

# Dasar-Dasar Statistika Ilmu Komputer : Pengenalan Statistika

Statistika dasar



## Dasar teori

### Pengenalan Statistika

Statistika (Statistics) pada dasarnya berbeda dari Statistik (Statistic). Statistika merupakan ilmu yang berhubungan dengan pengumpulan, analisis, penginterpretasian data serta melakukan prediksi, sementara Statistik adalah data, informasi yang didapat dari data serta perhitungan dari suatu algoritma tertentu, tetapi sebutan Statistik sudah umum digunakan untuk semua yang berkaitan dengan Statistika maupun Statistik.

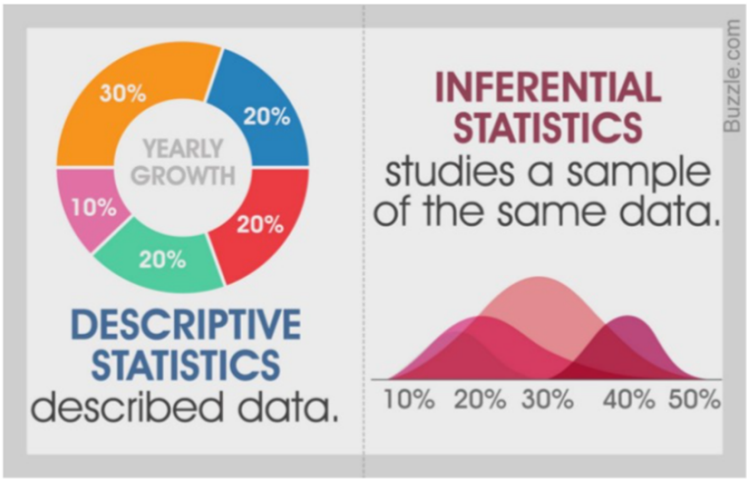
2 Jenis Statistik berdasarkan metodenya yaitu :

#### Statistik Deskriptif / Non-Eksperimental

Statistik deskriptif berkaitan dengan deskripsi data, menggambarkan informasi dari suatu data tersebut misalnya rata-rata, median, modus (mode), standard deviasi dan varian dari sekumpulan data yang dapat dianalisa dan divisualisasikan dengan tabel dan grafik agar mudah dibaca dan lebih bermakna.

#### Statistik Inferensial / Induktif / Probabilitas / Eksperimental.

Statistik inferensial digunakan untuk melakukan pengujian hipotesis, melakukan prediksi di masa depan dengan regresi, atau membuat klasifikasi suatu data dengan cara membuat model dan biasanya digunakan untuk melakukan pengambilan keputusan berdasarkan analisis data.



2 Jenis Statistik Inferensi berdasarkan parameternya yaitu :

#### Statistik Parametrik

Statistik ini digunakan untuk analisa dengan menggunakan parameter-parameter populasi seperti rata-rata dan sebagainya, biasanya digunakan untuk mengolah data kuantitatif (numerikal / diskrit / kontinu) yaitu data jenis interval atau ratio dengan memastikan distribusi data adalah distribusi normal atau mendekati normal.

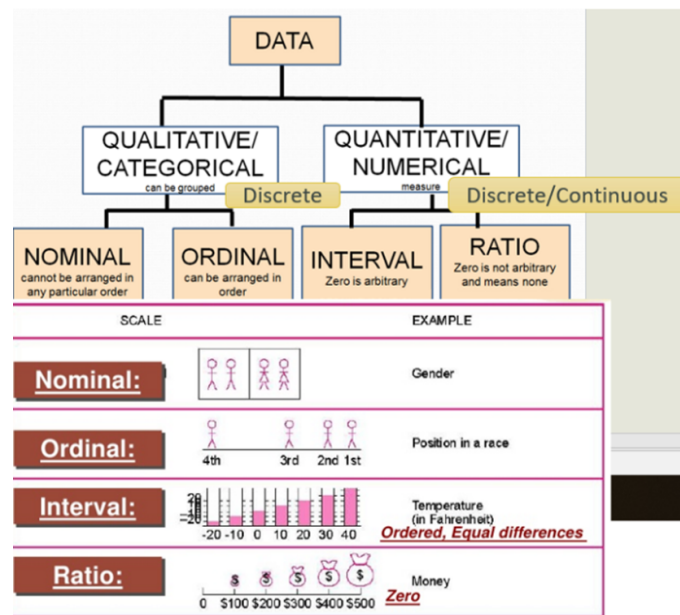
#### Statistik Nonparametrik

Statistik ini digunakan untuk analisa yang tidak menggunakan parameter-parameter populasi biasanya digunakan untuk

mengolah data kualitatif (kategorikal dan kategori bertingkat) yaitu data jenis nominal atau ordinal dengan distribusi data yang tidak diketahui atau tidak normal.

