



Institut Informatika & Bisnis
DARMAJAYA
Yayasan Alifian Husin



**Kampus
Merdeka**
INDONESIA JAYA

**MERDEKA
BELAJAR**

Exploratory Data Analysis

Session 2

SSD23407

Egi Safitri, S.Mat., M.Si



Teknik Pengumpulan Data

Observasi

- Observasi adalah pengamatan yang melibatkan semua indera (penglihatan, pendengaran, penciuman, peraba, perasa).
- Observasi dapat dilakukan secara langsung untuk mengumpulkan data empiris.

Contoh: Guru menggunakan indera penglihatan dan pendengaran untuk mengamati interaksi siswa.

Wawancara

- **Wawancara tidak terstruktur:**
 - Pertanyaan spontan dan alami.
 - Digunakan untuk menggali informasi lebih mendalam.
- **Wawancara terstruktur:**
 - Pertanyaan sudah disiapkan sebelumnya.
 - Digunakan untuk mendapatkan informasi yang lebih spesifik.
- **Jenis wawancara:**
 - Wawancara langsung (face to face).
 - Wawancara tidak langsung (telepon, online).

Teknik Pengumpulan Data

Kuisisioner

- Kuesioner adalah daftar pertanyaan tertulis yang diajukan kepada responden.
- Efisien untuk mendapatkan data jika variabel yang diukur sudah jelas.
- Dapat dilakukan secara langsung maupun online.

Pengukuran Fisik

- Data fisik berupa benda nyata, seperti tanah, bangunan, atau buku.
- Pengukuran dilakukan dengan menggunakan angka (meter, liter, dll).

Teknik Pengumpulan Data

Percobaan Laboratorium

- Pengumpulan data dilakukan di lingkungan laboratorium.
- Faktor-faktor dikendalikan secara ketat untuk menghindari bias pada hasil percobaan.
- Contoh: Mengukur pengaruh pH terhadap pertumbuhan tanaman.

Penyajian Data

- Penyajian data berfungsi untuk memberikan gambaran awal dari hasil pengumpulan data, membantu analisis data.
- Contoh: Penelitian pengaruh desain website pada kesediaan konsumen membeli produk.
- Peneliti dapat menggunakan berbagai metode penyajian data tergantung pada sifat data.

Tahapan Penyajian Data

- Tahapan penyajian data:
 1. Editing: Memeriksa kesalahan data mentah.
 2. Koding: Menempatkan angka atau simbol untuk kategori tertentu.
- Metode penyajian: Tabel dan grafik, untuk informasi yang lebih jelas dan informatif.

Tabel Satu Arah (One-Way Table)

- Tabel ini menunjukkan satu variabel atau karakteristik saja.
- Contoh: Jumlah responden berdasarkan umur.

Kategori Umur	Jumlah Responden
< 20 tahun	10
21 - 30 tahun	15
31 - 40 tahun	25
> 40 tahun	10
Total	60

Tabel Dua Arah (Two-Way Table)

- Menggabungkan dua variabel untuk menyajikan data lebih kompleks.
- Contoh: Jumlah responden berdasarkan umur dan jenis kelamin.

Kategori Umur	Laki-laki	Perempuan
< 20 tahun	5	5
21 - 30 tahun	12	3
31 - 40 tahun	10	15
> 40 tahun	2	8
Total	29	31

Tabel Frekuensi

- Tabel yang menunjukkan jumlah data dalam kelas-kelas tertentu.
- Langkah-langkah:
 1. Tentukan range, jumlah kelas, lebar interval.
 2. Hitung frekuensi dan frekuensi kumulatif.

