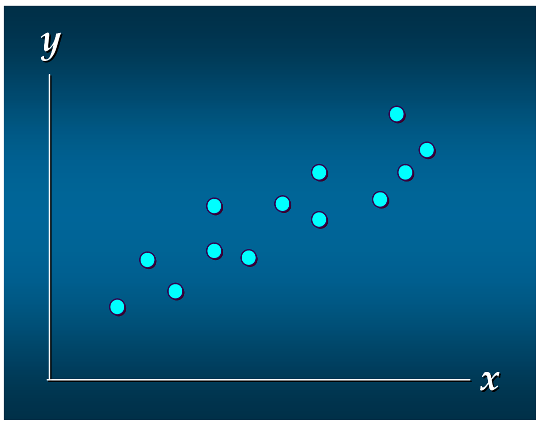
DIAGRAM SCATTER

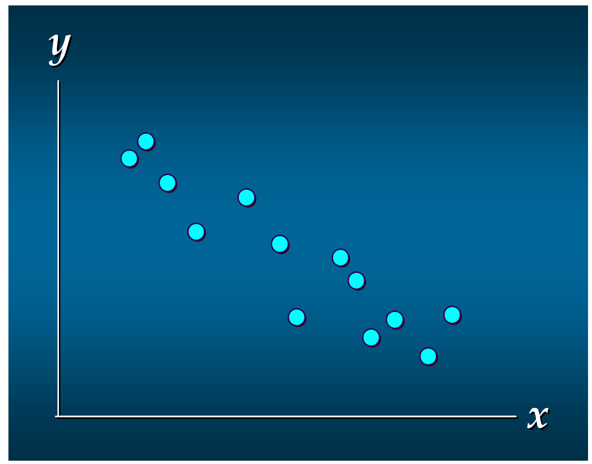
Diagram scatter (*scatter diagram*) merupakan metode presentasi secara grafis untuk menggambarkan hubungan antara dua variabel kuantitatif.

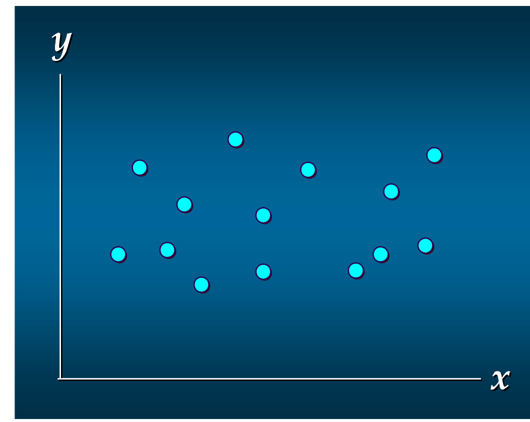
Salah satu variabel digambarkan pada sumbu horisontal dan variabel lainnya digambarkan pada sumbu vertikal.

Pola yang ditunjukkan oleh titik-titik yang ada menggambarkan hubungan yang terjadi antar variabel.

POLA HUBUNGAN PADA DIAGRAM SCATTER







Hubungan Positif
Jika X naik, maka Y juga naik dan jika X turun, maka Y juga turun

Hubungan Negatif
Jika X naik, maka Y akan turun dan jika X turun, maka Y akan naik

Tidak ada hubungan antara X dan Y

PROSEDUR PENGGUNAAN TABEL & GRAFIK

- Data
  - Data Kualitatif
    - Metode Tabel
      - Distr. Frekuensi
      - Distr. Frek. Relatif
      - % Distr. Frek.
      - Tabulasi silang
    - Metode Grafik
      - Grafik Batang
      - Grafik Lingkaran
  - Data Kuantitatif
    - Metode Tabel
      - Distr. Frekuensi
      - Distr. Frek. Relatif
      - Distr. Frek. Kum.
      - Distr. Frek. Relatif Kum.
      - Diagram Batang-Daun
      - Tabulasi silang
    - Metode Grafik
      - Plot Titik
      - Histogram
      - Ogive
      - Diagram Scatter

DISTRIBUSI FREKUENSI

- Merupakan tabel ringkasan data yang menunjukkan frekuensi/banyaknya item/obyek pada setiap kelas yang ada.

- Tujuan: mendapatkan informasi lebih dalam tentang data yang ada yang tidak dapat secara cepat diperoleh dengan melihat data aslinya.

DISTRIBUSI FREKUENSI RELATIF

- Merupakan fraksi atau proporsi frekuensi setiap kelas terhadap jumlah total.
- Distribusi frekuensi relatif merupakan tabel ringkasan dari sekumpulan data yang menggambarkan frekuensi relatif untuk masing-masing kelas.

Cara Membuat Distribusi Frequensi

- Data dikelompokkan dalam kelas interval
- Idealnya terdiri dari 5 sampai 15 kelas interval
  - Aturan Sturges : jml kelas k=1+3.222*log(n)
  - Lebar kelas = Rentang data/jml kelas
- Kelas Interval tidak saling overlap

Data Uang Kiriman Mhs UGM					
67	44	35	48	22	51
52	56	61	47	37	61
72	48	44	41	66	26
42	44	51	62	49	73
21	69	52	72	69	33
55	56	77	85	42	71
47	27	82	25	54	64
66	34	57	72	59	57
54	47	63	54	58	55
37	59	73	52	75	56
37	20	49	108	61	47
34	51	67	28	66	87
59	42	33	93	99	68
51	78	78	37	97	97

Distribusi Frekuensi Uang Kiriman Mhs UGM

Uang Kiriman (puluhan ribuan rupiah)	Banyak Mahasiswa
20,0 – 29,9	7
30,0 – 39,9	9
40,0 – 49,9	16
50,0 – 59,9	21
60,0 – 69,9	14
70,0 – 79,9	9
80,0 – 89,9	4
90,0 – 99,9	3
100,0 – 109,9	1
	84

Nilai Ujian Statistik	Banyaknya Mahasiswa
2	1
3	6
4	11
5	16
6	18
7	9
8	7
9	2
	70

Ukuran Tengah dan Deviasi

- Harga Tengah
  - Rata-rata (Mean)
  - Median
  - Modus
  - Mean Geometrik
- Harga Deviasi
  - Variansi
  - Standart Deviasi
  - Range
  - Standart Error

Rata-rata Data Tunggal

- Rumus Rata-rata (Mean)
  - Jumlah data dibagi banyak data

n angka, $X_1, X_2, \dots, X_n$

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$= \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Diketahui harga gula perkilodi 5 pasar di kecamatan minggiran
- Dapat dihitung rata-rata harga gula

4	4,2	4,1	4	4,1

$$\bar{X} = \frac{4 + 4{,}2 + 4{,}1 + 4 + 4{,}1}{5}$$

Rata-Rata Data Interval

- Rumus

$$\overline{X} = \frac{f_1\overline{X}_1 + f_2\overline{X}_2 + \ldots + f_n\overline{X}_n}{f_1 + f_2 + \ldots f_n}$$

$$= \frac{\sum_{i=1}^{n} f_i\overline{X}_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i\overline{X}_i}{n}$$

- Keterangan

$\overline{X}_i = nilai\ tengah\ kelas\ i$

$f_i = frekuensi\ \ kelas\ \ i$

Penghasilan	f_i	x_i	$f_i x_i$	$f_i x_i^2$
19,5 – 29,5	7	24,5	171,5	
29,5 – 39,5	9	34,5	310,5	
39,5 – 49,5	16	44,5	712,5	
49,5 – 59,5	21	54,5	1144,5	
59,5 – 69,5	14	64,5	903,5	
69,5 – 79,5	9	74,5	670,5	
79,5 – 89,5	4	84,5	338,0	
89,5 – 99,5	3	94,5	283,5	
99,5 – 109,5	1	104,5	104,5	
JUMLAH	84		4638,0	

x_i = titik tengah interval ke-i

Perbandingan

- Menggunakan rata-rata data interval

$$\overline{X} = \frac{\sum_{i=1}^{9} f_i x_i}{\sum_{i=1}^{9} f_i} = \frac{4638,0}{84} = 52,21$$

- Menggunakan rata-rata data tunggal

$$\overline{X} = \frac{67 + 52 + 72 + \ldots + 42 + 51}{84}$$

$$= 52,21$$

Ukuran Dispersi Data

- Deviasi rata-rata
  - Rata-rata sebaran data terhadap mean

$$\text{mean}: \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\text{d.r.} = \frac{1}{n}\sum_{i=1}^{n} \left|X_i - \bar{X}\right|$$

X_i	$(X_i - \bar{X})$	$\left\|X_i - \bar{X}\right\|$
200	-150	150
275	-75	75
300	-50	50
450	100	100
525	175	175
	0	550

$$\Rightarrow \text{d.r} = \frac{550}{5} = 110$$

Ukuran Dispersi

- Variansi

$$S^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \ldots + (X_n - \overline{X})^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} \; ; \textit{Var Sampel}$$

$$\sigma^2 = \sigma_x^2 = \frac{\sum_{i=1}^{N} (X_i - \mu_x)^2}{N}$$

Variansi Populasi

- Standar Deviasi
  - Akar dari Variansi
  - Dimensi sama dengan dimensi rata-rata sehingga bisa digunakan dengan lebih tepat
  - Rumus di samping untuk data tunggal

Contoh

$\Rightarrow$ Variansi : $s^2 = \dfrac{71250}{4}$

$= 17812,5$

$\Rightarrow$ Standar Deviasi

$s = 133,46$

X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
200	-150	150^2
275	-75	75^2
300	-50	50^2
450	100	100^2
525	175	175^2
	0	71250

Data Interval

- Rumus

$$s^2 = \frac{1}{n}\sum_{i=1}^{k} f_i\left(X_i - \overline{X}\right)^2$$

$$= \frac{n\sum_{i=1}^{k} f_i X_i^{\,2} - \left(\sum_{i=1}^{k} f_i X_i\right)^2}{n\,(n-1)}$$

$$\overline{X} = \frac{\Sigma f_i x_i}{n}, \qquad n = \Sigma f_i$$

Penghasilan	f_i	X_i	$\lvert X_i - \bar{X} \rvert$	$f_i \lvert X_i - \bar{X} \rvert$
19,5 – 29,5	7	24,5	30,71	214,97
29,5 – 39,5	9	34,5	20,71	186,39
39,5 – 49,5	16	44,5	10,71	171,36
49,5 – 59,5	21	54,5	0,71	14,91
59,5 – 69,5	14	64,5	9,29	130,06
69,5 – 79,5	9	74,5	19,29	173,61
79,5 – 89,5	4	84,5	29,29	117,16
89,5 – 99,5	3	94,5	39,29	117,87
99,5 – 109,5	1	104,5	49,29	49,29
Jumlah	84			1175,62

$$\overline{X} = \frac{\sum_{i=1}^{9} f_i X_i}{\sum_{i=1}^{9} f_i} = \frac{4638}{84} = 55,21$$

$$\Rightarrow \; s^2 = \frac{84 \times 283441,00 - (4638,0)^2}{83 \times 84}$$

$$= 329,6041$$

$$\Rightarrow \; s = \sqrt{329,6041} = 18,155$$

Tugas 1

- Data Kuantitatif
  - Kepala Sekolah SMA Maju berkeinginan melihat gambaran yang lebih jelas tentang distribusi penghasilan orang tua siswa. Untuk itu diambil 50 orang tua siswa sebagai sampel, kemudian dicatat penghasilan per bulannya (dalam puluhan ribu rupiah). Berikut hasilnya:

91	78	93	57	75	52	99	80	97	62
71	69	72	89	66	75	79	75	72	76
104	74	62	68	97	105	77	65	80	109
85	97	88	68	83	68	71	69	67	74
62	82	98	101	79	105	79	69	62	73

Buatlah : Distribusi frekeuensinya, histogram, ogive, dan rata-rata (mean). Coba saudara buat interpretasi dari data penghasilan orang tua tersebut di atas.