## Data

Data merupakan elemen utama dalam statistik, data dapat dikelompokkan menjadi beberapa kategori tergantung dari jenis datanya.



Jenis-jenis Data:

### 1. Data Kualitatif / Kategorikal (Qualitative / Categorical)

Nominal yaitu data kategori yang tidak memiliki tingkatan dan setara satu sama lain misalnya jenis kelamin, warna rambut, wilayah, golongan darah. Ordinal yaitu data kategori yang memiliki tingkatan dan tidak setara satu sama lain dan memiliki ranking misalnya posisi dalam balap, kategori nilai A-F (A lebih tinggi nilainya dari F), Ya dan Tidak.

### 2. Data Kuantitatif / Numerikal (Quantitative / Numerical)

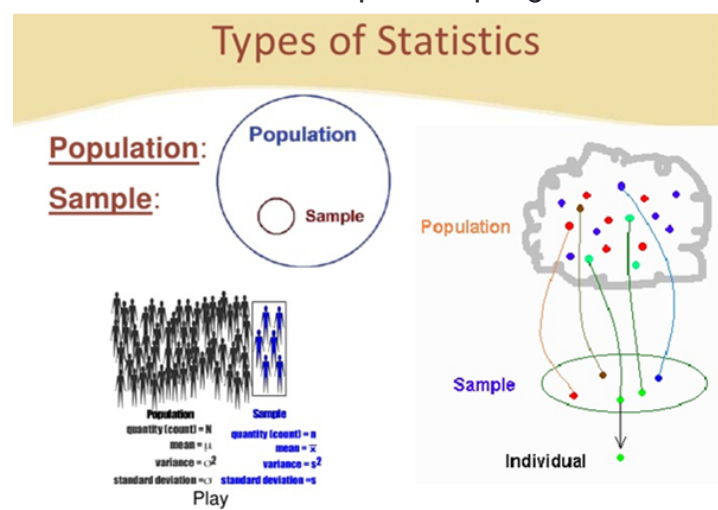
Interval yaitu data numerik yang menunjukkan suatu nilai berjenjang dan bukan menyatakan jumlah tetapi suatu nilai metrik, nilai 0 tidak berarti tidak ada (karena tidak menyatakan jumlah), misalnya nilai suhu / temperatur dimana 0 °C tidak berarti tidak ada suhu tetapi tetap menunjukkan suatu nilai/metrik. Ratio yaitu data numerik yang menunjukkan suatu nilai yang menyatakan jumlah, nilai 0 berarti tidak ada (karena menyatakan jumlah), misalnya nilai uang dimana 0 rupiah berarti tidak ada nilai uang.

Data Kuantitatif / Numerikal bisa berupa :

Diskrit / Discrete, nilai diskrit berarti nilainya berupa bilangan bulat dan jumlahnya biasanya terbatas, misalnya jumlah orang (tidak bisa 0.5 orang), jumlah mobil, dan sebagainya. Kontinu / Continuous, nilai kontinu berarti nilainya bisa berupa bilangan pecahan dan jumlahnya tidak terbatas, misalnya jumlah uang, tinggi badan (bisa 167.53 cm), berat badan, jarak dan sebagainya.

## Populasi dan Sampel

Data untuk statistik didapat dari pengambilan data dari populasi ataupun sampel.



