## Previous Topic

Previous topic

Frekuensi dan deskripsi statistics of two data

Ini dan data

## Draw a Conclusion

Dalam suatu penelitian, ada 2 variabel yang dibandingkan, contoh yang lagi dibahas mengenai hubungan pendapatan dan pengeluaran

# Mean

## Inferential Statistics

### What is Inferential Statistics

Tujuan utama dalam suatu statistik yang mau kita mengumpulkan sampel yang diagus mempresentasikan ini yang ada di populasi

### Example 1

PT SUKAMAJU

**Target Value per Customer**

Mean = 10 K

**Statistics Value**

Mean = 12 K
Standard Deviation = 1 K

Dalam kasus ini, PT Sukamaju ingin membandingkan apakah value per customer dari tahun lalu lebih dari target perusahaan. Kalau ini dibandingkan dengan sampel dari 10 pelanggan yang kita ambil secara random, kita dapatkan rata-rata dari sampel tersebut 12 K

### What We have to Know?

1. Sampling Method
2. Sampling Distribution & Distribution
3. Concept of Hypothesis Testing

## Sampling Method

Sampling adalah suatu cara untuk mengambil sampel dari populasi

### Type of Sampling Method

1. Probability Sampling

Mengambil sampel berdasarkan nilai di dalam sebaran. Sampel harus mewakili populasi. Probability sampling yang akan kita pelajari adalah simple random sampling, stratified random sampling, cluster random sampling, dan systematic random sampling

2. Non-probability Sampling

Tidak mempertimbangkan sebaran data, tidak ada jaminan bahwa sampel yang diambil akan mewakili populasi. Non-probability sampling yang akan kita pelajari adalah convenience sampling, purposive sampling, dan snowball sampling

#### Simple Random Sampling

Cara paling sederhana untuk mengambil sampel dari populasi dengan cara random. Setiap individu/item di populasi dengan peluang yang sama untuk dipilih. Contoh: Kita mau ambil sampel 100 mahasiswa dari populasi 1000 mahasiswa. Kita bisa pakai random number generator, atau undian, atau kocok nama dan ambil 100 pertama. Cocok untuk populasi yang homogen (seragam). Mudah dilakukan, tapi bisa jadi kurang efisien kalau populasi sangat besar atau tersebar luas

#### Stratified Sampling Vs Cluster Sampling

### Non-probability Sampling

1. Convenience Sampling

Mengambil sampel berdasarkan kemudahan akses. Contoh: mengambil sampel dari siswa yang kebetulan ada di kelas saat penelitian karena mereka lebih mudah dihubungi

2. Purposive Sampling

Sampel dipilih berdasarkan tujuan atau kriteria tertentu yang ditetapkan peneliti. Contoh: Peneliti ingin mengetahui pendapat ibu hamil tentang vitamin, maka peneliti hanya mengambil sampel dari ibu hamil yang ada di puskesmas. Berguna jika peneliti ingin fokus pada kelompok tertentu yang dianggap relevan. Bisa menghasilkan bias jika kriteria yang ditetapkan terlalu ketat

3. Snowball Sampling

Peneliti mencari responden awal, lalu responden tersebut diminta untuk mengajak teman-teman mereka jadi sampel. Contoh: Peneliti ingin meneliti pengalaman pengguna aplikasi ojek online, maka peneliti mencari beberapa pengguna awal, lalu mereka diminta untuk mengajak teman-teman mereka yang juga pengguna aplikasi ojek online. Berguna untuk populasi yang sulit dijangkau atau tersembunyi (misal: pengguna narkoba, pekerja seks, dll). Sangat bergantung pada responden awal yang "baik" untuk terus mengajak orang lain

## Error Type

1. Sampling Error

Kesalahan yang terjadi karena sampel yang diambil tidak sepenuhnya mewakili populasi secara sempurna. Bukan disengaja, tapi memang ada saja kesalahan kecil karena keterbatasan sampel

Formula di Bahan Ajaran juga ada di mission mencari nilai error (standard error) dengan rumus sampel error (standard error) di bawah ini

$$SE = \frac{\sigma}{\sqrt{n}}$$

$\sigma$ = simbol sampel

$n$ = jumlah populasi

2. Non-sampling Error

Kesalahan yang terjadi karena bukan karena sampel yang diambil tidak mewakili populasi, tapi karena ada masalah lain di proses pengumpulan data atau analisis. Contoh: Salah ngitung, salah input data, responden ngasih jawaban yang salah, dll