Permutasi dan Kombinasi

Permutasi Dan Kombinasi

- Faktorial

  Hasil kali semua bilangan bulat dari 1 hingga n
- Permutasi

  Penyusunan obyek ke dalam urutan tertentu.
- Kombinasi

  Penyusunan obyek tanpa memperhatikan urutan
- Koefisien Binomial

Contoh Permutasi

Tentukan jumlah Urutan yang mungkin jika **M**urid-**G**uru-**K**aryawan harus berbaris!

➔ Solusi:

MGK,
MKG,
GKM,
GMK,
KMG,
KGM.

Terdapat 6 Urutan

Posisi 1: ada 3 pilihan (M, G atau K)

Posisi 2: ada 2 pilihan (satu ketgori sudah dipakai di posisi 1)

Posisi 3: ada 1 pilihan (dua ketgori sudah dipakai di posisi 1 dan 2)

Jml Urutan

= **3 x 2 x 1**

= **3!**

= 6

Permutasi n obyek tanpa Pengembalian

A. Seluruhnya

Contoh:

Terdapat 4 macam buku statistis, 3 macam buku pemrograman dan 2 buku hardware. Ada berapa cara menyusun buku-buku tsb?

Solusi:

a. 4 Buku statistik ➔ 4P4 = 4! = 4 x 3 x 2 x 1 = 24 cara
b. 3 buku pemrograman ➔ 3P3 = 3! = 6 cara
c. 2 buku hardware ➔ 2P2 = 2! = 2 cara
d. Ketiga kelompok buku ➔ 3P3 = 3! = 6 cara
e. Seluruh buku = 24 x 6 x 2 x 6 = 1.728 cara

Permutasi n obyek tanpa Pengembalian

B. Sebagian

$$n\,\mathrm{Pr} = \frac{n!}{(n-r)!}$$

Contoh: Dari 6 orang pendiri suatu Partai, akan dipilih Ketua, Wakil Ketua, Sekretaris dan Bendahara. Ada berapa macam kemungkinan susunan struktur Pengurus Partai tersebut?

Solusi: n = 6

r = 4

Jumlah permutasi yang mungkin sebanyak

$${}_6P_4 = \frac{6!}{(6-4)!} = \frac{6x5x4x3x2x1}{2x1} = 360$$

Permutasi n obyek tanpa Pengembalian

C. Melingkar

$$P = (n - 1)!$$

Contoh:

Enam orang duduk mengelilingi meja bundar. Ada berapa kemungkinan urutan keenam orang tersebut?

Solusi: n = 6

$P = (n - 1)!$

$= 5! = 5 \times 4 \times 3 \times 2 \times 1$

= 120 cara

Permutasi n Obyek Dengan Pengembalian

$$_{n}P_{r} = n^{r}$$

Contoh:

Tentukan permutasi dari ABC sebanyak 2 unsur dengan pengembalian unsur yang terpilih

Solusi:

n = 3

r = 2

3P2 = n^r

= 3^2

= 9

AA, AB AC

BB, BA, BC

CC, CA, CB

Permutasi dari n obyek dengan perulangan

$$nPn_1, n_2, n_3, \ldots, n_k = \frac{n!}{n_1! n_2! n_3! \ldots n_k!}$$

Contoh:

Tentukan permutasi dari huruf-huruf "**STATISTIK**"

Solusi

n = 9

S ➔ $n_1 = 2$ ➔ $n_1! = 2$

T ➔ $n_2 = 3$ ➔ $n_2! = 6$

I ➔ $n_3 = 2$ ➔ $n_3! = 2$

$$9P2,3,2 = \frac{9!}{2!.3!.2!}$$

$$= \frac{362880}{2x6x2} = 15120$$

KOMBINASI

$$c_r^n = \frac{n!}{r!(n-r)!} \quad n \geq r$$

Contoh:

Dari 4 orang (***ABCD***) pendiri suatu Partai, akan dipilih Ketua, Wakil Ketua, dan Sekretaris. Ada berapa macam urutan pengurus partai tersebut yang mungkin terpilih?

Solusi $\Rightarrow$ $$C_3^4 = \frac{4!}{3!(4-3)!} = \frac{24}{6(1)} = 4$$

n = 4; r = 3

Urutan yang mungkin adalah:

ABC	ABD
ACD	BCD

DISTRIBUSI PROBABILITAS DISKRIT

Distribusi Uniform Diskrit

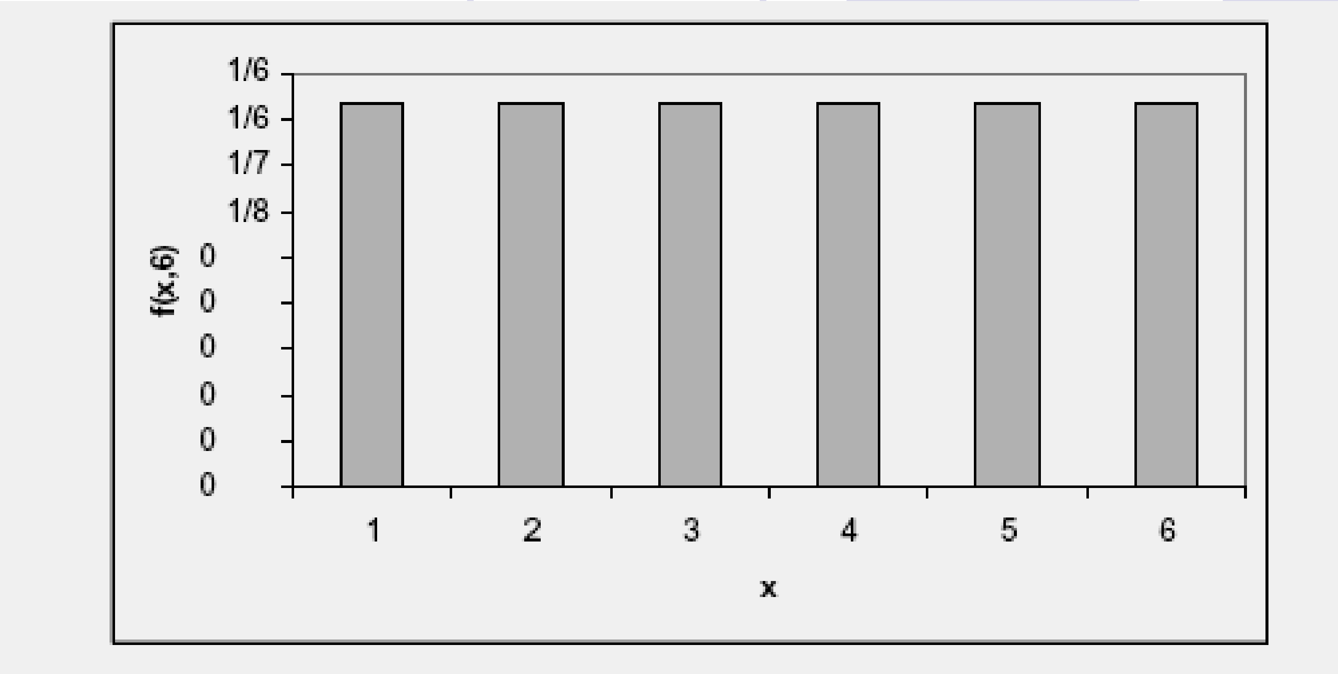
Bila variabel acak *X* memiliki nilai *x*1, *x*2, ... *x*k dengan probabilitas yang sama,maka **distribusi uniform diskrit** dinyatakan sebagai:

$$f(x,k) = \frac{1}{k}; \qquad x = x_1, x_1, \ldots\ldots x_k \qquad k = \text{parameter}$$

Bila lampu pijar dipilih secara acak dari suatu kotak yang berisi lampu pijar 40 watt, 60 watt, 75 watt, 100 watt, dengan probabilitas masingmasing ¼. Maka distribusi uniformnya adalah:

$$f(x,4)=\frac{1}{4}; \qquad x = 40,60,75,100$$

Secara grafis (histogram) ditunjukkan pada gambar berikut:



Rata-rata dan varians dari suatu distribusi uniform diskrit $f(x;k)$ adalah:

$$\mu = \frac{\sum_{i=1}^{k} x_i}{k} \qquad \text{dan} \qquad \sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k}$$

contoh 5.3: rata-rata dan varians dari soal 5.2 adalah:

$$\mu = \frac{1+2+3+4+5+6}{6} = 3.5$$

$$\sigma^2 = \frac{(1-3.5)^2 + (2-3.5)^2 + \dots\dots\dots + (6-3.5)^2}{6} = \frac{35}{12}$$

Distribusi Binomial dan Multinomial

Suatu percobaan binomial adalah suatu percobaan yang memiliki sifat-sifat berikut:

1. Percobaan terdiri atas n percobaan yang identik.
2. Setiap hasil percobaan dapat diklasifikasikan sebagai sukses atau gagal
3. Probabilitas suskes, dinyatakan dengan p, dan kegagalan dengan $q = 1 - p$.
4. Percobaan-percobaan yang dilakukan tidak saling bergantung (independent).

Distribusi Binomial

Distribusi probabilitas dari vaiabel acak binomial *X*:

$$P(X=x)=\binom{n}{x}p^x q^{n-x} \qquad x=0,1,2,3,\ldots$$

Probabilitas suatu alat tertentu akan tetap bertahan (tidak rusak) bila digetarkan adalah ¾. Tentukan probabilitas bahwa terdapat 2 dari 4 kompenen yang ditest akan bertahan.