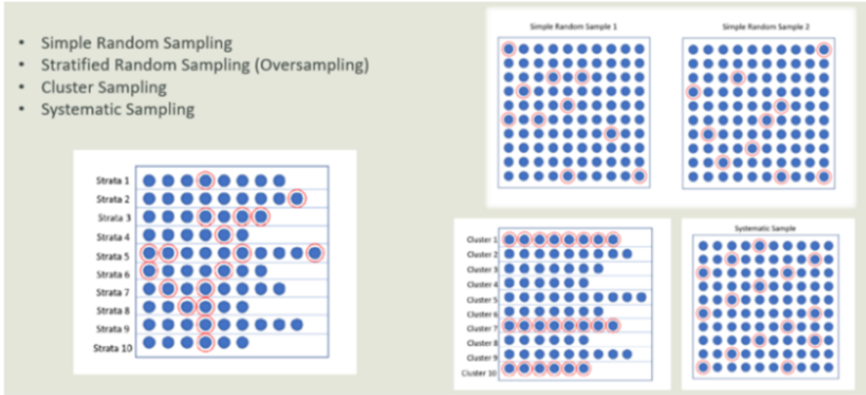
### Populasi

Adalah keseluruhan dari data secara lengkap dari suatu masalah yang sedang dianalisa.

### Sampel

Adalah sebagian dari data yang dianggap mewakili data populasi yang digunakan untuk dianalisa karena sulitnya mendapatkan data jika populasi berjumlah sangat besar.

Pengambilan Sample memiliki banyak metode, berikut metode yang umum digunakan :



Contoh Metode Sampling yang umum

Variabel

Adalah sekumpulan nilai yang seragam dari populasi atau sampel yang diukur dan dianalisa misalnya Variabel tinggi badan, atau berat badan, variabel warna rambut, variabel nilai uang dan sebagainya.

Jenis Analisa Statistik berdasarkan jumlah Variabelnya :

- 1 Variabel / Univariate
- 2 Variabel / Bivariate
- 3 Variabel atau lebih / Multivariate

Analisa yang digunakan tergantung dari jumlah variabel yang ada. Untuk dasarnya akan dibahas univariate terlebih dahulu.

Analisa 1 Variabel / Univariate :

Variabel berisi banyak nilai data yang memiliki rentang tertentu, nilai bisa sama bisa juga berbeda, ada beberapa nilai yang muncul berkali-kali, untuk itu ada namanya frekuensi (jumlah kemunculan).

Pengukur titik tengah :

Mean

(u atau  $\bar{x}$ ) yaitu rata-rata dari data populasi / sampel, contoh diatas 10 data, mean =(139 + 140 + 154 + 154 + 154 + 155 + 180 + 192 + 192 + 196) / 10 = 165

Population Mean	Sample Mean
$\mu = \frac{\sum_{i=1}^N x_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$
$N$ = number of items in the population	$n$ = number of items in the sample

Median

Yaitu titik tengah dari data populasi / sample, contoh diatas 10 data, median = (154+155)/2 = 154.5

Mode / Modus

Yaitu nilai yang paling sering muncul dari data populasi / sample, contoh diatas 10 data, mode adalah 154, mode bisa memiliki lebih dari 1 nilai.

Pengukur sebaran :

Range

Yaitu rentang data dihitung dari nilai tertinggi dikurangi nilai terendah (Max-Min)

Standard Deviation

( $\sigma$  atau  $s$ ) yaitu nilai sebaran distribusi data dihitung dari mean (positif dan negatif), semakin besar nilai SD, semakin lebar rentangnya / semakin tersebar, semakin kecil nilai SD, semakin kecil rentangnya / terfokus di nilai mean, SD juga dipakai sebagai batas kewajaran/normal suatu sebaran. Standard Deviation merupakan nilai akar dari Variance. Contoh 10 data diatas maka SD =  $\sqrt{(((139-165.6)^2 + \dots + (196-165.6)^2)/9)} = 4.336$  atau -4.336

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

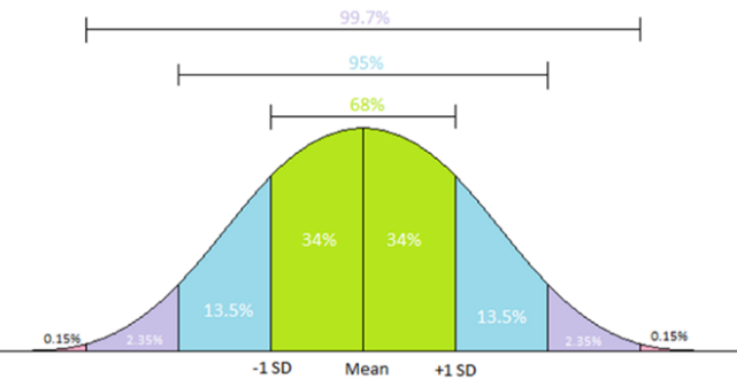
Rumus SD (Populasi)

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

Variance

( $\sigma^2$  atau  $s^2$ ) yaitu nilai kuadrat dari Standard Deviation, mirip dengan SD, variance juga menyatakan nilai positif sebaran dari mean, kuadrat digunakan untuk mempositifkan nilai negatif dan positif tapi memberikan dampak sensitifitas yang lebih besar untuk nilai yang semakin besar.