## Sample Size (Salah satu cara menentukan jumlah Sample)

Slovin :

$$n = \frac{N}{N \times e^2 + 1}$$

n = Number of Observation
N = Population

- jumlah sampel
- jumlah populasi

Error Tolerance (kesalahan error toleransi)

Contoh: Kita mau ambil sampel dari populasi 1000 mahasiswa dengan toleransi error 5%, maka rumusnya jadi: n = 1000 / (1000 × 0.05² + 1) = 1000 / (1000 × 0.0025 + 1) = 1000 / (2.5 + 1) = 1000 / 3.5 = 286 mahasiswa. Artinya, peneliti dengan toleransi 5% error bisa ambil sampel sekitar 286 mahasiswa

Sampling error juga sebuah bagian dari Sampling Distribution

Rumus CTS (central limit theorem) untuk menemukan jumlah sampel dengan toleransi

$$n = \left(\frac{z \times \sigma}{E}\right)^2$$

Keterangan:

- $n$ = ukuran sampel yang dicari
- $z$ = nilai z untuk tingkat kepercayaan (misal 1.96 untuk 95%)
- $\sigma$ = standar deviasi populasi
- $E$ = error tolerance (toleransi error)

Contoh: Kita mau ambil sampel dari populasi 1000 mahasiswa dengan toleransi error 5%, dengan tingkat kepercayaan 95% dan standar deviasi 10. Maka: n = (1.96 × 10 / 0.05)² = (19.6 / 0.05)² = (392)² = 153.664 ≈ 154 mahasiswa. Artinya, peneliti dengan toleransi error 5% dan tingkat kepercayaan 95% bisa ambil sampel sekitar 154 mahasiswa

## Concept of Probability

Probability adalah seberapa kemungkinan

Ini nilai antara angka antara nol sampai satu. Bisa dari 0 atau nol atau satu atau pecahan seperti 0.5 atau 0.75

Terkadang kita mau menghitung peluang dari suatu kejadian yang mungkin terjadi dari sebuah percobaan yang dilakukan. Misalnya, kita mau tahu peluang dari dua dadu yang dilemparkan dan ada angka yang muncul. Tetapi kita tidak dapat memastikan berapa dua kali lemparan dadu yang akan muncul

Peluang dapatkan angka ganjil dari dadu

Probability is a numerical description of how likely an event is to occur or how likely it is that a proposition is true

Usually denoted by Round

P(A) - Probability of event A

Range

- Nilai 0 = tidak mungkin terjadi
- Nilai 1 = pasti terjadi

Meaning: Meaning increasing from 0 to 1 means that the event is more likely to happen. Nilai 0 = tidak mungkin terjadi, nilai 1 = pasti terjadi, nilai antara 0 dan 1 = kemungkinan terjadi

### Probability - test

### Disjoint and overlapping in Probability

Diagram — Mutually exclusive

P(A or B) ketemu mana ke dua B

### Disjoint - test

P(Disjoint or D) = Kalau Disjoint/mutually exclusive maka P(event) = P(D) jadi 0.25 + 0.25 = 0.50

P(Disjoint or Red) = Kalau overlapping maka P(event) = P(Red) - P(Disjoint) jadi 0.25 + 0.25 - 0.25 = 0.25

## Probability Distribution

List of possible outcomes and the probabilities which they occur

| outcome | frequency | probability | value | value × probability |
|---|---|---|---|---|
| 1 | 1 | 1/6 | 1 | 1/6 |

Rumus:

- Ini ada ada 6
- Dari situ kita kalikan 6 aja?

## Probability Example:

Rolling dice

- Ketemu angka ganjil 1/2 or of dice is less than equal 3
- Ada angka ganjil 1/2 of dice is less than equal 3

What are you want to see, Kitaman is Rita?

Probability of 1/2 or 0.5 of dice is equal to 3?

