## Distribusi Sampling
## (Distribusi Penarikan Sampel)

### 1. Pendahuluan

- Bidang Inferensia Statistik membahas generalisasi/penarikan kesimpulan dan prediksi/ peramalan. Generalisasi dan prediksi tersebut melibatkan sampel/contoh, sangat jarang menyangkut populasi.
- Sensus = pendataan setiap anggota populasi
- Sampling = pendataan sebagian anggota populasi = penarikan contoh = pengambilan sampel
- Pekerjaan yang melibatkan populasi tidak digunakan, karena:
    1. mahal dari segi biaya dan waktu yang panjang
    2. ketelitian pekerjaan yang melibatkan sampel lebih tinggi dibanding pekerjaan yang melibatkan populasi
    3. populasi akan menjadi rusak atau habis jika disensus
       misal : dari populasi donat ingin diketahui rasanya, jika semua donat dimakan, dan donat tidak tersisa, tidak ada yang dijual?
- Sampel yang baik $\rightarrow$ Sampel yang representatif
  Besaran/ciri sampel (*Statistik Sampel*) memberikan gambaran yang tepat mengenai besaran/ciri populasi (*Parameter Populasi*)

Masih ingat beda antara *Statistik Sampel* Vs *Parameter Populasi*? Perhatikan tabel berikut:

| Ukuran/Ciri | Statistik Sampel | Parameter Populasi |
|---|---|---|
| Rata-Rata | $\bar{x}$ | μ : myu |
| Selisih 2 Rata-rata | $\left\|\bar{x}_1 - \bar{x}_2\right\|$ : nilai mutlak | $\left\|\mu_1 - \mu_2\right\|$ : nilai mutlak |
| Standar Deviasi = Simpangan Baku | s | σ : sigma |
| Varians = Ragam | $s^2$ | $\sigma^2$ |
| Proporsi | $\bar{p}$ atau $\hat{p}$ | π : phi atau p |
| Selisih 2 proporsi | $\left\|\bar{p}_1 - \bar{p}_2\right\|$ : nilai mutlak | $\left\|\pi_1 - \pi_2\right\|$ : nilai mutlak |

catatan : pada Nilai Mutlak, nilai negatif diabaikan misal : $|3 - 7| = |-4| = 4$
atau gunakan asumsi $\bar{p}_1$ adalah nilai yang selalu lebih besar dari $\bar{p}_2$ atau $\bar{p}_1 > \bar{p}_2$

- Sampel yang baik diperoleh dengan memperhatikan hal-hal berikut :
    1. keacakannya (*randomness*)
    2. ukuran
    3. teknik penarikan sampel (*sampling*) yang sesuai dengan kondisi atau sifat populasi

- Sampel Acak = Contoh Random $\rightarrow$ dipilih dari populasi di mana setiap anggota populasi memiliki peluang yang sama terpilih menjadi anggota sampel.

- Berdasarkan Ukurannya, maka sampel dibedakan menjadi :
  - a. Sampel Besar jika ukuran sampel (n) $\geq$ 30
  - b. Sampel Kecil jika ukuran sampel (n) < 30

- Beberapa Teknik Penarikan Sampel :

a. Penarikan Sampel Acak Sederhana (*Simple Randomized Sampling*)
Pengacakan dapat dilakukan dengan : undian, tabel bilangan acak, program komputer.

b. Penarikan Sampel Sistematik (*Systematic Sampling*)
Tetapkan interval lalu pilih secara acak anggota pertama sampel

Contoh : Ditetapkan interval = 20
Secara acak terpilih : Anggota populasi ke-7 sebagai anggota ke-1 sampel
maka :
Anggota populasi ke-27 menjadi anggota ke-2 sampel
Anggota populasi ke-47 menjadi anggota ke-3 sampel, dst.

c. Penarikan Sampel Berlapis (*Stratified Sampling*)
Populasi terdiri dari beberapa kelas/kelompok. Dari setiap kelas diambil sampel secara acak.

Perhatikan !!!!
Antar Kelas bersifat (cenderung) berbeda nyata (heterogen). Anggota dalam suatu kelas akan (cenderung) sama (homogen).