Kategori Umur	Frekuensi	Frekuensi Kumulatif	Frekuensi Relatif
19 - 26 tahun	4	4	0,20
27 - 34 tahun	3	7	0,35
35 - 42 tahun	5	12	0,60
43 - 50 tahun	6	18	0,90
51 - 58 tahun	2	20	1,00
Total	20	20	1,00

Grafik

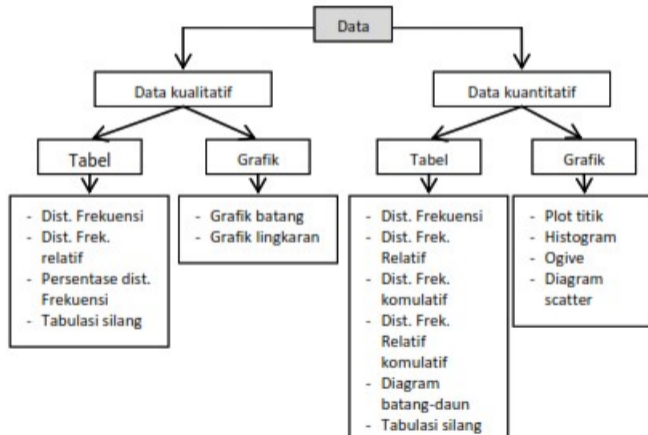
- Meringkas data, baik data kualitatif maupun kuantitatif.
- Data kualitatif berupa distribusi frekuensi, frekuensi relatif, persen distribusi, grafik batang, dan grafik lingkaran.
- Data kuantitatif berupa distribusi frekuensi, relatif frekuensi, persen distribusi frekuensi, diagram plot titik, distribusi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

Prosedur Penggunaan Tabel dan Grafik

Grafik adalah representasi visual dari data atau informasi yang disusun untuk mempermudah pemahaman pola, tren, hubungan, atau distribusi data. Grafik digunakan untuk menyajikan data secara lebih jelas dan intuitif, terutama ketika data tersebut terdiri dari banyak titik atau memiliki variasi yang kompleks.

- Tabel dan grafik berbeda penggunaannya, tergantung dari jenis data (kualitatif atau kuantitatif).
- Data kualitatif lebih cocok disajikan menggunakan distribusi frekuensi dan grafik batang atau lingkaran.
- Data kuantitatif dapat disajikan dengan distribusi frekuensi, histogram, ogive, plot titik, dan scatter diagram.

Prosedur Penggunaan Tabel dan Grafik



Latihan

Angka berikut menunjukkan hasil survei curah hujan tahunan pada suatu wilayah (dalam mm) selama 30 tahun antara tahun 1960-1990:

123	117	83	140	97	110
117	86	116	79	130	63
93	95	108	98	119	84
136	87	91	107	122	74
98	80	82	90	125	105
97					

- Buat tabel frekuensi dari data di atas!
- Kelompokkan data sesuai interval yang relevan.
- Hitung jumlah data dalam setiap interval untuk membentuk tabel frekuensi.

Box plot & Fungsinya

Box Plot, atau yang juga dikenal sebagai diagram kotak-garis, adalah representasi grafis yang digunakan untuk menampilkan distribusi data numerik berdasarkan kuartil. Diagram ini memberikan ringkasan visual tentang sebaran data, seperti rentang, median, serta kemungkinan outlier. Fungsi utama box plot adalah:

- Menampilkan Distribusi Data: Box Plot memungkinkan kita untuk melihat bagaimana data tersebar di seluruh rentang nilai.
- Mendeteksi Outlier: Box Plot sangat efektif untuk mendeteksi outlier (data yang terletak jauh dari kelompok nilai lainnya).
- Memahami Posisi Median: Median (nilai tengah) ditampilkan dengan jelas dalam Box Plot sebagai garis di dalam kotak.
- Membandingkan Kelompok Data: Box Plot sering digunakan untuk membandingkan sebaran data.

Penggunaan Umum Box Plot

1. Analisis Data Eksploratif: Untuk mendapatkan gambaran awal distribusi data.
2. Mendeteksi Skewness: Memahami apakah data simetris atau miring.
3. Perbandingan Dataset: Memudahkan perbandingan antara beberapa set data yang berbeda, misalnya distribusi nilai antara dua kelompok siswa atau hasil eksperimen dari beberapa percobaan.

