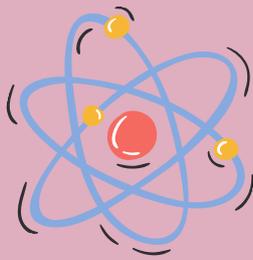


# Modul Pembelajaran

## Angkatan 59



### STATISTIKA DAN ANALISIS DATA

2022

Ekspresi Karya dan Legacy SR 58-Kabinet Altair  
Tutor Sebaya

# KATA PENGANTAR

Puji dan syukur kepada Allah, Tuhan Yang Maha Esa, karena berkat rahmat dan karunia-Nya, sehingga kami berhasil menyelesaikan “Modul Pembelajaran Angkatan 59”. Sholawat serta salam semoga senantiasa tercurahkan kepada Nabi Muhammad SAW.

Modul Pembelajaran Angkatan 59 ini berisi rangkuman cantik (rantik) tiap pekan yang bersumber dari materi kuliah Matematika dan Berpikir Logis, Fisika, Kimia ST, Biologi, Statistika dan Sains Data, Ekonomi, serta Kalkulus 1 yang telah diajarkan oleh Dosen dan Tim Pengajar pada perkuliahan tingkat Pendidikan Kompetensi Umum (PKU). Terdiri dari 2 sesi, yaitu sesi pra uts dan sesi pra uas.

Rangkuman cantik ini ditata ulang desain oleh Lini Club Tutor Sebaya Senior Resident Asrama Mahasiswa PKU. Rantik ini akan terus di perbaharui sesuai dengan materi kuliah.

Semoga Rantik ini dapat mempermudah mahasiswa Gatamala Agrinawa/Angkatan 59 memahami materi kuliah terutama 7 matakuliah tersebut.

Terima kasih kepada semua pihak yang telah membantu menyelesaikan menyusun rantik ini. Kritik, masukan serta saran tentunya sangat diperlukan untuk menyempurnakan rantik ini.

Sebagai bentuk apresiasi karya, kiranya tidak menggandakan rantik ini tanpa seizin tim penyusun rantik.

Penulis

Tim Penyusun

# PENYUSUN

**Pengarah:**

Toni Bakhtiar

**Penanggungjawab :**

Suratni

**Sekretariat**

1. Tri Suswantoro
2. Dadi Ahdi

**Supporting Staff**

1. Mega Silvianty
2. Irayanti Wahyuningsih

**Tim Penyusun Materi**

Statistika dan Analisis Data :

1. Handito Rahman
2. Jesika Br Girsang
3. Muhammad Zhafran Iffat Taqy Arif
4. Putri Qoonitah Dewi
5. Taffyana Izza Alifa

*Together we walk, we learn, we create  
the better version of education*

*-Tutor Sebaya 58-*

# SUMBER MODUL PEMBELAJARAN

Sumber modul pembelajaran yang dirangkum dan ditata ulang oleh Lini Club Tutor Sebaya Asrama Mahasiswa PKU berasal dari materi yang disampaikan oleh dosen/tim pengajar pada Perkuliahan Tingkat I Pendidikan Kompetensi Umum Angkatan 58 pada Semester Ganjil dan Genap 2021/2022.

Modul Pembelajaran yang terdiri dari 7 mata kuliah pokok, yaitu:

1. Biologi
2. Ekonomi
3. Fisika
4. Kimia ST
5. Kalkulus 1
6. Matematika dan Berpikir Logis
7. Statistika dan Analisis Data

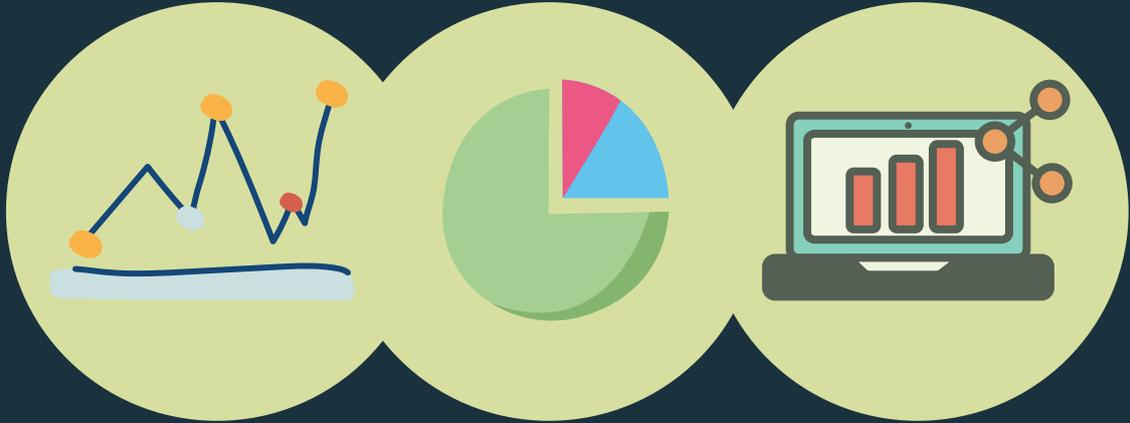
*Together we walk, we learn, we create  
the better version of education*

*-Tutor Sebaya 58-*



# **STATISTIKA**

## **SESI PRA UTS**



# RANTIK STATISTIKA

PEKAN 1 - SEMESTER 2



# Konsep Dasar Statistika

## Mengapa harus belajar Statistika?

★ Aktivitas keseharian manusia sejatinya merupakan aktivitas memproduksi data

- Data ada di mana-mana
- Data ada di sekitar kita
- Data perlu dimanfaatkan

Data → Diolah → Bermanfaat

## Pengertian Statistika

Statistika adalah ilmu yang mempelajari seluk-beluk data, terutama bagaimana:

- Cara merancang atau mengumpulkan
- Cara menganalisis atau memodelkan
- Cara menarik kesimpulan

## Populasi, Parameter, Sampel, dan Statistik

- Populasi → Keseluruhan objek/pengamatan yang menjadi pusat perhatian. ex: Seluruh Mahasiswa PPKU
- Parameter → Karakteristik numerik dari populasi. ex: Rata-rata pengeluaran seluruh Mahasiswa PPKU
- Sampel → Himpunan bagian dari populasi (mewakili). ex: Sebagian Mahasiswa PPKU yang menjadi objek pengukuran
- Statistik → Karakteristik numerik dari sampel. ex: Rata-rata pengeluaran Mahasiswa PPKU dari data sampel, misal kelas parallel 19
- Statistik merupakan penduga bagi parameter