Solusi:

$$P(X=2)=\binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2=\frac{4!}{2!2!}.\frac{3^2}{4^2}=\frac{27}{128}$$

Dalam perencanaan sistem pengendalian banjir suatu sungai, banjirtahunan maksimum adalah hal yg harus diperhatikan. Bila probabilitas dari banjirmaksimum tahunan melebihi ketinggian desain tertentu *h0* adalah 0,1. Berapaprobabilitas bahwa ketinggian *h0* akan terlampaui satu kali dalam 5 tahunmendatang:

Solusi:

$$P(X=1)=\binom{5}{1}(0.1)^1(0.9)^4=0.328$$

Rata-rata dan varians dari distribusi binomial adalah:

$$\mu = np \qquad \text{dan} \qquad \sigma^2 = npq$$

Distribusi Multinomial

Bila dari suatu percobaan diperoleh hasil *E1*, *E2*, ...*Ek*, dengan probabilitas *p1*,*p2*,.... *pk*, maka distribusi probabilitas dari suatu variabel acak *X1*, *X2*,... *Xk*, dalam *n* kali percobaan adalah:

$$f(x_1, x_2, \ldots\ldots\ldots x_k; p_1, p_2, \ldots\ldots\ldots p_k, n) = \binom{n}{x_1, x_2, \ldots x_k} p_1^{x_1}, p_2^{x_2}, \ldots p_k^{x_k}$$

dengan $\sum_{i=1}^{k} x_i = n$ dan $\sum_{i=1}^{k} p_i = 1$

Distribusi Geometrik

Bila percobaan yang berulang secara independent menghasilkan kesuksesan dengan probabilitas *p* dan gagal dengan probabilitas *q* = 1 – *p*. Maka distribusi probabilitas variabel acak *X* pada saat terjadinya kesuksesan pertama adalah:

$$P(X = x) = pq^{x-1}, \qquad x = 1,2,3,\ldots..$$

Pada suatu proses pembuatan alat tertentu diketahui bahwa setiap 100 item ada 1 yang rusak. Berapa probabilitas bahwa item ke 5 yang diawasi adalah yang pertama rusak.

$$P(X=5)=(0.01)(0.99)^4=0.0096$$

Rata-rata dan varians dari suatu variabel acak dengan distribusi geometrik

$$\mu=\frac{1}{p}, \qquad \sigma^2=\frac{1-p}{p^2}$$

Distribusi Binomial negatif

Bila percobaan yang berulang secara independent menghasilkan kesuksesandengan probabilitas p dan gagal dengan probabilitas $q = 1 - p$. Maka distribusi probabilitas variabel acak X pada saat terjadinya kesuksesan yang ke - k adalah:

$$P(X_k = x) = \binom{x-1}{k-1} p^k q^{x-k}, \qquad x = k, k+1, k+2, \ldots$$

Contoh:

Anggap suatu kabel terdiri dari beberapa kawat yang terususn secara independent.Kadang-kadang kabel tersebut dibebani dengan beban berlebih; pada saat itu probabilitas bahwa ada 1 kawat yang putus adalah 0.05. Asumsikan bahwa kegagalan 2 atau lebih kawat tidak sama. Kabel harus diganti bila 3 kawat sudah putus, tentukan probabilitas bahwa kabel dapat bertahan pada saat dibebani dengan beban berlebih paling tidak 5 kali sebelum kabel tersebut diganti :

Solusi:

$$P(X_k \geq 6) = 1 - P(X_3 < 6)$$

$$= 1 - \binom{2}{2}(0.05)^3 - \binom{3}{2}(0.05)^3(0.95) - \binom{4}{2}(0.05)^3(0.95)^2 = 0.99884$$

Proses Poisson dan Distribusi Poisson

Banyak masalah yang menjadi perhatian seorang insinyur adalah mengetahui kemungkinan terjadinya suatu peristiwa pada interval waktu tertentu. Contoh :
gempa dapat terjadi pada waktu tertentu, kecelakaan lalu lintas dapat terjadi pada rentan waktu tertentu di suatu jalan raya. Dalam kasus seperti ini kejadian suatu peristiwa lebih tepat bila dimodelkan dengan **Proses Poisson**. Asumsi proses poisson adalah sebagai berikut :

- Suatu peristiwa dapat terjadi secara acak dan pada interval waktu tertentu.
- Kejadian satu peristiwa dengan peristiwa lain pada interval waktu tertentu adalah independen (bebas).
- Probabilitas kejadian suatu peristiwa pada interval waktu Δt adalah proporsional terhadap Δt, dan dapat diberikan dengan $v\Delta t$, dimana v adalah rata-rata kejadian suatu peristiwa

Berdasarkan asumsi diatas, disribusi Poissson dinyatakan dengan rumus berikut :

$$P(X_t = x) = \frac{e^{-vt}(vt)^x}{x!}, \qquad x = 0,1,2,\ldots$$

Berdasarkan data, badai hujan di suatu kota selama 20 tahun, menunjukkan bahwa rata-rata terdapat 4 kali badai hujan per tahun. Asumsikan kejadian badai hujan adalah proses Poisson, berapa probabilitas bahwa tidak ada badai hujan tahun depan :

$$P(X_t = 0) = \frac{e^{-4}(4)^0}{0!} = 0.018$$

Probabilitas akan terjadi 4 kali badai hujan tahun depan adalah

$$P(X_t = 4) = \frac{e^{-4}(4)^4}{4!} = 0.195$$

Probabilitas akan terjadi 2 kali atau lebih badai hujan tahun depan adalah

$$P(X_t \geq 2) = 1 - \sum_{x=0}^{1} \frac{e^{-4}(4)^x}{x!} = 1 - 0.018 - 0.074 = 0.908$$

DISTRIBUSI SAMPEL RANDOM

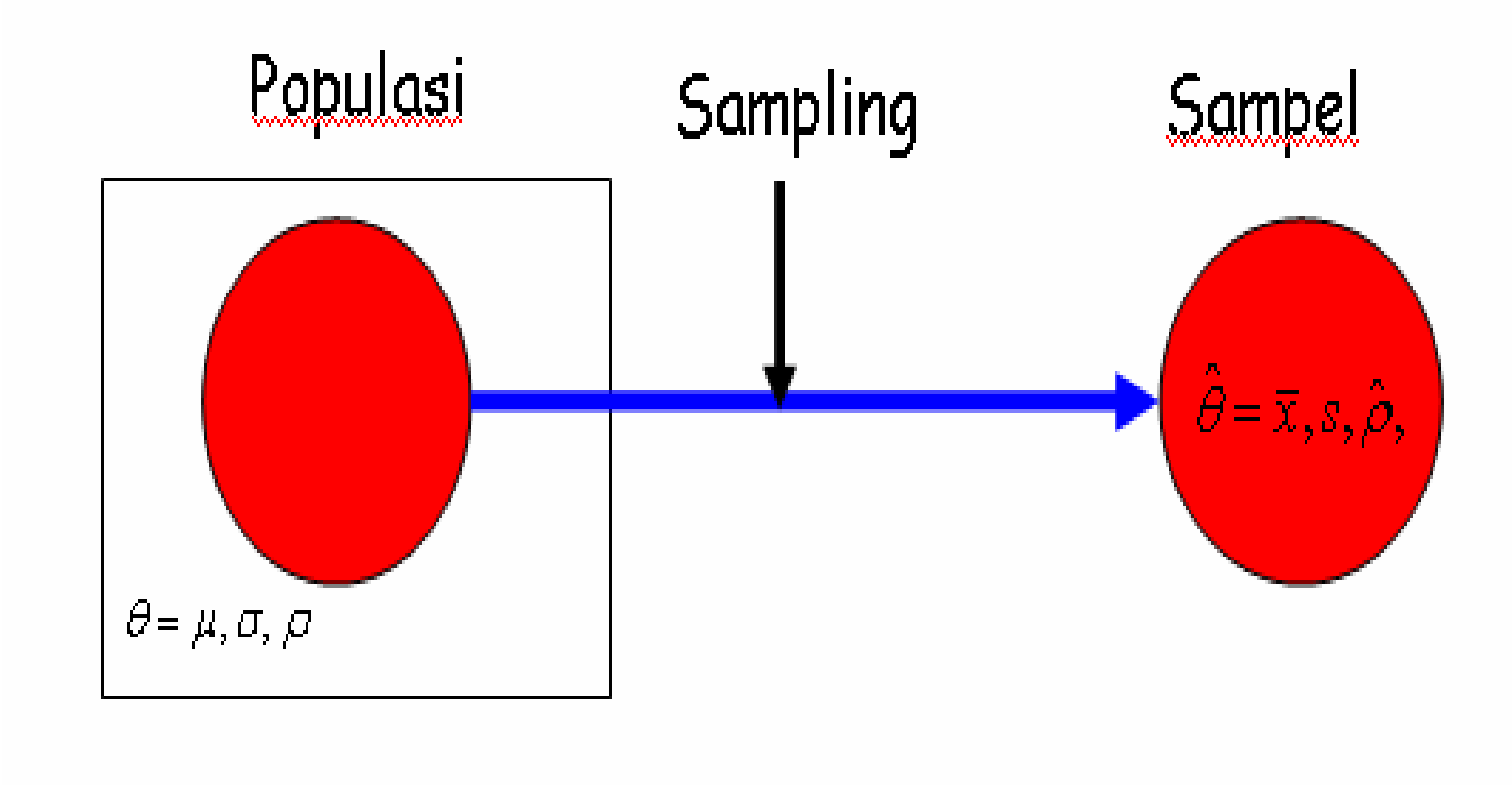
AGENDA

- Pengertian dan Konsep Dasar
- Distribusi Sampel Rata-rata
- Distribusi Sampel Proporsi
- Distribusi Sampel Beda Dua Rata-rata
- Distribusi Sampel Beda Dua Proporsi

1. Pengertian dan Konsep Dasar

- Populasi adalah banyaknya pengamatan
- Dua jenis populasi menurut ukurannya: terbatas (berhingga) dan tak terbatas (tak berhingga)
- Sifat/ciri dalam populasi disebut karakteristik populasi, hasil pengukuran karakteristik populasi disebut parameter populasi
- Sensus adalah cara untuk mengumpulkan data populasi
- Kelemahan sensus: biaya mahal, waktu lama, tenaga yang besar
- Kelemahan sensus diatasi dengan teknik sampel (sampling)
- Karakteristik sampel disebut statistik
- Keuntungan teknik sampel adalah biaya yang rendah serta waktu yang pendek tanpa mengurangi keakuratan

1. Pengertian dan Konsep Dasar



Gambar 1. Hubungan populasi dan sampel

1. Pengertian dan Konsep Dasar

Tabel 8.1. Karakteristik Populasi dan Sampel

No.	Karakteristik Populasi	Karakteristik Sampel
1.	Ukuran N	Ukuran n
2.	Parameter	Statistik
3.	Mean, μ	Rata-rata (mean), $\overline{X}$
4.	Standar deviasi, σ	Standar deviasi, S
5.	Proporsi, p	Proporsi, $\hat{p}$
6.	Populasi terbatas dan tak terbatas	Sampel besar dan kecil

1. Pengertian dan Konsep Dasar

- Terdapat gap antara populasi dan sampel yang disebut sebagai kesalahan (penyimpangan)
- Sebab kesalahan sampel: kesalahan pemilihan sampel, kesalahan hitung, dll
- Sampel yang representatif memiliki ciri: ukuran tertentu yang memakai syarat, kesalahan terkecil, dan dipilih dengan prosedur yang benar berdasarkan teknik sampel tertentu

1. Pengertian dan Konsep Dasar

- Teknik sampel acak sederhana
  - Setiap unit dalam populasi memiliki kesempatan yang sama terambil
  - Setiap ukuran sampel n mempunyai kesempatan yang sama terambil
  - Populasi bersifat uniform atau seragam
  - Sesuai untuk populasi yang kecil
  - Menggunakan tabel bilangan acak
- Teknik sampel acak sistematik
  - Mengambil setiap unsur ke-k dalam populasi