Distribusi



Frekuensi Distribusi

Gambar diatas adalah sebaran data Distribusi Normal, dimana sumbu x adalah nilai variabel tersebut dan sumbu y adalah nilai frekuensinya. Distribusi disebut normal jika : Berbentuk seperti bel dengan nilai frekuensi terbanyak di tengah (di mean) dan frekuensi kecil pada kedua ujung kiri (min) dan kanan (max).

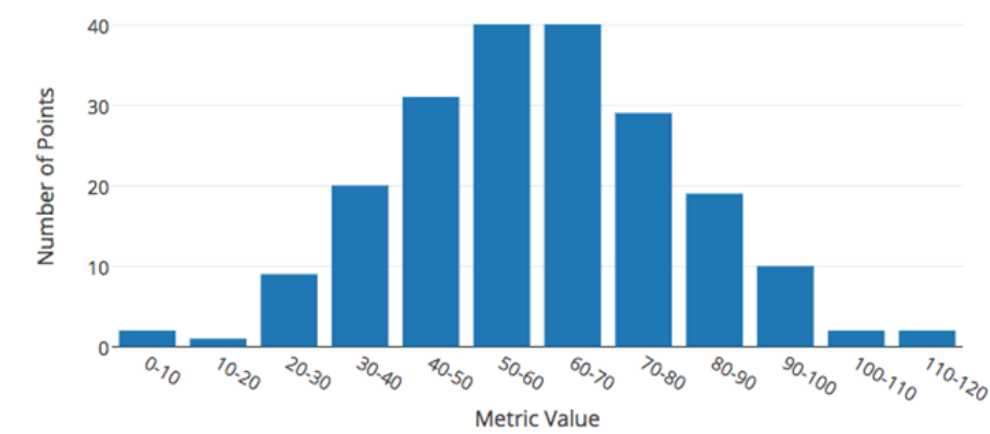
Nilai Mean dan Median dan Mode tidak berbeda jauh.

1 kali nilai SD baik positif maupun negatif mewaliki data kurang lebih sekitar 68% dari keseluruhan, 2 kali nilai SD mewakili kurang lebih 95% dari data keseluruhan. Distribusi normal adalah data paling ideal untuk dilakukan pengujian statistik, jika belum normal ada beberapa cara untuk memperbaiki : Sampling ulang lagi dengan data ditambahkan lagi jumlah samplanya dan dilakukan pengujian ulang.

Membuang outliers yang menjadi penyebab data tidak normal, semudian dilakukan pengujian ulang, pastikan data yang dibuang memang tidak penting bagi pengujian, jika outlier tersebut penting, jangan dibuang. Transformasi data, misalnya mengubah ke bentuk logaritma (log) atau kebentuk natural (Ln) atau bentuk lainnya dan dilakukan pengujian ulang. Jika masih tetap kurang normal / tidak normal berarti demikian keadaan datanya, ada metode lain untuk pengujiannya biasanya dengan nonparametrik, terutama untuk data kategorikal.

Bagaimana dengan data numerikal ratio yang sifatnya kontinu ? misalnya 42.5, 52.4 dan seterusnya dan tidak ada data yang sama, bukankah frekuensi setiap nilai selalu 1 ? untuk data jenis ini, analisa distribusinya sebaiknya dikelompokkan misalnya : 40–50, 50–60, 60–70 dst.

Sehingga kategori 40–50 akan memiliki banyak data frekuensi, dan biasanya dalam bentuk histogram.