## Peubah dan Skala Pengukuran

Peubah → Karakteristik dari objek yang diamati

Skala pengukuran peubah :

1. Nominal (kategorik): jenis kelamin, agama
2. Ordinal (kategorik): tingkat pendidikan, skala likert
3. Interval (numerik): skor intelegensi, suhu
4. Rasio (numerik): umur, berat badan



## Statistika Deskriptif vs Statistika Inferensia



Statistika Deskriptif	Statistika Inferensia
<p>Mempelajari teknik-teknik yang berguna dalam peringkasan data dan pemberian gambaran umum tentang data yang dimiliki.</p>	<p>Mempelajari kaidah-kaidah pengambilan keputusan statistika dari data yang dimiliki dengan menggunakan ilmu peluang.</p>
<p>Penyajian data (grafik, tabel).</p>	<p>Pendugaan (Menduga Pendapatan perkapita penduduk Indonesia menggunakan data pendapatan sampel)</p>
<p>Peringkasan data (min, max, mean, median, modus, ragam, simpangan baku).</p>	<p>Pengujian hipotesis (Menguji kebenaran anggapan yang menyatakan bahwa pendapatan perkapita penduduk Indonesia adalah lebih dari Rp 3 juta per bulan).</p>

### Apa itu Data?

★ Data merupakan gugus/himpunan hasil pengukuran karakteristik dari seluruh objek yang menjadi focus penelitian.

Data menurut periode pengukuran dibedakan menjadi:

- Data Cross-section: merupakan kumpulan hasil pengukuran terhadap beberapa objek data dan periode waktu tertentu.
- Data Time series: merupakan kumpulan hasil pengukuran terhadap suatu objek, yang dicatat mengikuti urutan waktu.
- Data panel: merupakan kumpulan hasil pengukuran terhadap beberapa objek, yang dicatat mengikuti urutan waktu.
- Data space-time adalah merupakan hasil pengukuran suatu objek, yang diamati pada beberapa lokasi dan mengikuti urutan waktu.



## Pengumpulan Data

1. Percobaan/ experiment
  - Data dibangkitkan dengan memberikan perlakuan terhadap suatu objek
  - Sering diterapkan pada penelitian pertanian/IPA/psikologi
2. Survei
  - Data sudah tersedia di alam, tinggal dikumpulkan
  - Banyak diterapkan pada bidang sosial ekonomi / non IPA
3. Administratif
  - Data yang diperoleh dari hasil pencatatan administrasi suatu institusi
  - Misal: Data transaksi perbankan dan data transaksi supermarket

Selain itu, ada data yang tersedia secara natural

- Data aktivitas manusia yang terekam secara otomatis
- Misalnya :
  - aktivitas pengguna aplikasi smartphone, browser, media sosial, belanja online, nonton youtube, dll
  - aktivitas jalan kaki juga dapat terekam (oleh GMaps, Google Fit, Samsung Health, dsb)
  - aktivitas yang terekam oleh satelit, radar, lalu lintas (oleh CCTV-nya NTMC Polri)

## Big Data

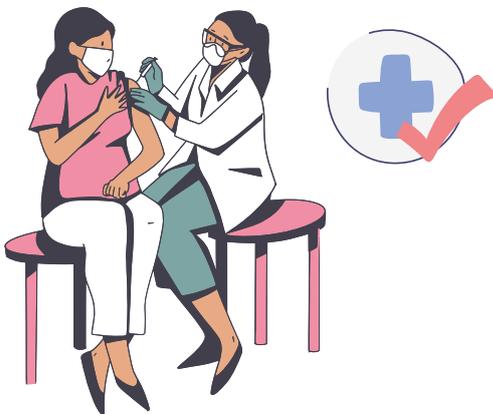
Data sangat besar dan kompleks, tidak dapat dikelola (capture, store, manage, analyze) dengan software dan tools pemrograman database biasa/konvensional sehingga dibutuhkan skill baru dalam analisis data

- Volume (Ukuran data sangat besar)
- Variety (banyak jenisnya : text, gambar, suara)
- Velocity (Berubah cepat, perlu analisis real time)

## Penerapan Statistika

### Bidang bisnis:

- Analisis tentang produk baru
- Prediksi target penjualan
- Analisis ketercapaian sebuah iklan
- Mengukur kerja unit/karyawan tertentu

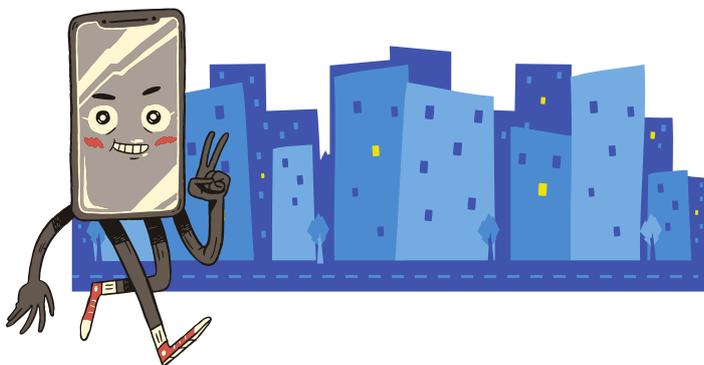


### Bidang Kesehatan:

- Uji klinis vaksin Covid-19
- Cek darah
- Prediksi penyakit dari DNA
- Pengujian obat

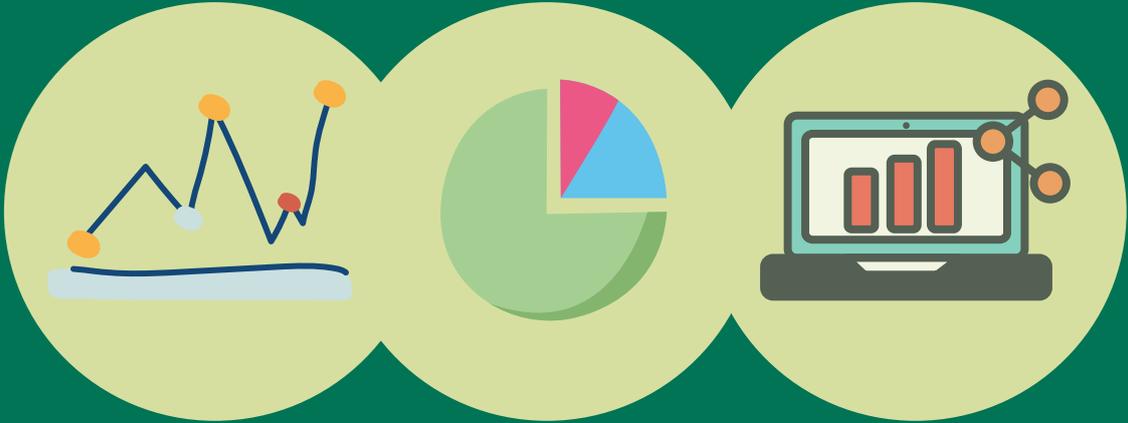
### Bidang politik:

- Survei opini public
- Quick count
- Penentuan titik kampanye kandidat kepala daerah



### Bidang big data:

- Media sosial
- Smart city



# RANTIK STATISTIKA

## PEKAN 2



# Data Understanding

## DESKRIPSI-PENYAJIAN DATA

### Review

Jenis data berdasarkan tipenya :

1. Kategorik
2. Numerik (diskrit dan kontinu)

Skala Pengukuran Peubah :

- Data Kuantitatif : Data Rasio, Data Interval
- Data Kualitatif : Data Ordinal, Data Nominal

### Penyajian Data

- Data dalam **bentuk mentah** biasanya tidak mudah digunakan untuk pengambilan keputusan, tetapi harus diolah terlebih dahulu dan disajikan ke audiens.
- Beberapa bentuk Penyajian yang dibutuhkan antara lain :
  - Ø Tabel
  - Ø Grafik / Diagram / Chart
- Tipe grafik yang digunakan tergantung dari peubah yang diringkaskan.

### Grafik Penyajian Data Berdasarkan Peubahnya

#### 1. Peubah Kategorik

- Tabel Distribusi Frekuensi
- Tabulasi Silang
- Diagram Batang
- Pie Chart
- Diagram Pareto

#### 2. Peubah Numerik

- Diagram Garis
- Tabel Distribusi Frekuensi
- Histogram
- Ogive
- Diagram Batang dan Daun
- Plot Pencar

## I. Peubah Kategorik

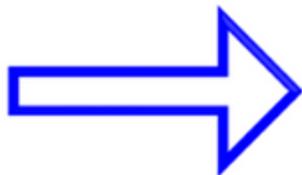
### - Tabel Distribusi Frekuensi

Tabel distribusi frekuensi merupakan alat penyajian data statistika yang berbentuk kolom dan baris yang didalamnya terdapat susunan data yang sudah dikelompokkan berdasarkan kategori tertentu. (<https://www.facebook.com/rory.stats>, 2021)

#### Contoh :

Dalam Suatu Survei, Mahasiswa baru Kelas X Universitas Y ditanyakan mengenai Kendala yang dihadapi ketika Kuliah dilaksanakan secara Online. Didapatkan respon sebagai berikut.

RESPONDEN	KENDALA
1	Sinyal
2	Sinyal
3	Sinyal
4	Kuota
5	Device (laptop/hp)
6	Kuota
7	Sinyal
8	Kuota
9	Kuota
...	...
...	...
...	...
125	Kuota
126	Device (laptop/hp)



Tabel Distribusi Frekuensi Kendala Kuliah Online Mahasiswa Baru Kelas X Universitas Y

KENDALA	JUMLAH
Device (laptop/hp)	10
Kuota	42
Sinyal	67
Tidak ada	3
Website terlalu banyak	1
Listrik	1
Jarang online	1



Peubahnya berupa Kategori

### - Tabulasi Silang (Cross Table)

Cross Tables (atau juga dikenal dengan Tabel Kontingensi) memuat daftar jumlah amatan dari setiap kombinasi dari nilai dua peubah kategorik. Jika ada r kategori untuk peubah pertama (baris) dan c kategori untuk peubah kedua (kolom), maka table tersebut dinamakan  $r \times c$  cross table.

#### Contoh :

Dalam Suatu Survei, Mahasiswa baru Kelas X Universitas Y selain ditanyakan mengenai Kendala yang dihadapi ketika Kuliah dilaksanakan secara Online. Juga ditanyakan mengenai Lokasi mereka berada. Didapatkan respon sebagai berikut.

RESPONDEN	KENDALA	LOKASI
1	Sinyal	Jawa
2	Sinyal	Sumatera
3	Sinyal	Jawa
4	Kuota	Jawa
5	Device (laptop/hp)	Jawa
6	Kuota	Sumatera
7	Sinyal	Jawa
8	Kuota	Jawa
9	Kuota	Jawa
...	...	...
...	...	...
...	...	...
125	Kuota	Jawa
126	Device (laptop/hp)	Jawa



		LOKASI					TOTAL
		Jawa	Kalimantan	Papua	Sulawesi	Sumatera	
KENDALA	Device (laptop/hp)	9				1	10
	Jarang online	1					1
	Kuota	29	1		3	9	42
	Listrik	1					1
	Sinyal	48		1	3	15	67
	Tidak ada	3					3
	Website terlalu banyak	1					1
	TOTAL	92	1	1	6	25	125

**- Diagram Batang**

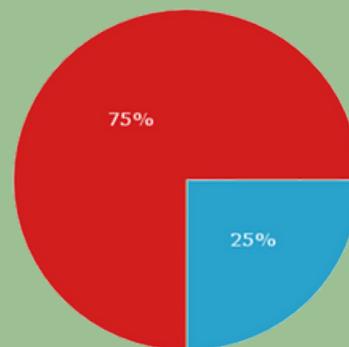
Diagram Batang sering digunakan untuk meringkas data kategorik (kualitatif) . Tinggi batang menunjukkan frekuensi atau persentase tiap kategori



(bobo.grid.id)