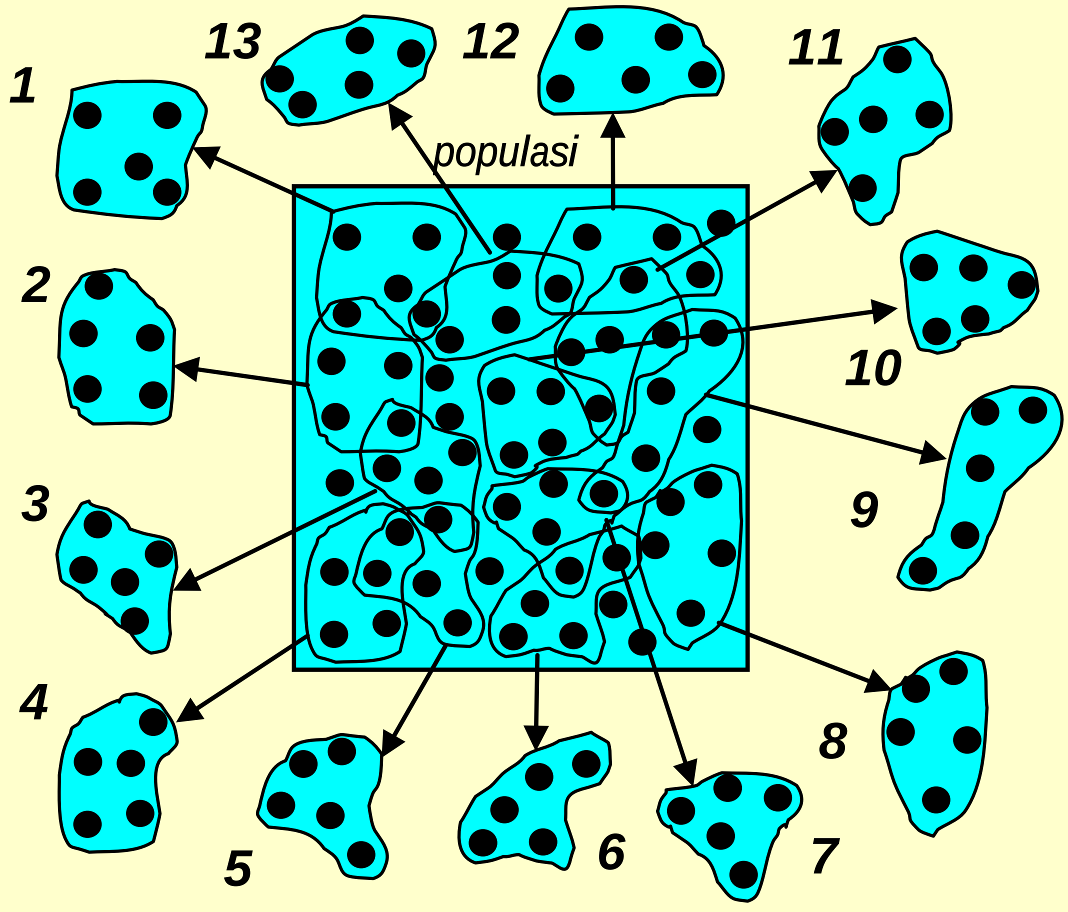
1. Pengertian dan Konsep Dasar

- Teknik sampel acak stratifikasi
  - Membagi populasi atas beberapa kelompok (strata) sehingga setiap kelompok menjadi uniform
  - Alokasi sebanding: mengambil sampel pada masing-masing kelompok populasi yang sebanding dengan ukuran populasi
- Teknik sampel acak cluster
  - Mengambil beberapa cluster
  - Sebagian atau seluruh unit dalam cluster sebagai sampel diambil secara acak

1. Pengertian dan Konsep Dasar

- Pengambilan sampel dengan pengembalian, maka ukuran populasi adalah tetap. Sesuai untuk ukuran populasi terbatas
- Pengambilan sampel dengan tanpa pengembalian maka ukuran populasi akan berkurang. Sesuai untuk populasi tak terbatas
- Distribusi sampel: statistik sampel yang diperoleh bersifat acak (variabel acak) yang mengikuti suatu distribusi tertentu

1. Pengertian dan Konsep Dasar



sampel	mean	Std.dev
1	$\bar{x}_1$	s_1
2	$\bar{x}_2$	s_2
3	$\bar{x}_3$	s_3
4	$\bar{x}_4$	s_4
5	$\bar{x}_5$	s_5
6	$\bar{x}_6$	s_6
7	$\bar{x}_7$	s_7
8	$\bar{x}_8$	s_8
9	$\bar{x}_9$	s_9
10	$\bar{x}_{10}$	s_{10}
11	$\bar{x}_{11}$	s_{11}
12	$\bar{x}_{12}$	s_{12}
13	$\bar{x}_{13}$	s_{13}
...	...	...
i	$\bar{x}_i$	s_i

2. Distribusi Sampel Rata-rata

Bila pada populasi terbatas berukuran N dengan rata-rata μ_x dan standar deviasi σ_x diambil sampel berukuran n secara berulang tanpa pengembalian, maka akan didapat distribusi sampel rata-rata yang memiliki rata-rata $\mu_{\bar{x}}$ dan standar deviasi $\sigma_{\bar{x}}$, yaitu

Rata-rata: $\mu_{\bar{x}} = \mu_x$ dan

Standar deviasi: $\sigma_{\bar{x}} = \dfrac{\sigma_X}{\sqrt{n}}\sqrt{\dfrac{N-n}{n-1}}$ populasi terbatas

$\sigma_{\bar{x}} = \dfrac{\sigma_X}{\sqrt{n}}$ populasi tak terbatas.

Bila pada populasi berukuran N baik terbatas dan tak terbatas dengan mean μ_x dan standar deviasi σ_x, maka distribusi rata-rata $\bar{X}$ akan mendekati distribusi normal dengan rata-rata $\mu_{\bar{X}}$ dan setandar deviasi $\sigma_{\bar{X}}$, sehingga variabel acak Z, diberikan oleh:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

Dimana $\sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}}$ bila populasi tak terbatas dan $\sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}$ bila populasi terbatas.

- Sampel 2, 3, 6, 8 dan 11 dapat dihitung mean populasi, deviasi standard populasi, mean dari distribusi penarikan sampel, dan deviasi standard distribusi mean penarikan sampel sebagai berikut dengan menggunakan definisi-definisi dasar statistik deskriptif sebagai berikut:

Populasi :

$$\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6,0$$

$$\sigma = \sqrt{\frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2+}{5}} = 3,29$$

Sampel :

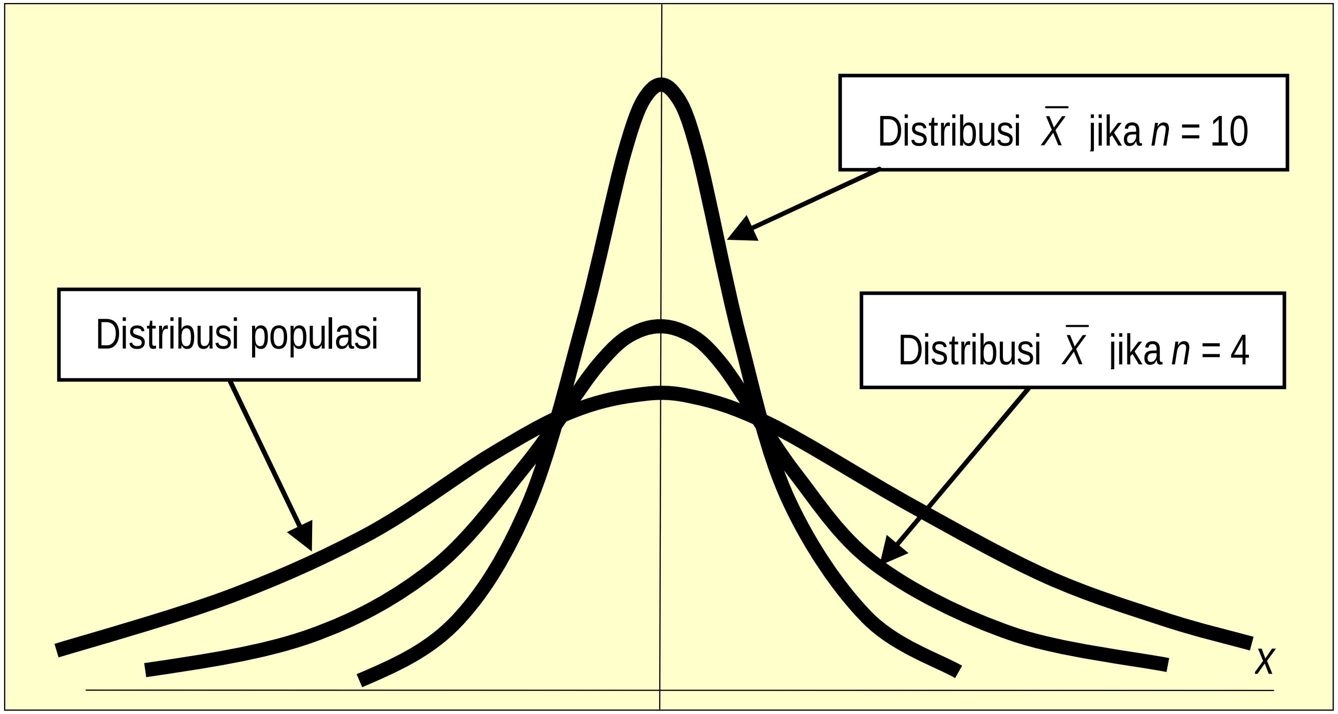
$$\mu_{\bar{x}} = \frac{\sum_{i=1}^{14} f_i \bar{x}_i}{\sum_{i=1}^{14} f_i} = \frac{(1)(2,5)+(1)(4)+\ldots+(1)(9,5)}{1+1+\ldots+1} = \frac{60}{10} = 6,0$$

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{14} f_i(\bar{x}_i - \mu_{\bar{x}})^2}{\sum_{i=1}^{14} f_i}} = \sqrt{\frac{(1)(2,5-6)^2+(1)(4-6)+\ldots+(1)(9,5-6)}{10}} = 2,01$$

Terlihat bahwa $\mu_{\bar{x}} = \mu = 6,0$ dan dengan n = 2, N = 5 dapat ditunjukkan:

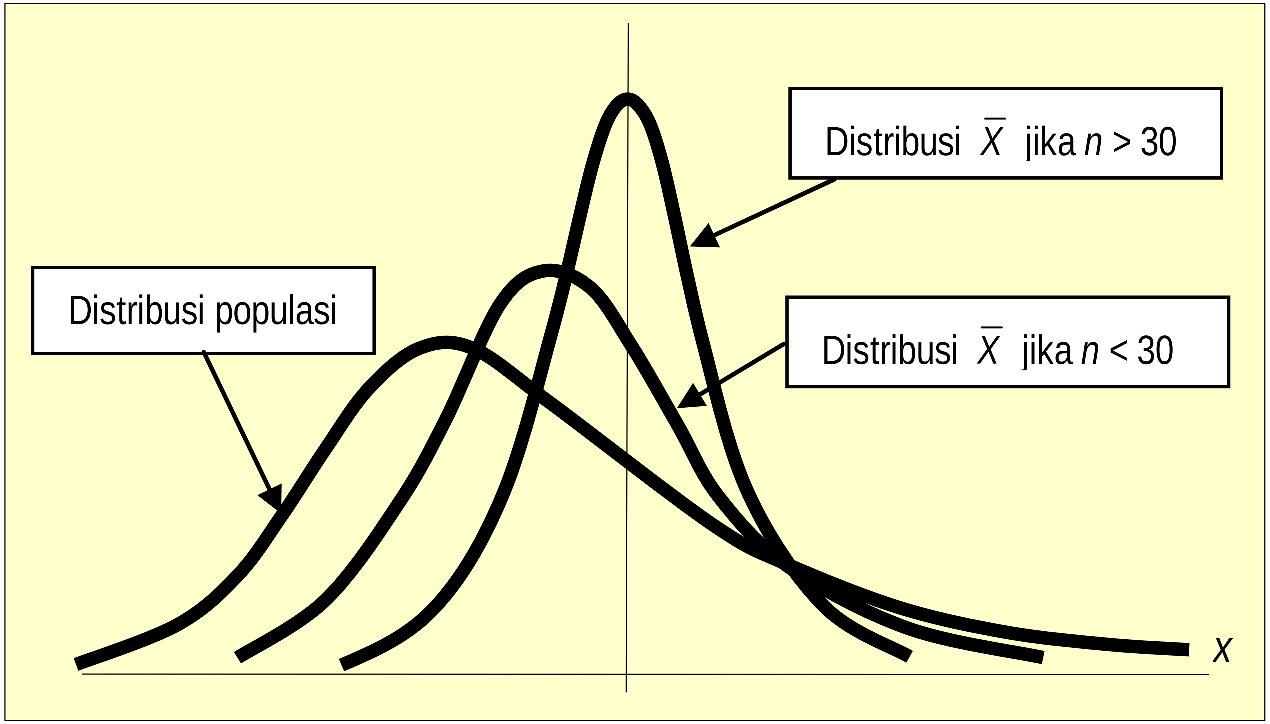
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{3,29}{\sqrt{2}}\sqrt{\frac{(5-2)}{5-1}} = 2,01$$