Contoh :
Dari 1500 penumpang KA (setiap kelas memiliki ukuran yang sama) akan diambil 150 orang sebagai sampel, dilakukan pendataan tentang tingkat kepuasan, maka sampel acak dapat diambil dari :

| | |
|---|---|
| Kelas Eksekutif | : 50 orang |
| Kelas Bisnis | : 50 orang |
| Kelas Ekonomi | : 50 orang |

d. Penarikan Sampel Gerombol/Kelompok (*Cluster Sampling*)
Populasi juga terdiri dari beberapa kelas/kelompok
Sampel yang diambil berupa kelompok bukan individu anggota

Perhatikan !!!!
Antar Kelas bersifat (cenderung) sama (homogen). Anggota dalam suatu kelas akan (cenderung) berbeda (heterogen).

Contoh :

Terdapat 40 kelas untuk tingkat II Jurusan Ekonomi-GD, setiap kelas terdiri dari 100 orang. Populasi mahasiswa kelas 2, Ekonomi-UGD = 40 × 100 = 4000.
Jika suatu penelitian dilakukan pada populasi tersebut dan sampel yang diperlukan = 600 orang, maka sampel dapat diambil dari 6 kelas.... Dari 40 kelas, ambil secara acak 6 kelas.

e. Penarikan Sampel Area (*Area Sampling*)
Prinsipnya sama dengan *Cluster Sampling*.
Pengelompokan ditentukan oleh letak geografis atau administratif.

Contoh : Pengambilan sampel di daerah JAWA BARAT, dapat dilakukan dengan memilih secara acak KOTAMADYA tempat pengambilan sampel, misalnya terpilih, Kodya Bogor, Sukabumi dan Bandung.

Sampel acak menjadi dasar penarikan sampel lain. Selanjutnya, pembahasan akan menyangkut Penarikan Sampel Acak.

- Penarikan Sampel Acak dapat dilakukan dengan 2 cara, yaitu :

a. Penarikan sampel tanpa pemulihan/tanpa pengembalian: setelah didata, anggota sampel tidak dikembalikan ke dalam ruang sampel

b. Penarikan sampel dengan pemulihan : bila setelah didata, anggota sampel dikembalikan ke dalam ruang sampel.

Distribusi Penarikan Sampel = Distribusi Sampling
- Jumlah Sampel Acak yang dapat diambil dari suatu populasi adalah sangat banyak.
- Nilai setiap Statistik Sampel akan bervariasi/beragam antar sampel.
- Suatu statistik dapat dianggap sebagai peubah acak yang besarnya sangat tergantung dari sampel yang kita ambil.
- Karena statistik sampel adalah peubah acak maka ia mempunyai distribusi peluang yang kita sebut sebagai : Distribusi peluang statistik sampel = Distribusi Sampling = Distribusi Penarikan Sampel

Statistik sampel yang paling populer dipelajari adalah Rata-Rata ( $\bar{x}$ )

## 2. Distribusi Sampling 1 Nilai Rata-Rata

Beberapa notasi :

| | | | |
|---|---|---|---|
| n | : ukuran sampel | N | : ukuran populasi |
| $\bar{x}$ | : rata-rata sampel | $\mu$ | : rata-rata populasi |
| s | : standar deviasi sampel | $\sigma$ | : standar deviasi populasi |
| $\mu_{\bar{x}}$ | : rata-rata dari semua rata-rata sampel | | |
| $\sigma_{\bar{x}}$ | : standar deviasi antar semua rata-rata sampel = standard error = galat baku | | |

Dalil 1

JIKA

Sampel:
berukuran = n ≥ 30
rata-rata = $\bar{x}$

} diambil DENGAN PEMULIHAN dari

{ Populasi berukuran = N
Terdistribusi NORMAL
Rata-rata = μ ; simpangan baku = σ

MAKA

Distribusi Rata-rata akan mendekati distribusi Normal dengan :

$$\mu_{\bar{x}} = \mu \qquad \text{dan} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \text{dan nilai } z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Dalil 2

JIKA

Sampel:
berukuran = n ≥ 30
rata-rata = $\bar{x}$

} diambil TANPA PEMULIHAN dari

{ Populasi berukuran = N
Terdistribusi NORMAL
Rata-rata = μ ; simpangan baku = σ

MAKA

Distribusi Rata-rata akan mendekati distribusi Normal dengan :

$$\mu_{\bar{x}} = \mu \qquad \text{dan} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad \text{dan nilai } z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})\sqrt{\frac{N-n}{N-1}}}$$