**- Pie Chart**

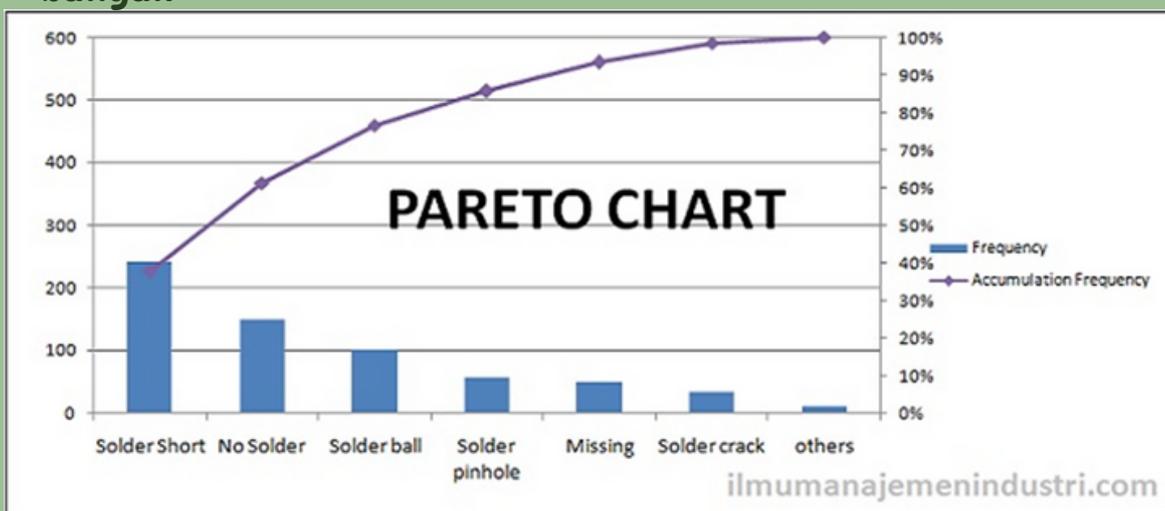
Ukuran 'potongan pie' menunjukkan frekuensi atau persentase tiap kategori.



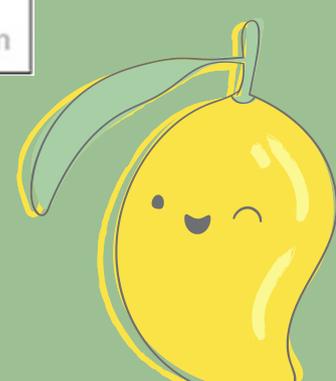
(sederet.com)

**- Diagram Pareto**

- Diagram batang, di mana kategori ditampilkan dalam urutan frekuensi yang menurun
- Poligon kumulatif sering kali ditampilkan dalam grafik yang sama
- Digunakan untuk memisahkan "beberapa penting" dari "sepele banyak"



(ipqi.com)



## I. Peubah Numerik

### - Diagram Garis

Line chart (Plot Deret Waktu) digunakan untuk menunjukkan nilai-nilai peubah seiring berjalannya waktu. Waktu ditampilkan pada axis horizontal, sedangkan Peubah yang sedang dibahas ditampilkan pada axis vertikal.



### - Tabel Distribusi Frekuensi

Selain digunakan dengan peubah kategorik, tabel distribusi frekuensi juga dapat digunakan dengan peubah numerik, dengan ketentuan sebagai berikut:

- Dalam pembuatan Distribusi Frekuensi dikenal istilah Class Interval dan Class Boundaries, dimana
- Setiap Kelas selalu memiliki lebar (Class Interval) yang sama.
- Class Interval dihitung dengan membagi selisih nilai maksimum dan nilai minimum dengan jumlah kelas yang diinginkan.

$$w = \text{interval width} = \frac{\text{max} - \text{min}}{\text{jumlah kelas yang diinginkan}}$$

- Jumlah kelas sebaiknya lebih dari 5 namun tidak lebih dari 10-15 atau bisa juga menggunakan Sturgis's Rule dimana jumlah kelas =  $3.3 \log(n) + 1$ , dimana n adalah jumlah data.
- Class Interval tidak pernah tumpang tindih.
- Class Interval dibulatkan sehingga memudahkan pembuatan Class interval.

DATA	INTERVAL
12	10 ≤ x < 20
13	
17	
21	20 ≤ x < 30
24	
24	
26	
27	
27	30 ≤ x < 40
30	
32	
35	
37	
38	40 ≤ x < 50
41	
43	
44	
46	50 ≤ x < 60
53	
58	

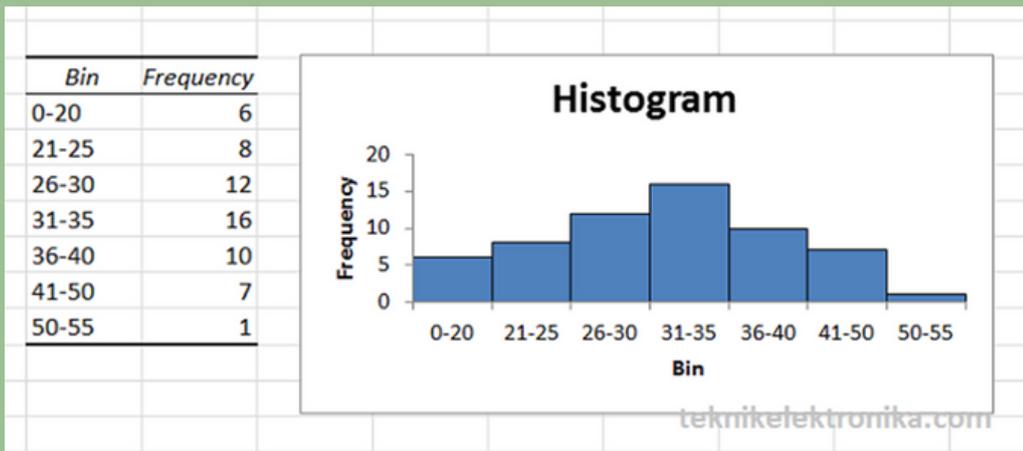
  

INTERVAL	FREKUENSI	FREKUENSI RELATIF	PERSENTASE
10 ≤ x < 20	3	0.15	15
20 ≤ x < 30	6	0.30	30
30 ≤ x < 40	5	0.25	25
40 ≤ x < 50	4	0.20	20
50 ≤ x < 60	2	0.10	10
<b>Total</b>	<b>20</b>	<b>1.00</b>	<b>100</b>



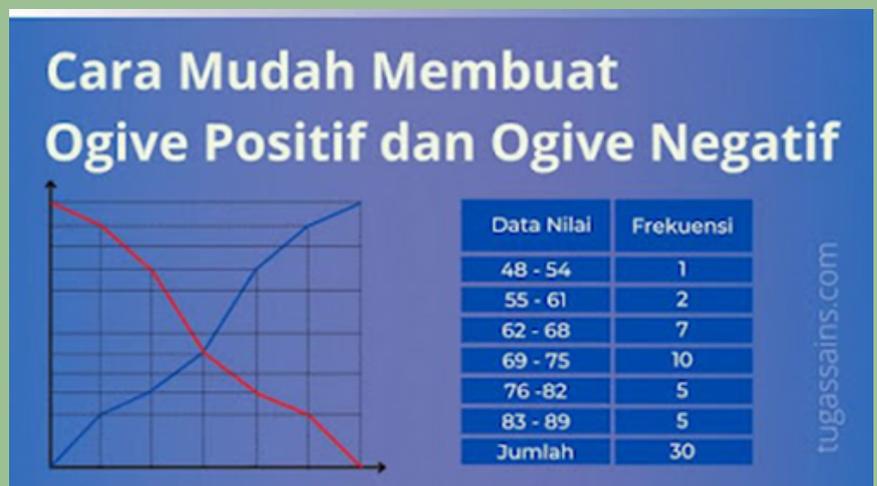
**- Histogram**

- Grafik dari Tabel Distribusi Frekuensi disebut histogram.
- Axis Horizontal menampilkan garis akhir interval (interval endpoints)
- Axis Vertikal bisa berupa frekuensi, frekuensi relatif, dan persentase.
- Batang dengan tinggi yang sesuai digunakan untuk mewakili jumlah amatan yang ada di setiap kelas.