2. Distribusi Sampel Rata-rata



Gambar 2. Distribusi sampel rata-rata pada populasi terdistribusi normal

2. Distribusi Sampel Rata-rata



Gambar 3. Ilustrasi teorema limit pusat (central limit theorem)

- Lima ratus cetakan logam memilki berat rata-rata 5,02 N dan deviasi standard 0,30 N. Probabilitas bahwa suatu sampel acak terdiri dari 100 cetakan yang dipilih akan mempunyai berat total antara 496 sampai 500 N dapat ditentukan dengan sebagai berikut:

Distribusi mean penarikan sampel persoalah diatas memiliki mean dan deviasi standard:

$$\mu_{\bar{x}} = \mu = 5{,}02 \text{ N}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{0{,}3}{\sqrt{100}}\sqrt{\frac{500-100}{500-1}} = 0{,}027$$

Jika seratus sampel cetakan memiliki berat total 496 sampai 500 N maka meannya adalah 4,96 sampai 5,00 N. Dengan mengingat kembali konsep-konsep distribusi normal yang telah dibahas, probabilitas mean tersebut dapat dicari dengan menggunakan tabel distribusi normal standard dimana skor *z* (*z score*)nya didefinisikan sebagai:

$$z_{\bar{x}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

maka:

$$\begin{aligned}
&P(4{,}96 \le \bar{X} \le 5{,}00) \\
&= P\left(\frac{4{,}96-5{,}02}{0{,}027} \le Z_{\bar{x}} \le \frac{5{,}00-5{,}00}{0{,}027}\right) \\
&= P\left(-2{,}22 \le Z_{\bar{x}} \le -0{,}74\right) \\
&= \Phi\left(-0{,}74\right) - \Phi\left(-2{,}22\right) \\
&= 0{,}2296 - 0{,}0132 = 0{,}2164 = 21{,}64\%
\end{aligned}$$

3. Distribusi Sampel Proporsi

Dalam suatu populasi berukuran N terdapat jenis tertentu dengan proporsi $p = \frac{X}{N}$ dan pada populasi tersebut diambil sampel berukuran n yang mengandung jenis tertentu dengan proporsi $\hat{p} = \frac{x}{n}$, maka statistik $\hat{p} = \frac{x}{n}$ yang bersifat acak sehingga mempunyai suatu distribusi yang disebut distribusi sampel proporsi dengan mean dan standar deviasi sebagai berikut.

Rata-rata: $\mu_{\hat{p}} = \mu_p = \frac{X}{N}$ dan

Standar deviasi: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{n-1}}$ populasi terbatas

$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ populasi tak terbatas.

Untuk sampel besar, distribusi sampel proporsi merupakan distribusi normal hingga variabel Z diberikan oleh

$$Z = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$$

3. Distribusi Sampel Proporsi

- Divisi pengendalian mutu pabrik perkakas mesin mencatat bahwa 2 % dari mata bor yang diproduksi mengalami cacat. Jika dalam pengiriman satu batch produk terdiri dari 400 mata bor, tentukan probabilitas banyaknya mata bor yang cacat 3 % atau lebih?

  Distribusi proporsi penarikan sampel persoalan diatas memiliki mean dan deviasi standard :

$$\mu_P = \pi = 0{,}02$$

$$\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0{,}02(1-0{,}02)}{400}} = 0{,}007$$

Faktor koreksi variabel diskrit = 1/(2*n*)= 1/800 = 0,00125

Proporsi (3 %) setelah dikoreksi, *P* = 0,03 - 0,00125 = 0.02875

Maka probabilitas mata bor yang cacat dengan proporsi lebih dari 3 % adalah

$$\begin{aligned} P(P > 0{,}02875) &= 1 - P(P \leq 0{,}02875) \\ &= 1 - P\left(Z_P \leq \frac{0{,}02875 - 0{,}02}{0{,}007}\right) \\ &= 1 - P(Z_P \leq 1{,}25) = 1 - \Phi(1{,}25) \\ &= 1 - 0{,}8944 = 0{,}1056 = 10{,}56\% \end{aligned}$$

4. Distribusi Sampel Beda Dua Rata-rata

Misal dua populasi, populasi I dan II. Populasi I sebanyak N_1 memiliki mean μ_1 dan standar deviasi σ_1, populasi II sebanyak N_2 memiliki mean μ_2 dan standar deviasi σ_2. Dari populasi I diambil sampel acak sebanyak n_1 dan memiliki rata-rata $\bar{X}_1$, dari populasi II diambil sampel acak sebanyak n_2 dan memiliki rata-rata $\bar{X}_2$. Kedua sampel diasumsikan saling bebas. Bila dari kedua sampel disampel secara acak maka akan didapat distribusi sampel beda dua rata-rata $\left(\bar{X}_1 - \bar{X}_2\right)$. Rata-rata dan standar deviasi beda dua rata-rata $\left(\bar{X}_1 - \bar{X}_2\right)$ diberikan oleh

Rata-rata: $\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$ dan

Standar deviasi: $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}} \sqrt{\dfrac{(N_1 + N_2) - (n_1 + n_2)}{(N_1 + N_2) - 1}}$ populasi terbatas

$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ populasi tak terbatas.

Untuk sampel besar, distribusi sampel beda dua rata-rata merupakan distribusi normal hingga variabel Z diberikan oleh

$$Z = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}}$$

- Lampu bohlam produksi perusahaan A memiliki daya tahan pakai rata-rata 1400 jam dan deviasi standard 200 jam, sementara yang diproduksi perusahaan B memiliki daya tahan pakai rata-rata 1200 jam dengan deviasi standard 100 jam. Jika dari masing-masing produk dipilih 125 bohlam sebagai sampel yang diuji, maka probabilitas bahwa bohlam produksi A memiliki daya tahan pakai sekurang-kurangnya 160 jam lebih lama dibandingkan bohlam produksi B dapat ditentukan sebagai berikut.

Statistik yang dibicarakan dalam persoalan ini adalah mean dari daya tahan pakai bohlam A dan B ($\overline{x}_A$ dan $\overline{x}_B$) yang akan ditentukan perbedaannya $(\overline{x}_A - \overline{x}_B)$.

Maka mean dari distribusi perbedaan penarikan sampel daya tahan pakai bohlam A dan B:

$$\mu_{\overline{x}_A - \overline{x}_B} = \mu_{\overline{x}_A} - \mu_{\overline{x}_B} = \mu_{x_A} - \mu_{x_B} = 1400 - 1200 = 200$$

Deviasi standardnya adalah:

$$\sigma_{\overline{x}_A - \overline{x}_B} = \sqrt{\sigma^2_{\overline{x}_A} + \sigma^2_{\overline{x}_B}} = \sqrt{\frac{\sigma^2_{x_A}}{n_A} + \frac{\sigma^2_{x_B}}{n_B}} = \sqrt{\frac{(100)^2}{125} + \frac{(200)^2}{125}} = 20$$

Skor z untuk perbedaan mean adalah:

$$z_{\overline{x}_A - \overline{x}_B} = \frac{(\overline{x}_A - \overline{x}_B) - (\mu_{\overline{x}_A - \overline{x}_B})}{\sigma_{\overline{x}_A - \overline{x}_B}} = \frac{(\overline{x}_A - \overline{x}_B) - 200}{20}$$

Skor z untuk perbedaan mean 160 jam adalah:

$$z_{\overline{x}_A - \overline{x}_B} = \frac{(\overline{x}_A - \overline{x}_B) - 200}{20} = \frac{160 - 200}{20} = -2$$

Jadi probabilitas yang akan ditentukan adalah:

$$\begin{aligned} P((\overline{x}_A - \overline{x}_B) > 160) = P(z_{\overline{x}_A - \overline{x}_B} > -2) &= 1 - P(z_{\overline{x}_A - \overline{x}_B} < -2) \\ &= 1 - 0{,}0228 = 0{,}9772 = 97{,}72\% \end{aligned}$$

4. Distribusi Sampel Beda Dua Proporsi

Misal dua populasi, populasi I dan II. Populasi I sebanyak N_1 mengandung jenis X_1 dengan proporsi $p_1 = \frac{X_1}{N_1}$ dan Populasi II sebanyak N_2 mengandung jenis X_2 dengan proporsi $p_2 = \frac{X_2}{N_2}$. Dari kedua populasi diambil sampel acak sebanyak n_1 dan n_2, maka sampel I akan mengandung jenis X_1 dengan proporsi $\hat{p}_1 = \frac{X_1}{n_1}$ dan sampel II akan mengandung jenis X_2 dengan proporsi $\hat{p}_2 = \frac{X_2}{n_2}$. Bila kedua sampel diambil secara acak dan saling bebas, maka diperoleh distribusi sampel beda dua proporsi $(\hat{p}_1 - \hat{p}_2)$. Ditribusi sampel beda dua proporsi $(\hat{p}_1 - \hat{p}_2)$ memiliki rata-rata dan standar deviasi sebagai berikut:

Rata-rata: $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$ dan

Standar deviasi: $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \sqrt{\frac{(N_1 + N_2) - (n_1 + n_2)}{(N_1 + N_2) - 1}}$ populasi terbatas

$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$ populasi tak terbatas.