Histogram

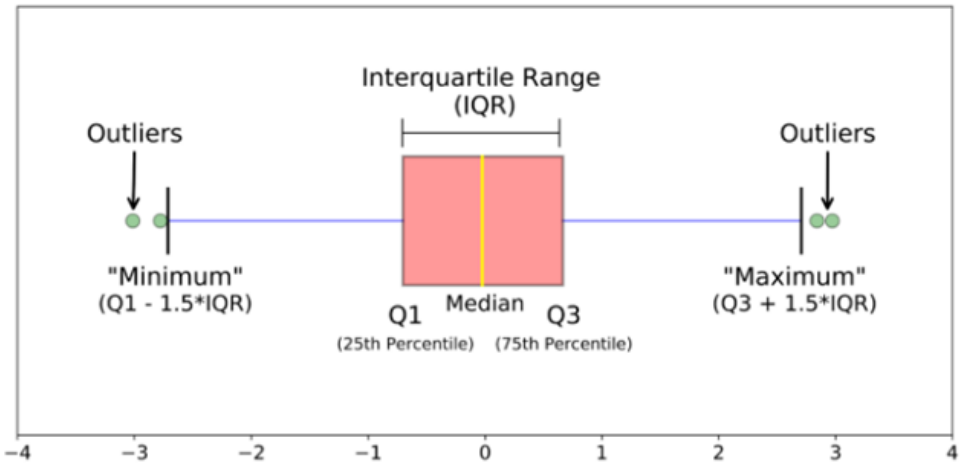
Pengukur lainnya

Pengukur lain yang digunakan umumnya adalah Q1 (Kuartar 1 atau 25%), Q3 (Kuartar 3 atau 75%), Percentile (dibagi 1–99%). Outliers adalah nilai yang menjadi pengganggu analisa distribusi normal, nilai ini bersifat anomali karena berbeda dengan data lainnya dan tentunya dapat merusak rata-rata data. Contoh : data 1,2,2,3,4,4,5,6,1000.

Nilai 1000 adalah outliers karena dapat merusak pola data yang ada, jika dirata-rata :  $(1+2+2+3+4+4+5+6+1000)/9 = 114.11$ , padahal sebaran data nilainya relatif kecil 1–6 tetapi memiliki rata-rata sebesar 114. Data anomali ini sebaiknya tidak disertakan dalam analisa statistik. Bagaimana cara menentukan data mana yang tergolong outliers? Ada beberapa cara salah satunya dengan metode IQR, dan Standard Deviation. Rumus terlampir digambar atas.

Boxplot

Boxplot digunakan untuk melihat sebaran data dengan menampilkan informasi berupa Min, Median, Max, Q1, Q3 serta titik outliers.



Skewness dan Kurtosis

Skewness adalah tingkat kemiringan suatu distribusi, nilai positif berarti puncak berada di kiri dan menurun ke kanan, negatif sebaliknya, digunakan sebagai pengukur juga apakah distribusi masih tergolong normal atau tidak. Skewness normal berada pada rentang -2 sampai 2.

Kurtosis adalah tingkat ketinggian frekuensi serta sebarannya, nilai 0 adalah normal, nilai diatas 0 memiliki kurva tinggi/frekuensi tinggi (data terpusat), dibawah 0 maka kurva landai dan tersebar. Kurtosis normal berada pada rentang -2 sampai 2.

Quiz

KONTAK KAMI

📍 Eduplex Coworking Space, Jln. Ir. H. Juanda Dago no. 84 Bandung, Jawa Barat, Indonesia

☎ +62-8211-6654-087

✉ bisaaimail@gmail.com

IJIN PENYELENGGARAAN

LEMBAGA PELATIHAN KERJA (LPK)

### BISA AI ACADEMY

PROGRAM

- MASTERCLASS ON JOB TRAINING GARANSI KERJA
- MENTORSHIP PERSIAPAN RISET ILMU KOMPUTER S2/S3
- PERSIAPAN DAN SERTIFIKASI OKUPASI BNSP

- MASTERCLASS ON JOB TR
- MAGANG PREMIUM
- PELATIHAN EXCLUSIVE

Jl. Pasir Salam I No.5B, Ancol, Kec. Regol, Kota Bandung, Jawa Barat

Website : <https://bisa.ai>

Email : [bisaaimail@gmail.com](mailto:bisaaimail@gmail.com)

CASE STUDY

- [Webinar](#)  
[Kompetisi](#)  
[Freelance](#)  
[Bootcamp](#)  
[Diskusi Private](#)
- [Kunjungan Industri](#)  
[Job Fair](#)  
[Event Sosial](#)  
[Ujian](#)  
[Master Class + OJT](#)

OFFICIAL PARTNERSHIP