**- Ogive**

Ogive dalam Statistika merupakan grafik yang tersusun dari Tabel Distribusi Frekuensi Kumulatif. Ogive sendiri memiliki dua jenis, yaitu Ogive positif (grafik ke atas) dan Ogive negatif (grafik ke bawah). (naufal al majid, 2021)



(tugassains.com)

**- Diagram Batang dan Daun (Steam and Leaf Diagram)**

Diagram Batang dan Daun adalah salah satu cara mudah untuk melihat sebaran data secara detil suatu set data.

Langkah : Pisahkan seri data yang diurutkan menjadi digit terdepan (batang) dan digit berikutnya (daun)



Contoh : Data yang telah diurutkan : 21, 24, 24, 26, 27, 27, 30, 32, 38, 41

	DAHAN	DAUN
21 ditampilkan sebagai	2	1
38 ditampilkan sebagai	3	8

Diagram Dahan Daun secara Lengkap

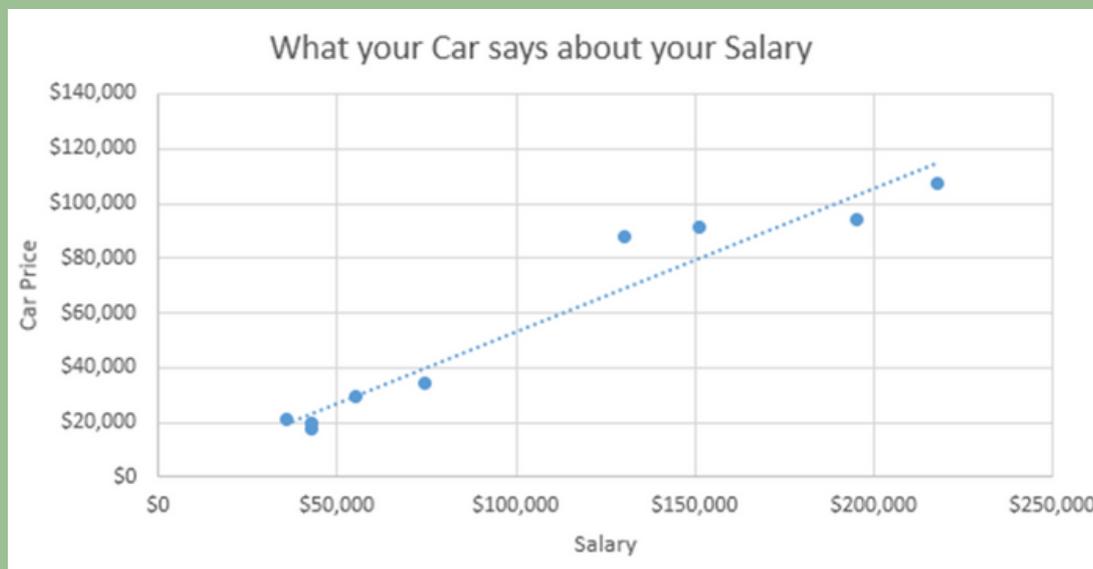
	DAHAN	DAUN
(6)	2	1 4 4 6 7 7
4	3	0 2 8
1	4	1

N = 10    Leaf Unit : 1.0

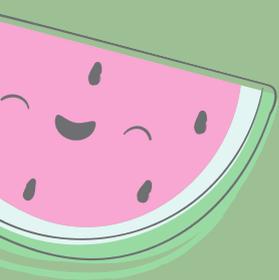
### - Plot Pencar (Scatter Plot)

Scatter Plot digunakan untuk observasi berpasangan yang diambil dari dua peubah numerik.

Peubah yang satu ditaruh di axis vertikal dan yang lain di axis horizontal.

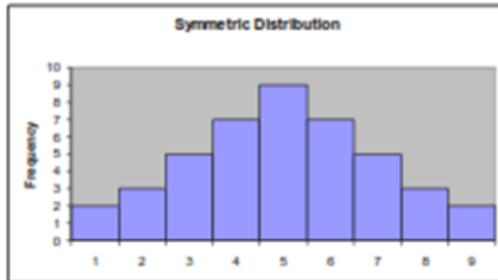


([excel-easy.com](http://excel-easy.com))



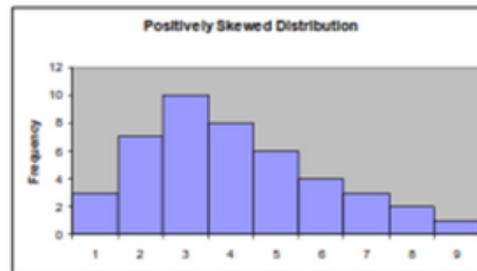
## Bentuk Distribusi Data

Bentuk distribusi datanya dikatakan simetris jika pengamatan seimbang atau merata ke arah tengah/ pusat.

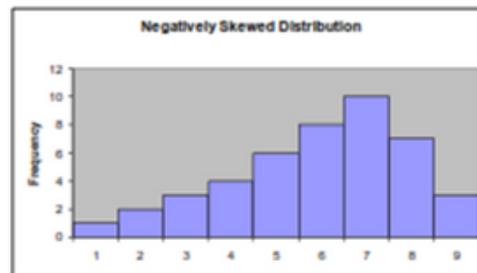


Sedangkan, bentuk distribusi data dikatakan miring (asimetris) jika observasi tidak didistribusikan secara simetris di sekitar pusat. Berikut adalah contoh bentuk distribusi data asimetris.

Ø Miring positif (miring ke kanan) : memiliki ekor yang memanjang ke kanan ke arah nilai positif.