Kalikan 1 × 1/6 = 1/6

Tambah 1/6 + 1/6

Probability using Mean:

Estimasi = 1/2 × 1 = 0.5

Estimasi = 1/2 × 1 = 0.5

## Common Distribution

Distribusi yang sering muncul dalam data, mirip dengan kita ketemu bentuk-bentuk pola yang sama berulang kali. Ini bukan cuma kebetulan, tapi ada rumus matematika yang ngatur kenapa bentuk-bentuk tertentu muncul. Misalnya tinggi manusia (cenderung ke tengah), nilai ujian (cenderung ke tengah), waktu tunggu antrian (cenderung pendek), dan lain-lain. Nah, kita bisa prediksi seperti apa bentuk data kita sebelum ngumpulin datanya, cukup pakai rumus yang sesuai

Meaning: Ngetrain/ngelatih algoritma supaya bisa ngenal pola-pola tertentu dari data. Contoh: kita ngajarin AI buat ngenalin "bentuk orang" dari ribuan foto, supaya dia bisa ngenali orang di foto yang belum pernah dilihat sebelumnya

Bentuk-bentuk distribusi yang sering muncul:

- Normal Distribution (bentuk lonceng, paling sering muncul)
- Binomial Distribution (cuma dua kemungkinan: sukses atau gagal)
- Poisson Distribution (berapa kali suatu kejadian terjadi dalam waktu tertentu)
- Uniform Distribution (semua kemungkinan sama rata)

Kenapa Bentuk Kurva Normal Penting?

- Banyak data alam mengikuti bentuk ini (tinggi manusia, berat badan, nilai ujian, dll)
- Rumus matematikanya sangat rapi dan bisa dipakai untuk prediksi
- Dasar dari banyak metode statistik yang lebih canggih

Bentuk-bentuk Kurva Normal:

- Kurva lonceng standard (mean = 0, standard deviation = 1)
- Kurva lonceng dengan mean yang geser
- Kurva lonceng dengan standard deviation yang berbeda

Contoh Distribusi Binomial:

- Distribusi Probability dari 3 dadu (Distribusi Binomial dengan parameter n dan p dikali 3 - Binyak)
  - n = banyaknya percobaan
  - p = kemungkinan sukses
- Distribusi Probability dari 3 dadu dikali dengan 2 (Distribusi Binomial dengan parameter n dan p dikali 3 - Binyak)
  - n = 3
  - p = 1/6 (sukses)

Contoh Distribusi Poisson:

- Distribusi Probability dari 3 pelanggan (Distribusi Poisson dengan parameter lambda dan 3 - Poisson)
  - lambda = rata-rata kedatangan
  - 3 = pengamatan

Distribusi Uniform:

- Distribusi Probability dari 3 dadu (Distribusi Uniform dengan parameter n dan 3 - Uniform)
  - n = banyaknya pilihan

## Distribusi Normal

Distribusi normal adalah distribusi yang paling penting dan paling sering dalam statistika

Kurva Distribusi Normal dengan mean = 0 dan standard deviation = 1

- Kurva lonceng yang paling sering muncul dalam data
- Simetris di sekitar mean
- Ekor kanan dan kiri mendekati nol tapi tidak pernah benar-benar nol
- Total luas di bawah kurva = 1 (atau 100%)

Kenapa Distribusi Normal Penting?

- Banyak data alam mengikuti bentuk ini (tinggi manusia, berat badan, nilai ujian, dll)
- Rumus matematikanya sangat rapi dan bisa dipakai untuk prediksi
- Dasar dari banyak metode statistik yang lebih canggih

Kurva Distribusi Normal dengan mean = 0 dan standard deviation = 1

- Kurva lonceng standard
- Mean = 0
- Standard deviation = 1

Kurva Distribusi Normal dengan mean yang geser

- Kurva lonceng dengan mean yang berbeda
- Bentuk tetap sama, cuma posisi di sumbu x yang berubah

Kurva Distribusi Normal dengan standard deviation yang berbeda

- Kurva lonceng dengan standard deviation yang berbeda
- Lebar kurva berubah, tapi bentuk dasar tetap sama

Rumus Kurva Normal

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Keterangan:

- $f(x)$ = tinggi kurva pada nilai x
- $\mu$ (mu) = mean
- $\sigma$ (sigma) = standard deviation
- $\pi$ (pi) = 3.14159...
- $e$ = 2.71828...