Elemen-Elemen Box Plot

Median(Q2)

Median adalah nilai tengah data yang telah diurutkan. Pada box plot, median ditunjukkan oleh garis horizontal di dalam kotak. Median memisahkan data menjadi dua bagian yang sama besar. Setengah dari data berada di bawah median, dan setengahnya berada di atas median.

Kuartil 1 (Q1)

Kuartil 1 (Q1) adalah nilai yang membagi 25% pertama dari data setelah diurutkan. Ini merupakan nilai tengah dari setengah pertama data. Q1 membantu menunjukkan batas bawah distribusi data. Pada box plot, Q1 merupakan tepi bawah kotak.

Kuartil 3 (Q3)

Kuartil 3 (Q3) adalah nilai yang memisahkan 75% data pertama dari data yang diurutkan. Ini adalah nilai tengah dari setengah kedua data. Q3 menunjukkan batas atas distribusi data. Pada box plot, Q3 adalah tepi atas kotak.

Elemen-Elemen Box Plot

Rentang antar kuartil (IQR)

IQR adalah jarak antara Kuartil 1 (Q_1) dan Kuartil 3 (Q_3). Rumusnya adalah:

$$\text{IQR} = Q_3 - Q_1$$

IQR digunakan untuk mengukur penyebaran data di sekitar median. Semakin besar IQR, semakin tersebar data tersebut. Kotak pada box plot menunjukkan IQR.

Whisker

Whiskers adalah garis vertikal di luar kotak yang memanjang hingga batas minimum dan maksimum data, tetapi tidak termasuk outlier. Whiskers menunjukkan seberapa jauh data tersebar sebelum dianggap outlier. Batas atas whisker adalah nilai terbesar yang masih berada dalam IQR, dan batas bawah whisker adalah nilai terkecil yang masih dalam IQR. Rumus whisker:

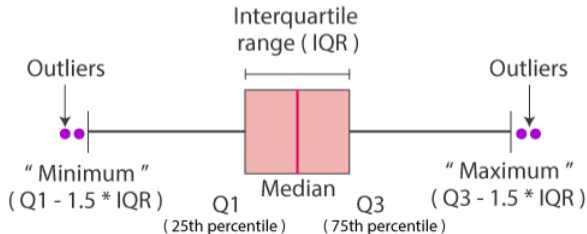
$$\text{Max} = Q_3 + 1.5 * \text{IQR}$$

$$\text{Min} = Q_1 - 1.5 * \text{IQR}$$

Elemen-Elemen Box Plot

Outlier

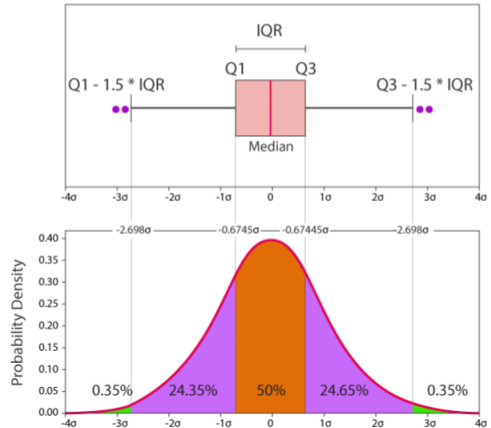
Outliers adalah data yang berada di luar whiskers, yaitu nilai-nilai yang secara signifikan lebih besar atau lebih kecil daripada data lain. Outliers dapat menunjukkan data yang tidak biasa atau kesalahan dalam data. Dalam box plot, outliers sering ditunjukkan dengan titik di luar whiskers.



Distribusi Box Plot

Distribusi kotak plot akan menjelaskan seberapa rapat data dikelompokkan, bagaimana data tersebut miring, dan juga tentang simetri data.

- Miring Positif : Jika jarak dari median ke nilai maksimum lebih besar daripada jarak dari median ke nilai minimum, maka diagram kotak miring positif.
- Miring Negatif : Jika jarak dari median ke minimum lebih besar daripada jarak dari median ke maksimum, maka diagram kotak miring negatif.
- Simetris : Kotak plot dikatakan simetris jika median berjarak sama dari nilai maksimum dan minimum.