Ø Miring negatif (miring ke kiri) : memiliki ekor yang memanjang ke kiri ke arah nilai negatif.



Sumber :

<https://www.facebook.com/rory.stats> (2021). *Tabel Distribusi Frekuensi*. [online] Rumus Statistik. Available at: [https://www.rumusstatistik.com/2021/04/tabel-distribusi-](https://www.rumusstatistik.com/2021/04/tabel-distribusi-frekuensi.html#:~:text=Tabel%20distribusi%20frekuensi%20adalah%20alat,telah%20di)

[frekuensi.html](https://www.rumusstatistik.com/2021/04/tabel-distribusi-frekuensi.html#:~:text=Tabel%20distribusi%20frekuensi%20adalah%20alat,telah%20di)#:~:text=Tabel%20distribusi%20frekuensi%20adalah%20alat,telah%20di kelompokkan%20menurut%20kategori%20tertentu. [Accessed 3 Feb. 2022].

naufal al majid (2021). *Cara Mudah Membuat Ogive Positif dan Ogive Negatif*. [online] tugassains.com. Available at: <https://www.tugassains.com/2021/11/cara-mudah-membuat-ogive-positif-dan-negatif.html> [Accessed 3 Feb. 2022].



**IPB University**  
— Bogor Indonesia —



# Data Understanding

**RANTIK STATISTIKA  
PEKAN 3**

# Ukuran Pemusatan Data

Menggambarkan posisi pusat data



## MEAN

- Nilai rata-rata
- Menimbang data menjadi 2 kelompok yang memiliki massa yang sama
- Dipengaruhi oleh nilai ekstrim (pencilan)

• Untuk **POPULASI** sejumlah  $N$  :

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

← Nilai-Nilai Amatan pada Populasi  
← Jumlah Populasi

• Untuk **SAMPEL** sejumlah  $n$  :

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

← Nilai-Nilai Amatan pada Sampel  
← Jumlah Sampel

## MEDIAN

- Membagi data yang sudah diurutkan menjadi 2 bagian yang sama banyak
- Tidak dipengaruhi oleh nilai pencilan (outliers)

• Jika  $n_{med}$  bulat, maka  $Median = X_{\frac{(n+1)}{2}}$

• Jika  $n_{med}$  pecahan, maka  $Median = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$

## KUARTIL (Q1,Q2,Q3)

- Membagi data yang telah diurutkan menjadi 4 bagian, dengan jumlah data per segmen sama.
- Cara mencari kuartil

### Metode Belah dua

- Urutkan data dari kecil ke besar
- Cari posisi kuartil
  - $n_{Q2} = (n+1)/2$
  - $n_{Q1} = (n_{Q2} + 1)/2 = n_{Q3}$ ,  $n_{Q2}$  posisi kuartil dua terpangkas (pecahan dibuang)
- Nilai kuartil 2 ditentukan sama seperti mencari nilai median. Kuartil 1 dan 3 prinsipnya sama seperti median tapi kuartil 1 dihitung dari kiri, sedangkan kuartil 3 dihitung dari kanan.

### Metode Interpolasi

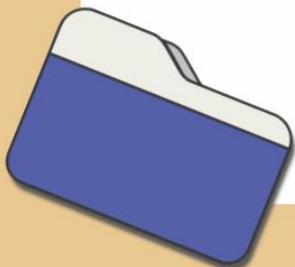
- Urutkan data dari kecil ke besar
- Cari posisi kuartil
  - $n_{q1} = (1/4)(n+1)$
  - $n_{q2} = (2/4)(n+1)$
  - $n_{q3} = (3/4)(n+1)$
- Nilai kuartil dihitung sebagai berikut:
  - $X_q = X_{a_i} + h_i (X_{b_i} - X_{a_i})$
  - $X_{a_i}$  = pengamatan sebelum posisi kuartil ke  $i$ ,  $X_{b_i}$  = pengamatan setelah posisi kuartil ke  $i$  dan  $h_i$  adalah nilai pecahan dari posisi kuartil

## MODUS

- Nilai yang paling sering muncul
- Tidak dipengaruhi pencilan
- Unimodus (memiliki 1 modus), bimodus (memiliki 2 modus), multi modus (memiliki 3 modus)

**NOTE** : Jika tidak terdapat pencilan, ukuran pemusatan data yang umum digunakan adalah mean

Jika terdapat pencilan, maka median yang digunakan



# Ukuran Penyebaran Data

Memberikan informasi tentang sebaran atau keragaman nilai-nilai suatu data



## RANGE (JANGKAUAN)

- Penyebaran data paling sederhana
- Penyebaran amatan terbesar dan terkecil ( $R = X_{\max} - X_{\min}$ )
- Tidak memperhatikan pola distribusi data
- Sensitif dengan pencilan

## JANGKAUAN INTERKUARTIL (IQR)

- Dapat mengatasi pencilan
- Hanya memperhitungkan 50% data yang berada di tengah
- $IQR = Q_3 - Q_1$

## RAGAM

- Ragam populasi : rata-rata dari kuadrat jarak antara nilai dan mean

Population variance: 
$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

Dimana  $\mu$  = population mean  
 $N$  = population size  
 $x_i$  =  $i^{\text{th}}$  value of the variable  $x$

- Ragam contoh : rata-rata (pendekatan) dari kuadrat jarak antara nilai amatan dan mean

Sample variance: 
$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Dimana  $\bar{X}$  = arithmetic mean  
 $n$  = sample size  
 $x_i$  =  $i^{\text{th}}$  value of the variable  $x$

## SIMPANGAN BAKU

- Standard Deviation
- Ukuran penyebaran data yang paling sering digunakan
- Menunjukkan keragaman terhadap mean
- Memiliki satuan yang sama dengan data awal
- Tiap nilai pada data digunakan dalam perhitungan
- Nilai yang jauh dari mean diberikan bobot yang lebih (karena selisih dari mean dikuadratkan)

Population standard deviation :

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

Sample standard deviation:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

## The Emperical Rule

Jika data menyebar normal, maka :

- ada 68% data di antara  $\mu \pm \sigma$
- ada 95% data di antara  $\mu \pm 2\sigma$
- Ada 99,7% data di antara  $\mu \pm 3\sigma$