Distribusi sampel beda dua proporsi mempunyai distribusi normal dengan statistik Z diberikan oleh:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sigma_{\hat{p}_1 - \hat{p}_2}}$$

Tabel 8.2. Nilai Rata-Rata dan Standar Deviasi dari Distribusi Sampel

No	Distribusi Sampel	Parameter Distribusi	Statistik Distribusi
1.	**Rata-rata**	Rata-rata: $\mu_{\bar{x}} = \mu_x$ dan Standar deviasi: $\sigma_{\bar{x}} = \frac{\sigma_X}{\sqrt{n}}\sqrt{\frac{N-n}{n-1}}$ populasi terbatas $\sigma_{\bar{x}} = \frac{\sigma_X}{\sqrt{n}}$ populasi tak terbatas.	Untuk sampel besar $(n \geq 30)$ $Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$ Untuk sampel kecil (n < 30) $t = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$
2.	**Proporsi**	Rata-rata: $\mu_{\hat{p}} = \mu_p = \frac{X}{N}$ Standar deviasi: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{n-1}}$ populasi terbatas $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ populasi tak terbatas.	Untuk sampel besar $(n \geq 30)$ $Z = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$

3.	**Beda Dua Rata-Rata**	Rata-rata: $\mu_{\bar{X}_1-\bar{X}_2}=\mu_1-\mu_2$ Standar deviasi: $\sigma_{\bar{X}_1-\bar{X}_2}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\sqrt{\frac{(N_1+N_2)-(n_1+n_2)}{(N_1+N_2)-1}}$ populasi terbatas $\sigma_{\bar{X}_1-\bar{X}_2}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ populasi tak terbatas.	Untuk sampel besar $(n\geq 30)$ $Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{X}_1-\bar{X}_2}}$ Untuk sampel besar $(n<30)$ $t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{X}_1-\bar{X}_2}}$
4.	**Beda Dua Proporsi**	Rata-rata: $\mu_{\hat{p}_1-\hat{p}_2}=p_1-p_2$ Standar deviasi: $\sigma_{\hat{p}_1-\hat{p}_2}=\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}\sqrt{\frac{(N_1+N_2)-(n_1+n_2)}{(N_1+N_2)-1}}$ populasi terbatas $\sigma_{\hat{p}_1-\hat{p}_2}=\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$ populasi tak terbatas.	Untuk sampel besar $(n\geq 30)$ $Z=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sigma_{\hat{p}_1-\hat{p}_2}}$

Besar Sampel pada Estimasi Parameter

Proses Desain Sampel

Define the Target Population

Identify the Sampling Frame

Choose the Sampling Method

Determine the Sample Size

Gather the Data

Konsep

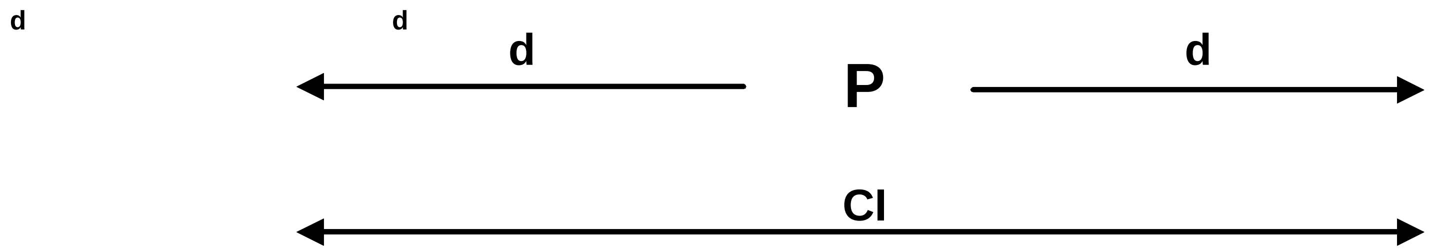
- Desain sampel harus memenuhi kaidah:
  - Validitas
    - Bergantung pada cara pengambilan sampel
    - Bergantung pada kerangka sampel
  - Presisi
    - Bergantung pada jumlah sampel
- Cara pengambilan sampel sering kurang mendapat perhatian dibandingkan besar sampel

Contoh Masalah Kerangka Sampel

Konsep

- Sampel hanya bisa dirancang dan dihitung jika ada informasi awal tentang hal yang diteliti dan populasi
- Secara garis besar desain dan besar sampel dapat dibagi menurut:
    - Estimasi parameter populasi
    - Uji Hipotesis
- Kesalahan yang sering terjadi: selalu menganggap penelitian sebagai estimasi parameter padahal sebenarnya uji hipotesis

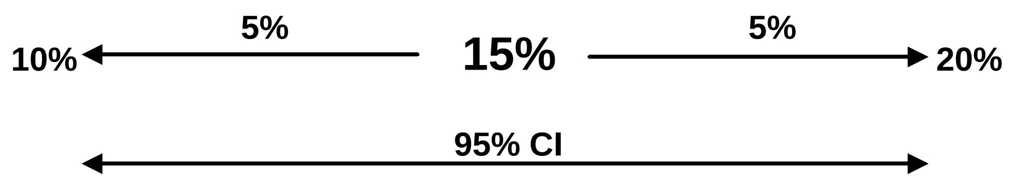
Terminologi pada Perhitunga Besar Sampel u/ Estimasi Parameter



P = Estimasi proporsi
d = Simpangan
CI = Confidence Interval

Contoh

- Diketahui prevalensi diare di Jabar 15%
- Simpangan yang dapat diterima 5%
- Derajat kepercayaan 95%



- Peneliti 95% percaya bahwa prevalensi diare di Jawa Barat berkisar antara 10% sampai dengan 20%

Besar sampel estimasi proporsi, simpangan mutlak

$$n = \frac{z^2_{1-\alpha/2} P(1-P)}{d^2}$$

P=Estimasi proposi

d=simpangan mutlak

z=nilai z pada derajat kepercayaan 1-α/2

Besar sampel estimasi proporsi, simpangan mutlak

$$n = \frac{z^2_{1-\alpha/2} P(1-P)}{d^2}$$

- Digunakan untuk estimasi proposi
- Tidak tepat digunakan untuk uji hipotesis
- Asumsi desain: populasi tak terbatas dan sampel srs

Besar sampel estimasi proporsi, simpangan mutlak

$$n = \frac{z^2_{1-\alpha/2} P(1-P)}{d^2}$$

Contoh penggunaan:

Survei cakupan imunisasi

Survei prevalensi gizi buruk di masyarakat

Penelitian prevalensi hipertensi di masyarakat

Besar sampel estimasi proporsi: Contoh

- Contoh:
  Suatu survei dilakukan untuk mengetahui prevalensi diare pada Balita di Kabupaten Bogor. Berapa jumlah sampel yang diperlukan untuk survei ini?
- Untuk menghitung jumlah sampel,peneliti perlu tahu:
  - Perkiraan prevalensi diare di kab. Bogor
  - Simpangan yang dapat diterima
  - Derajat kepercayaan

Besar sampel estimasi proporsi: Contoh

- Misalkan:
  - Diketahui prevalensi diare di Jabar 15%
  - Simpangan yang dapat diterima 5%
  - Derajat kepercayaan 95%
- Berarti:
  - Peneliti memperkirakan prevalensi diare di kab. Bogor 15%
  - Peneliti 95% yakin bahwa prevalensi diare di kab. Bogor berkisar antara 10-20%
  - Ada 5% kemungkinannya prevalensi diare berada di luar kisaran 10-20%

Besar sampel estimasi proporsi: Contoh

$$n = \frac{1,96^2 * 0,15(1-0,15)}{0,05^2}$$

$$n = 196$$

Besar sampel estimasi proporsi, simpangan relatif

$$n = \frac{z^{2}_{1-\alpha/2}(1-P)}{\varepsilon^{2}P}$$

P=Estimasi proposi

ε=Simpangan relatif

z=nilai z pada derajat kepercayaan 1-α/2

Besar sampel estimasi proporsi, simpangan relatif

- Contoh:
  Suatu penelitian dilakukan untuk mengetahui prevalensi karies dentis pada anak sekolah kelas 6 SD di Jakarta Barat
- Peneliti menggunakan asumsi
  - Prevalensi karies dentis pada anak SD di Indonesia 73%
  - Simpangan relatif 10%
  - Derajat kepercayaan 95%
- Berarti: Peneliti 95% yakin prevalensi karies dentis pada anak kelas 6SD di Jakbar berkisar 65,7-80,3%

Besar sampel estimasi proporsi, simpangan relatif

$$n = \frac{1{,}96^2(1-0{,}73)}{0{,}10^2 * 0{,}73}$$

$$n = 143$$

Besar sampel estimasi rata-rata, simpangan mutlak

$$n = \frac{z_{1-\alpha/2}^{2}\sigma^{2}}{d^{2}}$$

σ=simpang baku

d=simpangan mutlak dari rata-rata

z=nilai z pada derajat kepercayaan 1-α/2

Besar sampel estimasi rata-rata, simpangan mutlak

$$n = \frac{z^2_{1-\alpha/2}\sigma^2}{d^2}$$

- Digunakan untuk estimasi rata-rata
- Tidak tepat digunakan untuk uji hipotesis
- Asumsi desain: populasi tak terbatas dan sampel srs

Besar sampel estimasi rata-rata, simpangan mutlak

- Contoh:
  Suatu penelitian dilakukan untuk mengetahui rata-rata tek. Darah sistolik orang dewasa di Jakarta
- Asumsi yang digunakan:
  - Rata-rata tek. Darah 120 mmHg
  - Simpang baku dari penelitian sebelumnya (referensi) 20 mmHg
  - Simpangan mutlak 10 mmHg
  - Derajat kepercayaan 95%
- Berarti peneliti 95% yakin bahwa rata-rata tek. Darah sistolik di populasi berkisar 110-130 mmHg

Besar sampel estimasi rata-rata, simpangan mutlak

$$n = \frac{1{,}96^2 20^2}{10^2}$$

$$n = 16$$

Besar sampel estimasi rata-rata, simpangan relatif

$$n = \frac{z_{1-\alpha/2}^{2}\sigma^{2}}{\varepsilon^{2}\mu^{2}}$$

σ=simpang baku
ε=simpangan relatif dari rata-rata
μ=rata-rata
z=nilai z pada derajat kepercayaan 1-α/2

Besar sampel estimasi rata-rata, simpangan relatif

Contoh: Penelitian tentang rata-rata Hb pada bumil. Dari referensi: rata-rata 9,8 g/dl, simpang baku 3,3 g/dl. Peneliti menginginkan simpangan relatif 15% dan derajat kepercayaan 95%

$$n = \frac{1{,}96^2 * 3{,}3^2}{0{,}1^2 * 9{,}8^2} = 46$$

Masalah

- Tidak mungkin digunakan srs (misalnya survei) → jumlah dikoreksi dengan design effect
- Estimasi tidak hanya pada satu variabel, misal Survei Kesehatan Ibu dan anak
  - Hitung sampel untuk masing-masing variabel
  - Ambil jumlah sampel yang terbesar
- Jumlah sampel adalah jumlah sampel yang bisa diambil datanya bukan rumah atau orang yang perlu didatangi
  - Contoh untuk sampel 100 balita, mungkin pewawancara harus datang ke 150 rumah tangga

Besar Sampel pada Uji Hipotesis

Besar sampel uji hipotesis beda proporsi

$$n=\frac{\left(z_{1-\alpha/2}\sqrt{2\bar{P}(1-\bar{P})}+z_{1-\beta}\sqrt{P_1(1-P_1)+P_2(1-P_2)}\right)^2}{(P_1-P_2)^2}$$

- Untuk beda proporsi 2 kelompok
- P1 dan P2 bergantung pada desain
- Jumlah untuk masing-masing kelompok
- P1-P2 = beda minimal yang dianggap bermakna secara substansi

P1 dan P2 pada eksperimen, kohort & cross-sectional

Sebab	Keluaran		Total
	+	**-**	
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	a+b+c+d

- P1 = a/(a+b)
- P2 = c/(c+d)

P1 dan P2 pada kasus-kontrol

Sebab	Keluaran		Total
	+	-	
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	a+b+c+d