Contoh Hitung Probability dari Kurva Normal:

- Kita mau tahu probability dari nilai ujian di antara 70 dan 80
- Kita hitung luas di bawah kurva antara 70 dan 80
- Hasilnya = 0.3413 (34.13%)

Contoh Hitung Probability dari Kurva Normal:

- Kita mau tahu probability dari nilai ujian di atas 90
- Kita hitung luas di bawah kurva dari 90 ke kanan
- Hasilnya = 0.0228 (2.28%)

Contoh Hitung Probability dari Kurva Normal:

- Kita mau tahu probability dari nilai ujian di bawah 60
- Kita hitung luas di bawah kurva dari ujung kiri ke 60
- Hasilnya = 0.1587 (15.87%)

Contoh Hitung Probability dari Kurva Normal:

- Kita mau tahu probability dari nilai ujian di antara 65 dan 75
- Kita hitung luas di bawah kurva antara 65 dan 75
- Hasilnya = 0.2718 (27.18%)

## Teorema Limit Pusat

Teorema limit pusat adalah salah satu konsep paling penting dalam statistika. Intinya, kita bisa ambil sampel random, hitung rata-rata dari sampel tersebut, dan ulangi proses ini berkali-kali, maka distribusi dari rata-rata sampel tersebut akan mendekati kurva normal

Ini ini bergantungan sampel tersebut juga memiliki kesebaran yang juga tergap bentuk pola distribusi normal




Ambil pola Sampel, semua pencerminan kesebaran/sampel Bisa yang juga mengambil pola mirip/mirip dari populasi, sebagai contoh kita mau ambil sampel dari populasi distribusi binomial, maka

Maka, kesebaran bentuk kesebaran sampel juga, kan, dari rumus kesebaran atau kurva normal

Ini juga atau tidak, semua bentuk kurva yang mirip dengan kesebaran bentuk kesebaran tersebut, maka, kesebaran bentuk kesebaran tersebut juga

Meaning: Ngetrain/ngelatih algoritma supaya bisa ngenal pola-pola tertentu dari data. Contoh: kita ngajarin AI buat ngenalin "bentuk orang" dari ribuan foto, supaya dia bisa ngenali orang di foto yang belum pernah dilihat sebelumnya

## Hypothesis Testing

## Presumption of Innocence




Meaning: Ngetrain/ngelatih algoritma supaya bisa ngenal pola-pola tertentu dari data. Contoh: kita ngajarin AI buat ngenalin "bentuk orang" dari ribuan foto, supaya dia bisa ngenali orang di foto yang belum pernah dilihat sebelumnya

## We want to Prove




**H0 : H0 dalam pengadilan**
**H1 : H1 adalah pembuktian**

## How about Statistics

Ngetrain/ngelatih algoritma supaya bisa ngenalin/ngenali

Ngetrainnya cara hypothesis testing

Hypothesis testing/H1 dimasukkan ke cara kita dikitabkan. Dalam bahasa Tersebut Contoh kita ini, dia diajarkan bisa mendeteksi bentuk di sebuah gambar. H1 dimasukkan ke dalamnya kan?

**Null hypothesis (H0)**
H0 adalah pernyataan yang dianggap benar sampai ada cukup bukti untuk membantahnya. Ini adalah "status quo" atau "tidak ada perubahan".

**Alternative hypothesis (H1)**
H1 adalah lawan dari H0. Ini adalah "status quo" atau "tidak ada perubahan". Ini adalah "status quo" atau "tidak ada perubahan".

## Hypothesis Testing




Dalam pengadilan, ada prinsip sederhana bahwa lebih baik bebas satu orang bersalah daripada satu orang dihukum salah. Ini mirip dengan hypothesis testing

Type I Error = Salah tolak H0 (false positive)
Type II Error = Gagal tolak H0 (false negative)

## Example




## P-value

Nilai probabilitas dipakain dan ingat signifikansi / peluang/ sebuah/ampuan/ yaitu/ ketahuan/ ancaman/ kata/ Dalam/ Dalam/ dan/ dan/ Dalam/ dan/ dan/ dan