## COEFFICIENT OF VARIATION

- Mengukur variasi relatif antara ragam terhadap mean
- Membandingkan 2 atau lebih set data yang diukur dalam satuan yang berbeda

$$CV = \left( \frac{s}{\bar{x}} \right) \cdot 100\%$$

## COVARIANCE (KORAGAM)

- Keragaman bersama antara peubah satu dengan peubah lainnya
- Mengukur kekuatan hubungan linear antara dua peubah, semakin besar nilai koragam menunjukkan hubungan linier kedua peubah semakin kuat
- Tidak mengindikasikan hubungan sebab akibat
- $Cov(x,y) > 0 \rightarrow x$  dan  $y$  cenderung bergerak ke arah yang sama
- $Cov(x,y) < 0 \rightarrow x$  dan  $y$  cenderung bergerak ke arah berlawanan
- $Cov(x,y) = 0 \rightarrow x$  dan  $y$  saling bebas

• Covariance Populasi:

$$Cov(x,y) = \sigma_{xy} = \frac{\sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)}{N}$$

• Covariance Contoh:

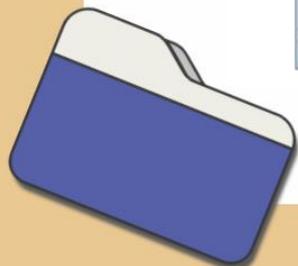
$$Cov(x,y) = s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

## CONTOH SOAL

Ragam contoh selalu lebih besar dari ragam populasi karena rumus ragam contoh yang menjadi pembagi adalah  $(n-1)$ . (Benar/Salah)

Jawaban : Salah

Pembahasan : Meskipun pembagi ragam contoh adalah  $(n-1)$ , bisa saja ragam populasi lebih besar karena perbedaan nilai pada data.





IPB University  
— Bogor Indonesia —



# Data

# Understanding

Explorasi Data

RANTIK STATISTIKA  
PEKAN 4



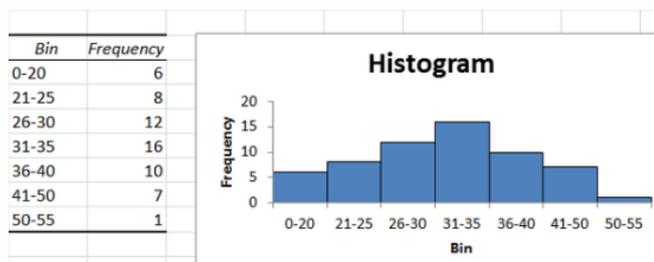
# Eksplorasi Data



**Eksplorasi Data** merupakan langkah awal sebelum analisis data dilakukan dengan memvisualisasikan data. Berfungsi untuk lebih memahami karakteristik dari data, agar lebih mudah eksplorasi data bisa divisualisasikan dalam bentuk grafik, *chart*, atau format visual lainnya.

## HISTOGRAM

- Pengertian : Tampilan bentuk desain grafis untuk menunjukkan distribusi data secara visual. (Ilmu Manajemen Industri, 2021)



- Fungsi Histogram :

1. Melihat ukuran penyebaran dan ukuran pemusatan data
2. Melihat adanya data *outlier* (pencilan)
3. Mendeteksi adanya bimodus (2 nilai data yang paling sering muncul) atau tidak.

## BOXPLOT

- *Boxplot* merupakan suatu kotak yang digunakan dalam statistik deskriptif untuk menggambarkan secara grafik dari data numeris melalui lima ukuran sebagai berikut:

- Nilai observasi terkecil,
- Kuartil terendah atau kuartil pertama (Q1), yang memotong 25% dari data terendah

- Median (Q2) atau nilai pertengahan,
- Kuartil tertinggi atau kuartil ketiga (Q3), yang memotong 25%



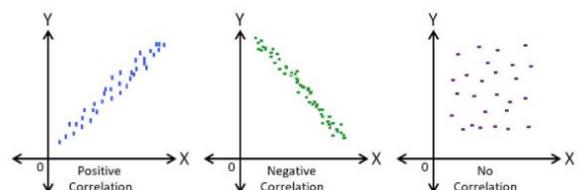
- Fungsi *Boxplot* :

- Melihat ukuran penyebaran dan ukuran pemusatan data.
- Melihat adanya data pencilan.
- Sebagai alat perbandingan sebaran dua kelompok data atau lebih.

## SCATTER PLOT

- Scatter Plot adalah grafik yang menggunakan titik untuk mewakili nilai dua peubah numerik yang berbeda. Scatter plot biasanya digunakan untuk mengamati hubungan antar variabel.

Scatter Plots & Correlation Examples



- Posisi setiap titik pada sumbu horizontal dan vertikal menunjukkan nilai untuk satu titik data serta menunjukkan korelasi pada data yang bersangkutan.

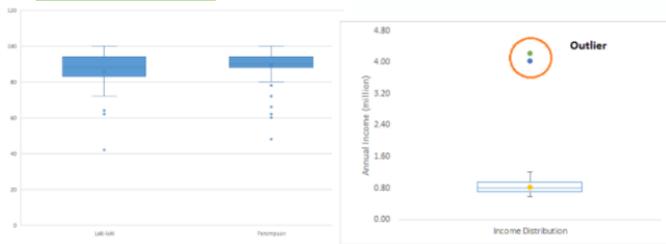
# Eksplorasi Kualitas Data



## IDENTIFIKASI NILAI EKSTREM/PENCILAN/OUTLIER

- Nilai ekstrem (atau dikenal sebagai 'pencilan' [outlier]) adalah titik data yang tersebar di ekor distribusi suatu data.
- Nilai-nilai ekstrim tersebut akan memengaruhi analisis statistik. Untuk mendeteksi nilai outlier, kita dapat menggunakan metode sebagai berikut :

### BOXPLOT



Catatan

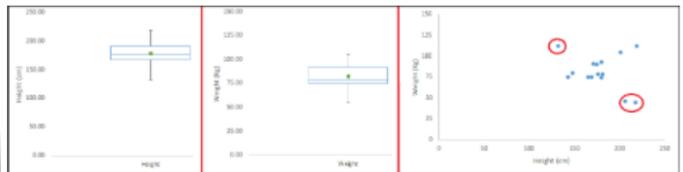
Untuk mencari nilai pencilan, maka gunakan rumus berikut (digunakan untuk menghitung pencilan secara manual dengan *boxplot*):

1. Statistik lima serangkai : Min, Q1, Q2, Q3, Max
2. Pagar Dalam Atas (PDA) :  $Q3 + 1.5(Q3 - Q1) = Q3 + 1.5(IQR)$
3. Pagar Dalam Bawah (PDB) :  $Q1 - 1.5(Q3 - Q1) = Q1 - 1.5(IQR)$
4. Identifikasi data : Jika data ada yang kurang dari PDB atau lebih dari PDA maka data dikatakan memiliki Pencilan ( $IQR = Interquartile Range = Q3 - Q1$ )