- $P1 = a/(a+c)$
- $P2 = b/(b+d)$

Contoh P1 dan P2

- “Hubungan antara anemia dengan BBLR”
  - Desain kohort/cross sectional
    - P1: Proposi BBL R pada ibu anemia
    - P2: Proposi BBLR pada ibu tidak anemia
  - Desain kasus-kontrol
    - P1: Proporsi ibu anemia pada BBLR
    - P2: Proporsi ibu anemia pada non BBLR
- Kesalahan penetapan P1 dan P2 sering terjadi pada desain kasus-kontrol

Contoh Kohort

- Pada contoh ini P1-P2=20%
- Beda minimal proporsi BBLR yang dianggap bermakna 20% antara ibu anemia vs ibu non anemia
- → Jika nantinya ada beda BBLR 20% atau lebih pada n sampel yang diambil → Uji statistik “signifikan”
- → Jika nantinya ada beda BBLR kurang dari 20% pada n sampel yang diambil → Uji statistik “tidak signifikan”
- → Signifikan uji statistik dirancang berdasarkan pengertian tentang substansi
- INGAT: Perbedaan berapapun dapat dirancang untuk “signifikan” secara statistik asal jumlah sampel terpenuhi

Contoh Kohort

$$n = \frac{\left(1{,}96\sqrt{2*0{,}2(1-0{,}2)} + 0{,}84\sqrt{0{,}3(1-0{,}3) + 0{,}1(1-0{,}1)}\right)^2}{(0{,}3-0{,}1)^2}$$

$$n = 62/\mathit{kelompok}$$

- Jadi sampel yang dibutuhkan 62 ibu anemia dan 62 ibu non anemia
- Bukan berarti diambil sampel 124 ibu hamil → tidak menjamin diperoleh 62 ibu hamil anemia dan 62 ibu hamil non anemia

Contoh Kohort

- Suatu penelitian dilakukan untuk mengetahui hubungan antara anemia pada ibu hamil dengan BBLR dengan desain kohort
- Asumsi untuk besar sampel:
  - Proporsi BBLR pada ibu anemia: 30%
  - Proporsi BBLR pada ibu non anemia: 10%
  - → Peneliti menganggap beda minimal BBLR 20% antara ibu anemia vs ibu non anemia
  - Derajat kemaknaan: 5%
  - Kekuatan uji: 80%
  - Maka P=(0,3+0,1)/2 = 0,2

Contoh Kohort

- Berapa sampel ibu hamil yang perlu diambil agar dapat diperoleh 62 ibu hamil anemia & 62 ibu hamil non anemia?
  - Tergantung proporsi anemia pada ibu hamil
    - 60% bumil anemia, 40% tidak anemia
  - Jadi dihitung 62 = 40/100 * n’
  - n’ = 155 → akan diperoleh 93 bumil anemia & 62 bumil non anemia
    - Dari 93 bumil anemia dapat dipilih secara acak atau kuota 62 bumil saja

Contoh Kasus-Kontrol

- Suatu penelitian dilakukan untuk mengetahui hubungan antara anemia pada ibu hamil dengan BBLR dengan desain kasus kontrol
- Asumsi untuk besar sampel:
  - Proporsi anemia pada BBLR: 80%
  - Proporsi anemia pada non BBLR: 60%
  - → Peneliti menganggap beda minimal proporsi ibu anemia 20% antara bayi BBLR vs non BBLR
  - Derajat kemaknaan: 5%
  - Keluatan uji: 80%
  - Maka P=(0,8+0,6)/2 = 0,7

Contoh Kasus Kontrol

$$n = \frac{\left(1{,}96\sqrt{2*0{,}7(1-0{,}7)} + 0{,}84\sqrt{0{,}8(1-0{,}8) + 0{,}6(1-0{,}6)}\right)^2}{(0{,}8-0{,}6)^2}$$

$$n = 82/\,kelompok$$

- Jadi sampel yang dibutuhkan 82 bayi BBLR dan 82 bayi non BBLR
- Bukan berarti diambil sampel 164 bayi → tidak menjamin diperoleh 82 bayi BBLR dan 82 bayi non BBLR

Contol Kasus Kontrol

- Proporsi bayi yang BBLR 15%, 85% non BBLR
- 82 = 15/100 * n’
- n’ = 547
- Peneliti perlu mengikutsertakan 547 bayi sebagai sampel agar diperoleh 82 bayi BBLR

Masalah

- Jika hipotesis tidak fokus
  - Faktor-faktor yang berpengaruh pada kejadian BBLR
- P1 dan P2 yang mana ?
- Solusi:
  - Pilih faktor utama saja, faktor lain dianggap confounder
  - Hitung sampel untuk tiap faktor utama
- Perbedaan P1 dan P2 harus berdasarkan perbedaan yang dianggap secara subtansi bermakna, bukan hanya dari penelitian terdahulu saja

Contoh

- Penelitian “Faktor-faktor yang berhubungan dengan BBLR”
- Faktor utama yang ingin diuji
  - Anemia
  - Merokok
  - Hipertensi
  - Status Ekonomi

Contoh

- Maka perlu informasi tentang:
  - Prop BBLR pada anemia vs prop BBLR pada non anemia
  - Prop BBLR pada perokok vs prop BBLR pada non perokok
  - Prop BBLR pada hipertensi vs prop BBLR pada non hipertensi
  - Prop BBLR pada ibu miskin vs prop BBLR pada ibu non miskin
- Sampel terbesar yang diambil

Besar sampel uji hipotesis beda rata-rata

$$n = \frac{2\sigma^2 \left[z_{1-\alpha/2} + z_{1-\beta} \right]^2}{\left(\mu_1 - \mu_2 \right)^2}$$

$$\sigma^2 = \frac{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 \right]}{(n_1 - 1) + (n_2 - 1)}$$

Contoh

- Suatu penelitian dilakukan untuk membandingkan efek asupan natrium terhadap tek. darah orang dewasa
- Asumsi (dari ref):
  - Pada kel. Natrium rendah:
    - Rata-rata TD: 72 mmHg, SD:10 mmHg, n=20
  - Pada kel. Natrium tinggi:
    - Rata-rata TD: 85 mmHg, SD:12 mmHg, n=20
  - Perbedaan minimal yg ingin dideteksi: 10 mmHg
  - Derajat kemaknaan:5%
  - Kekuatan uji:80%

Contoh

$$\sigma^2 = \frac{\left[(20-1)10^2 + (20-1)12^2\right]}{(20-1)+(20-1)} = 122$$

$$n = \frac{2 * 122^2 \left[1{,}96 + 0{,}84\right]^2}{(10)^2} = 20$$

Dibutuh sampel 20 orang dengan asupan natrium tinggi
Dan 20 orang dengan asupan natrium rendah

Penggunaan data sekunder

- Tetap harus dihitung apakah data yang ada memadai dari segi jumlah dan pengambilan sampel
- Diambil sesuai jumlah sampel vs diambil seluruh data yang tersedia
- Alternatif: perhitungan power of the test dari pada perhitungan jumlah sampel

Uji Statistik *Chi-Square*

- Dibuat oleh Karl Pearson (1899)
- Sering disebut *Pearson's Chi-Square*
- Digunakan untuk :

Goodness of Fit &
Test for Independence

Rumus Umum

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E}$$

Goodness of Fit

- Uji Chi-Square mengenai perbedaan frekuensi yang diobservasi dengan frekuensi yang diharapkan.

KEJADIAN	**A1,A2,A3,...Ak**
FREKUENSI YANG DIOBSERVASI	**o1,o2,o3,...ok**
FREKUENSI YANG DIHARAPKAN	**e1,e2,e3,...ek**

Dalam hal ini kita ingin mengetahui APAKAH ADA PERBEDAAN BERMAKNA ANTARA *FREKUENSI* YANG DI*OBSERVASI* DENGAN *FREKUENSI* YANG DI*HARAPKAN*

- Sebuah uang logam dilemparkan sebanyak 100 kali. Hasilnya adalah 58 kali muncul sisi muka dan 42 kali sisi belakang. Ujilah hipotesis bahwa uang logam itu simetri dengan memakai taraf signifikansi $\alpha = 0{,}05$ dan $\alpha = 0{,}01$

- n = banyaknya lemparan = 100
- p = propabilitas muncul sisi muka = ½ dan
- q = probabilitas munculnya sisi belakang yaitu

  1-p = 1 - ½ = ½

Frekuensi HARAPAN munculnya sisi muka = n x p = 100 X ½ = 50

Frekuensi HARAPAN munculnya sisi belakang = n x q = 100 X ½ = 50

MAKA KITA AKAN MENDAPATKAN TABEL SEPERTI DIBAWAH INI

KEJADIAN MUNCUL SISI MATA UANG LOGAM	A 1 (sisi muka) ,	A2 (sisi belakang)
FREKUENSI YANG DIOBSERVASI	58	42
FREKUENSI YANG DIHARAPKAN	50	50

Uang logam simetris kalau probabilitas munculnya sisi muka sama dengan sisi belakang yaitu P(sisi muka) = P(sisi belakang) = ½

H0 : P(muka) = P(belakang) = ½
Ha : P(muka) ≠ P(belakang)
$\alpha = 0{,}05$ dan $\alpha = 0{,}01$
Kategori kejadiannya ada dua yaitu munculnya sisi muka dan munculnya sisi belakang, maka k = 2. *Degree of Freedom* (df) nya adalah k-1 = 2-1 = 1.
Nilai kritis $\chi 2$ untuk $\alpha = 0{,}05$ dan df = 1 adalah 3,841 dan $\alpha = 0{,}01$ adalah 6,635 (lihat tabel)

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E}$$

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E} = \frac{(O1 - E1)^2}{E1} + \frac{(O2 - E2)^2}{E2}$$

$$\chi^2 = \frac{(58-50)^2}{50} + \frac{(42-50)^2}{50} = 1,28 + 1,28 = 2,56$$

Pada **α = 0,05; nilai χ2 hit = 2,56 < 3,841**

Pada **α = 0,01; nilai χ2 hit = 2,56** < 6,635

Kesimpulannya H0 TIDAK DITOLAK (diterima) artinya UANG LOGAM ITU SIMETRI

Karena besar kecilnya nilai χ**2** pada dasarnya menunjukkan **kesesuaian** antara frekuensi yang **diobservasi** dengan frekuensi yang **diharapkan,** maka seringkali uji χ**2** disebut **UJI KEBAIKAN SUAI** (***Goodness of Fit***)

Test for Independence

- **Uji Chi-Square kebebasan dua faktor.**
- **Uji hipotesis mengenai ada / tidaknya hubungan (asosiasi) atau kaitan antara dua faktor.**
- **Misal : <u>Apakah Prestasi Belajar Mahasiswa Dalam Mata Kuliah Statistika Ada Hubungannya Dengan Ketepatan Waktu Penyelesaian Karil ?</u>**

- Jika **TIDAK ADA HUBUNGAN** antara dua faktor itu, maka dikatakan bahwa dua faktor itu **SALING BEBAS** atau **INDEPENDEN**. Lebih tepatnya **INDEPENDEN** *secara statistik*.

→ Hipotesis nya selalu menyatakan bahwa kedua faktor saling bebas / independen (tidak terikat, tidak berkaitan, tidak berhubungan).