- $\sqrt{\frac{N-n}{N-1}}$ disebut sebagai FAKTOR KOREKSI populasi terhingga.
- Faktor Koreksi (FK) akan menjadi penting jika sampel berukuran n diambil dari populasi berukuran N yang terhingga/ terbatas besarnya
- Jika sampel berukuran n diambil dari populasi berukuran N yang sangat besar maka FK akan mendekati 1 $\rightarrow \sqrt{\frac{N-n}{N-1}} \approx 1$, hal ini mengantar kita pada dalil ke-3 yaitu DALIL LIMIT PUSAT = DALIL BATAS TENGAH = *THE CENTRAL LIMIT THEOREM*

Dalil 3 DALIL LIMIT PUSAT

JIKA

Sampel:
berukuran = n
rata-rata = $\bar{x}$

$\left.\begin{array}{l} \text{Sampel:} \\ \text{berukuran} = n \\ \text{rata-rata} = \bar{x} \end{array}\right\}$ diambil dari

$\left\{\begin{array}{l} \text{Populasi berukuran} = N \text{ yang BESAR} \\ \text{distribusi : SEMBARANG} \\ \text{Rata-rata} = \mu \text{ ; simpangan baku} = \sigma \end{array}\right.$

MAKA

Distribusi Rata-rata akan mendekati distribusi Normal dengan :

$\mu_{\bar{x}} = \mu$ dan $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ dan nilai $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

- Dalil Limit Pusat berlaku untuk :
  - penarikan sampel dari populasi yang sangat besar,
  - distribusi populasi tidak dipersoalkan

- Beberapa buku mencatat hal berikut : Populasi dianggap BESAR jika ukuran sampel KURANG DARI 5 % ukuran populasi atau $\boxed{\dfrac{n}{N} < 5\%}$

Dalam pengerjaan soal DISTRIBUSI SAMPLING RATA-RATA perhatikan asumsi-asumsi dalam soal sehingga anda dapat dengan mudah dan tepat menggunakan dalil-dalil tersebut!

Contoh 1:

PT AKUA sebuah perusahaan air mineral rata-rata setiap hari memproduksi 100 juta gelas air mineral. Perusahaan ini menyatakan bahwa rata-rata isi segelas AKUA adalah 250 ml dengan standar deviasi = 15 ml. Rata-rata populasi dianggap menyebar normal.

1. Jika setiap hari diambil 100 gelas AKUA sebagai sampel acak DENGAN PEMULIHAN, hitunglah:
   a. standard error atau galat baku sampel tersebut?
   b. peluang rata-rata sampel akan berisi kurang dari 253 ml?
2. Jika sampel diperkecil menjadi 25 gelas, hitunglah :
   a. standard error atau galat baku sampel tersebut?
   b. peluang rata-rata sampel akan berisi lebih dari 255 ml?

Jawab:

1. Diselesaikan dengan DALIL 1 → karena PEMULIHAN
   Diselesaikan dengan DALIL 3 → karena POPULASI SANGAT BESAR

N = 100 000 000 $\mu_{\bar{x}} = \mu = 250$ $\sigma = 15$ n = 100

P( $\bar{x}$ < 253) = P(z < ?)

GALAT BAKU = $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = \frac{15}{10} = 1.5$

$$z = \frac{253 - 250}{1.5} = \frac{3}{1.5} = 2.0$$

Jadi P( $\bar{x}$ < 253) = P(z < 2.0) = 0.5 + 0.4772 = 0.9772

2. Diselesaikan dengan DALIL 3 → karena POPULASI SANGAT BESAR

N = 100 000 000 $\mu_{\bar{x}} = \mu = 250$ $\sigma = 15$ n = 25

P( $\bar{x}$ > 255) = P(z > ?)