Manfaat Box Plot

Boxplots dapat membantu kita dalam memahami karakteristik dari distribusi data. Selain untuk melihat derajat penyebaran data (yang dapat dilihat dari tinggi/panjang boxplot) juga dapat digunakan untuk menilai kesimetrisan sebaran data. Panjang kotak menggambarkan tingkat penyebaran atau keragaman data pengamatan, sedangkan letak median dan panjang whisker menggambarkan tingkat kesimetrisannya.

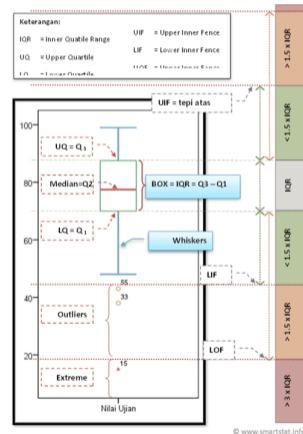
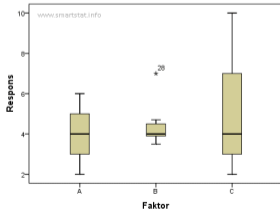
- Jika data simetris (berasal dari distribusi normal):
 - garis median akan berada di tengah box dan whisker bagian atas dan bawah akan memiliki panjang yang sama serta tidak terdapat nilai outlier ataupun nilai ekstrim.
 - diharapkan nilai-nilai pengamatan yang berada di luar whiskers tidak lebih dari 1%.

Manfaat Box Plot

- Jika data tidak simetris (miring), median tidak akan berada di tengah box dan salah satu dari whisker lebih panjang dari yang lainnya.
 - Adanya outlier di bagian atas boxplot yang disertai dengan whisker bagian atas yang lebih panjang, menunjukkan bahwa distribusi data cenderung menjulur ke arah kanan (positive skewness).
 - Sebaliknya, adanya outlier di bagian bawah boxplot yang disertai dengan whisker bagian bawah yang lebih panjang, menunjukkan bahwa distribusi data cenderung menjulur ke arah kiri (negatif skewness).

Kemiringan (Skewness)

Keunggulan Boxplots dibanding dengan Histogram, dotplot, dan stemplot sangat terasa pada saat kita ingin membandingkan sebaran beberapa kelompok data secara bersamaan. Sebagai contoh, perhatikan gambar berikut:



Soal 1

- Diberikan data berikut mengenai nilai ujian dari sekelompok siswa:

65, 70, 75, 80, 85, 90, 95, 100, 105, 110

- Langkah-langkah:
 1. Urutkan data dari yang terkecil hingga terbesar.
 2. Tentukan nilai minimum, kuartil pertama (Q1), median (Q2), kuartil ketiga (Q3), dan nilai maksimum.
 3. Gambarkan Box Plot berdasarkan hasil yang ditemukan.

Soal 2

- Diberikan data berat badan (dalam kg) dari 12 orang berikut:

45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100

- Buatlah Box Plot berdasarkan data ini dengan mengikuti langkah-langkah berikut:
 1. Urutkan data dari yang terkecil hingga terbesar.
 2. Tentukan nilai minimum, Q1, Q2 (median), Q3, dan nilai maksimum.
 3. Gambarkan Box Plot menggunakan informasi tersebut.

Soal 3

- Berikut adalah data tinggi badan (dalam cm) dari 10 siswa:

150, 155, 160, 165, 170, 175, 180, 185, 190, 210

- Buatlah Box Plot dari data ini dan identifikasilah apakah terdapat outlier.
- Langkah-langkah:
 1. Hitung Q1, Q2, dan Q3.
 2. Hitung IQR (Interquartile Range).
 3. Tentukan apakah ada data yang berada di luar $1.5 \times \text{IQR}$ dari Q1 atau Q3.

Thank You!



Institut Informatika & Bisnis
DARMAJAYA
Yayasan Alfian Husin



**Kampus
Merdeka**
INDONESIA JAYA

**MERDEKA
BELAJAR**