Oleh karena itu, bentuk Ho : Tidak ada hubungan/asosiasi antara X dan Y

Sebuah penelitian dilakukan untuk memperoleh informasi mengenai hubungan antara Prestasi Belajar MA.Statistika dengan Ketepatan Waktu Penyelesaikan KARIL pada mahasiswa FKG UI Angkatan 2006. Hasil penelitian tersebut tertuang dalam tabel dibawah ini (dengan α = **0,05**)

Prestasi Belajar MA.Statistika	**Ketepatan Waktu KARIL**		**Total Baris**
	Tepat Waktu	Waktu Lebih	
Memuaskan	55 a	20 b	75 ab
Kurang Memuaskan	10 c	15 d	25 cd
Total Kolom	65 ac	35 bd	100 N

Langkah-Langkah Penyelesaian

- Tentukan nilai EXPECTED setiap sel.

- Sel a $= \dfrac{(total\_baris)\,X\,(total\_kolom)}{total\_observasi}$

$$= \frac{75\ X\ 65}{100} = 48{,}75$$

Sel c = 65 – 48,75 = 16,25

Sel b = 75 – 48,75 = 26,25

Sel d = 35 – 26,25 = 8,75

Sehingga kita mendapatkan tabel seperti dibawah

Prestasi Belajar MA.Statistika	**Ketepatan Waktu KARIL**		**Total Baris**
	Tepat Waktu	Waktu Lebih	
Memuaskan	O : 55 ; E : 48,75 (O-E)= 6,25 $\chi 2 = 0{,}801$	O : 20 ; E : 26,25 (O-E)= - 6,25 $\chi 2 = 1{,}488$	75
Kurang Memuaskan	O : 10 ; E : 16,25 (O-E)= - 6,25 $\chi 2 = 2{,}404$	O : 15 ; E : 8,75 (O-E)= 6,25 $\chi 2 = 4{,}464$	25
Total Kolom	65	35	100

$$\chi 2 = 0{,}801 + 1{,}488 + 2{,}404 + 4{,}464 = 9{,}157$$

Prestasi Belajar MA.Statistika	Ketepatan Waktu KARIL	
	Tepat Waktu	Waktu Lebih
Memuaskan	O : 55 ; E : 48,75 (O-E)= 6,25 $\chi 2 = 0{,}801$	O : 20 ; E : 26,25 (O-E)= - 6,25 $\chi 2 = 1{,}488$
Kurang Memuaskan	O : 10 ; E : 16,25 (O-E)= - 6,25 $\chi 2 = 2{,}404$	O : 15 ; E : 8,75 (O-E)= 6,25 $\chi 2 = 4{,}464$

$$\chi 2 = 0{,}801 + 1{,}488 + 2{,}404 + 4{,}464 = 9{,}157$$

Hasil dan Simpulan

- **$\alpha = 0{,}05$**
- **df = (b-1) x (k-1) = (2-1)(2-1) = 1**
- **$\chi 2$ tab = 3,841**
- **$\chi 2$ hit = 9,157**
- **$\chi 2$ hit > $\chi 2$ tab → Ho ditolak. → Kedua faktor TIDAK BEBAS / INDEPENDEN satu dengan yang lain.**
- **Artinya :** Ada Hubungan antara Prestasi Belajar Mahasiswa Dalam Mata Kuliah Statistika Ada Hubungannya Dengan Ketepatan Waktu Penyelesaian Karil
- Didalam pembahasan hasil penelitian, simpulannya ditulis <u>Prestasi Belajar Mahasiswa Dalam Mata Kuliah Statistika BERHUBUNGAN Dengan Ketepatan Waktu Penyelesaian Karil</u>

YATE'S CORRECTION

- Bila jumlah sampel kecil, penggunaan rumus *chi-square* dikoreksi oleh FRANK YATE (Ahli Statistik Inggris)

$$\chi^2_{\text{Yates}} = \sum_{i=1}^{N} \frac{(|O_i - E_i| - 0.5)^2}{E_i}$$

Catt :

1. Koreksi dari Yate cenderung berlebihan (*overcorrect*) sehingga ada sebagian ahli yang menyatakan sebaiknya tetap menggunakan rumus Pearson walaupun jumlah sampelnya hanya 20.
2. *Untuk melakukan chi-square tes, sel TIDAK BOLEH ada angka NOL (0)*

REGRESI LINEAR

Latar Belakang

- Terdapat kejadian– kejadian , kegiatan-kegiatan, atau masalah- masalah yang saling berhubungan satu sama lain
- Dibutuhkan analisis hubungan antara kejadian tersebut
- Perlu dibahas mengenai bentuk hubungan yang ada atau diperkirakan ada antara kedua perubah tersebut

APA YANG DIUKUR DARI HUBUNGAN TERSEBUT

- Bagaimana hubungan fungsional dua kejadian tersebut atau bagaimana persamaan matematis yang mempresentasikan hubungan dua kejadian tersebut (analisis regresi)
- Bagaiman kekuatan atau keeratan hubungan dua kejadian tersebut (analisis korelasi)

Dua variabel dalam regresi

- Variabel bebas → X
- Variabel terikat → Y

UKURAN DALAM REGRESI

- **Koefisien Regresi**

  → mengukur besarnya pengaruh X terhadap Y
- **Koefisien korelasi**

→ mengukur Kuat tidaknya hubungan X dan Y

UJI HIPOTESIS DALAM REGRESI

- uji keberartian koefisien regresi
- Uji keberartian model regresi / Uji linearitas
- Uji Korelasi

JENIS REGRESI LINEAR SEDERHANA

- Linear positif
- Linear negatif

APA ITU GARIS REGRESI?

- Garis linear yang menunjukan pola hubungan antara dua variabel misalnya variabel X dan Y sebenarnya hanya merupakan garis taksiran yang dipakai untuk mewakili pola sebaran data tersebut

MODEL REGRESI LINEAR SEDERHANA

$$\mathbf{Y} = \alpha + \beta\mathbf{X} + \varepsilon$$

Dimana

- ε adalah error random (kasalahan pengganggu)
- $\varepsilon \sim$ N (0, σ2).

$\sum (Y - \hat{Y})^2$

METODE KUADRAT TERKECIL

- kesalahan tidak dapat dihilangkan sama sekali, maka resiko betapapun kecilnya selalu ada.
- Resiko hanya bisa diperkecil dengan memperkecil kesalahan
- persamaan garis regresi yang paling baik adalah persamaan garis regresi yang mempunyai total kuadrat kesalahan kecil

TOTAL KUADRAT KESALAHAN

$$\sum (Y - \hat{Y})^2$$

No. Subyek	Var. Bebas (X)	Var. Terikat (Y)
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
n	x_n	y_n

MODEL DARI n DATA

- $y_i = \alpha + \beta x_i + \varepsilon_i$, untuk $i = 1,2, . . ., n$
- $\varepsilon_i = y_i - \alpha - \beta x_i$
- $(\varepsilon_i)^2 = (y_i - \alpha - \beta x_i)^2$
- J= $\sum_{i=1}^{n}(\varepsilon_i)^2 = \sum_{i=1}^{n}\left(y_i - \alpha - \beta x_i\right)^2$

J Diturunan terhadap α dan β

$$\frac{\partial J}{\partial \alpha} = -2\sum_{i=1}^{n} (y_i - \alpha - \beta x_i) = 0$$

$$\frac{\partial J}{\partial \beta} = -2\sum_{i=1}^{n} (y_i - \alpha - \beta x_i) x_i = 0$$

Persamaan baru

$$\sum_{i=1}^{n} y_i - n\alpha - \beta \sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} y_i x_i - \alpha \sum_{i=1}^{n} x_i - \beta \sum_{i=1}^{n} x_i^{\ 2} = 0$$

a dan b taksiran dari α dan β

$$\sum_{i=1}^{n} y_i - na - b\sum_{i=1}^{n} x_i = 0$$

$$na = \sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i$$

$$a = \sum_{i=1}^{n} \frac{y_i}{n} - b\sum_{i=1}^{n} \frac{x_i}{n}$$

$$a = \overline{y} - b\overline{x}$$

AKIBAT

$$\sum_{i=1}^{n} y_i x_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^{\ 2} = 0$$

$$\sum_{i=1}^{n} y_i x_i - (\bar{y} - b\bar{x})\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^{\ 2} = 0$$

$$\sum_{i=1}^{n} y_i x_i - \bar{y}\sum_{i=1}^{n} x_i + b\bar{x}\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^{\ 2} = 0$$

$$\sum_{i=1}^{n} y_i x_i - \frac{1}{n}\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i + b\frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^{\ 2} = 0$$

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

Hasil

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

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

RUMUSAN LAIN

$$b = \frac{\sum_{i=1}^{n} y_i x_i - n\overline{y}\overline{x}}{\sum_{i=1}^{n} x_i^{\,2} - n\overline{x}^{\,2}}$$

$$b = \frac{\sum_{i=1}^{n} (y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

SIMPANGAN KUADRAT X DAN Y

$$J_{xx} = S_{xx} = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$$

$$J_{yy} = S_{yy} = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$$

$$J_{xy} = S_{xy} = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i$$

AKIBAT

$$b = \frac{J_{xx}}{J_{xy}}$$

JUMLAH KUADRAT

- Jumlah kuadrat total (JKT)
- Jumlah kuadrat regresi (JKR)
- Jumlah kuadrat sesatan (JKS)

JUMLAH KUADRAT TOTAL (JKT)

$$J_{yy} = S_{yy} = \sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} y_i \right)^2$$

JUMLAHKUADRAT REGRESI (JKR)

$$\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2=\sum_{i=1}^{n}\left(\left(a+bx_i\right)-\left(a+b\bar{x}\right)\right)^2$$

$$=\sum_{i=1}^{n}\left(bx_i-b\bar{x}\right)^2$$

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

$= b^2J_{xx}$

$= b\ J_{xy}$

JUMLAH KUADRAT SESATAN (JKS)

$$\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n}\left(y_i - a - bx_i\right)^2$$

$$= \sum_{i=1}^{n}\left(y_i - \left(\bar{y} - b\bar{x}\right) - bx_i\right)^2$$

$$= \sum_{i=1}^{n}\left(\left(y_i - \bar{y}\right) - b(x_i - \bar{x})\right)^2$$

$$= \sum_{i=1}^{n}\left\{(y_i - \bar{y})^2 - 2(y_i - \bar{y})b(x_i - \bar{x}) + b^2(x_i - \bar{x})^2\right\}$$

LANJUTAN JKS

$$= \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 - 2b\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right) + b^2\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

$$= J_{yy} - 2bJ_{xy} + b^2 J_{xx}$$

$$= J_{yy} - bJ_{xy}$$

=*Jyy - bJxy*

HUBUNGAN JKT, JKR, JKS

- JKT = JKR + JKS

DERAJAT KEBEBASAN(DK) MASING-MASING JK

- Derajat kebebasan untuk JKT adalah *n* -1
- Derajat kebebasan untuk JKR adalah 1
- Derajat kebebasan untuk JKS adalah n -2

HUBUNGAN DK

- (n -1) = (n -2) + 1

RATA-RATA JUMLAH KUADRAT (RJK)

- kuadrat tengah / kuadrat rata- rata /rata-rata jumlah kuadrat didefinisikan dengan jumlah kuadrat dibagi oleh derajat bebasnya dinamakna

$$RJK.. = \frac{JK..}{DK..}$$

JENIS-JENIS RJK

- RJK REGRESI (RJKR)

RJKR = JKR

- RJK SESATAN (RJKS)

$$RJKS = \frac{JKS}{n-2}$$