GALAT BAKU = $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{25}} = \frac{15}{5} = 3.0$

$$z = \frac{255 - 250}{3.0} = \frac{5}{3.0} = 1.67$$

Jadi P( $\bar{x}$ > 255 ) = P(z > 1.67) = 0.5 - 0.4525 = 0.0475

Contoh 2 :

Dari 500 mahasiswa FE-GD diketahui rata-rata tinggi badan = 165 cm dengan standar deviasi = 12 cm, diambil 36 orang sebagai sampel acak. Jika penarikan sampel dilakukan TANPA PEMULIHAN dan rata-rata tinggi mahasiswa diasumsikan menyebar normal, hitunglah :

a. galat baku sampel?
b. peluang sampel akan memiliki rata-rata tinggi badan kurang dari 160 cm?

Diselesaikan dengan DALIL 2 → TANPA PEMULIHAN
N = 500 $\mu_{\bar{x}} = \mu = 165$ $\sigma = 12$ n = 36

Catatan $\frac{n}{N} = \frac{36}{500} = 0.072 = 7.2\% > 5\%$ → Dalil Limit Pusat tidak dapat digunakan

$P(\bar{x} < 160) = P(z < ?)$

$$FK = \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{500-36}{500-1}} = \sqrt{\frac{464}{499}} = \sqrt{0.929\ldots} = 0.964\ldots$$

$$\text{GALAT BAKU} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \text{ x FK } = \frac{12}{\sqrt{36}} \times 0.964\ldots = 2 \text{ x } 0.964\ldots = 1.928\ldots$$

$$z = \frac{160-165}{1.928\ldots} = -2.59\ldots$$

$P(\bar{x} < 160) = P(z < -2.59) = 0.5 - 0.4952 = 0.0048$

## 3. Distribusi Sampling Bagi Beda 2 Rata-rata

Dalil 4

JIKA

Dua (2) Sampel
berukuran $n_1$ dan $n_2$
rata-rata = $\bar{x}_1$ dan $\bar{x}_2$

diambil dari

Dua (2) Populasi berukuran BESAR
Rata-rata $\mu_1$ dan $\mu_2$
Ragam $\sigma_1^2$ dan $\sigma_2^2$

MAKA

Distribusi Rata-rata akan mendekati distribusi Normal dengan :

$$\mu_{\bar{x}_1-\bar{x}_2} = |\mu_1 - \mu_2| \quad \text{dan} \quad \text{standard error} = \sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \text{dan}$$

nilai z

$$z = \frac{|\bar{x}_1 - \bar{x}_2| - |\mu_1 - \mu_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- Beda atau selisih 2 rata-rata = $|\mu_1 - \mu_2|$ $\rightarrow$ ambil nilai mutlaknya atau tetapkan bahwa $\mu_1 > \mu_2$
- Melibatkan 2 populasi yang BERBEDA dan SALING BEBAS
- Sampel-sampel yang diambil dalam banyak kasus (atau jika dilihat secara akumulatif) adalah sampel BESAR

Contoh 4:
Diketahui rata-rata IQ populasi mahasiswa Eropa = 125 dengan ragam = 119 sedangkan rata-rata IQ populasi mahasiswa Asia = 128 dengan ragam 181. Diasumsikan kedua populasi berukuran besar
Jika diambil 100 mahasiswa Eropa dan 100 mahasiswa Asia sebagai sampel, berapa peluang terdapat perbedaan IQ kedua kelompok akan kurang dari 2?

Jawab :

| Parameter | populasi ke-1 (Mhs. Eropa) | populasi ke-2 (Mhs. Asia) |
|---|---|---|
| Rata-rata ($\mu$) | $125 = \mu_2$ | $128 = \mu_1$ |
| Ragam ($\sigma^2$) | $119 = \sigma^2_2$ | $181 = \sigma^2_1$ |

Beda 2 Rata-rata = $\mu_{\bar{x}_1-\bar{x}_2} = |\mu_1 - \mu_2|$ = 128 - 125 = 3

Sampel : $n_1$ = 100 $n_2$ = 100

P( $|\bar{x}_1 - \bar{x}_2|$ <2 ) = P ( z < ?)

$$z = \frac{|\bar{x}_1 - \bar{x}_2| - |\mu_1 - \mu_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{2-3}{\sqrt{\frac{181}{100} + \frac{119}{100}}} = \frac{-1}{\sqrt{3}} = -0.577... \approx -0.58$$

P(z <-0.58) = 0.5 - 0.2190 = 0.2810

selesai