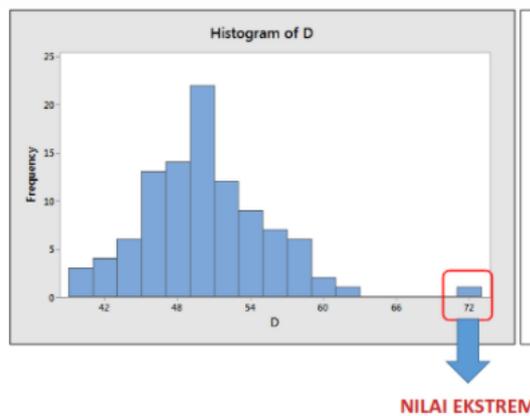
### SCATTER PLOT

Pencilan dapat terdiri dari dua jenis: **Univariat** dan **Multivariat**

Misalkan kita memahami hubungan antara tinggi dan berat. Dari kedua boxplot untuk tinggi dan berat badan, tidak terlihat adanya pencilan. Namun ketika kedua peubah divisualisasikan dengan menggunakan Scatter plot, terlihat adanya pencilan. Dua nilai di bawah dan satu di atas rata-rata dalam segmen berat dan tinggi tertentu.



### HISTOGRAM

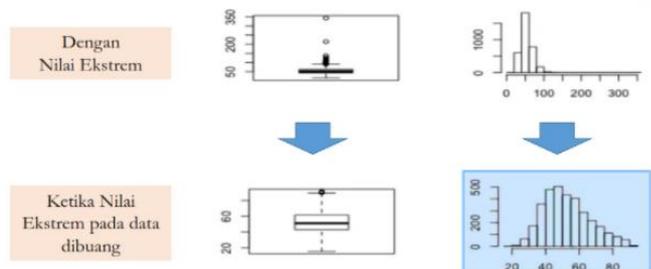


Pencilan yang terlihat dari diagram menggambarkan batang histogram terjal.

### GABUNGAN HISTOGRAM DAN BOXPLOT

Pencilan diidentifikasi sebagai nilai terbesar dalam kumpulan data dan muncul sebagai lingkaran di sebelah kanan boxplot maupun histogram.

- Pencilan harus diselidiki dengan hati-hati.
- Seringkali pencilan berisi informasi berharga tentang proses yang sedang diselidiki atau proses pengumpulan dan pencatatan data.
- Sebelum mempertimbangkan kemungkinan penghapusan pencilan dari data, terlebih dahulu harus dicoba untuk dipahami mengapa pencilan tersebut muncul dan apakah kemungkinan nilai serupa akan terus muncul.



# Eksplorasi Pola Sebaran Data



## HISTOGRAM

**Histogram** : Salah satu alat yang sering digunakan untuk melihat distribusi dari suatu data.

- Untuk distribusi yang menjulur, arah kemenjulurannya ditunjukkan oleh arah ekor yang lebih panjang.
- Untuk distribusi yang menjulur ke kanan, ekor panjang meluas ke kanan sementara sebagian besar nilai mengelompok di sebelah kiri (Histogram of % Fat).
- Untuk distribusi yang menjulur ke kiri sebaliknya (Histogram of Left Skew).





IPB University  
— Bogor Indonesia —



# Data

# Understanding

Explorasi Data

RANTIK STATISTIKA  
PEKAN 5

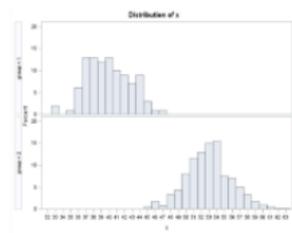
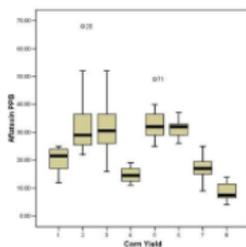


# Eksplorasi Perbandingan



## Antar Grup

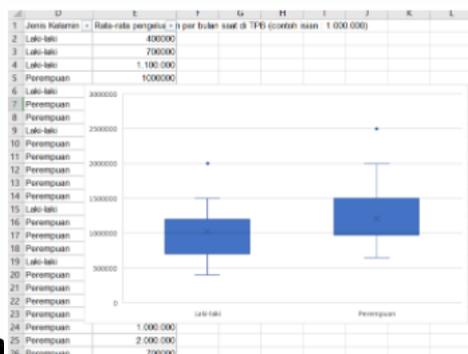
- Eksplorasi perbandingan antar grup dapat dilakukan dengan menggunakan Boxplot dan Histogram serta Statistik lima serangkai. Membuat histogram yang dikelompokkan pada dasarnya membuat histogram individual secara terpisah untuk setiap grup dan menempatkannya pada sumbu yang sama dan menggunakan lebar kotak yang sama. Skala yang digunakan harus identik pada sumbu y maupun sumbu x sehingga kita dapat membandingkan nilai di seluruh grup yang dibandingkan.



**Statistik Lima Serangkai:**  
1. Minimum  
2. Kuartil pertama  
3. Median  
4. Kuartil ketiga  
5. Maksimum

- Saat ingin membandingkan beberapa grup dengan histogram, ada beberapa hal yang perlu dibandingkan yaitu
  1. Bentuk sebaran data
  2. Nilai rata-rata, median, dan modus dari masing-masing grup
  3. Nilai pencilan
  4. Variasi dari sebaran grup

## Contoh Boxplot

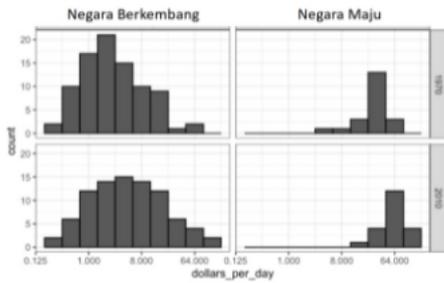


Boxplot ini menunjukkan bahwa median pengeluaran per bulan mahasiswa TPB perempuan lebih tinggi dari mahasiswa TPB laki-laki. Jarak antar kuartil dari kedua boxplot cukup mirip (seperti yang ditunjukkan oleh panjang kotak), meskipun terlihat bahwa secara keseluruhan boxplot untuk mahasiswa TPB perempuan lebih besar dibandingkan mahasiswa TPB laki-laki. Boxplot mahasiswa TPB perempuan tampak menjulur ke kanan, sementara boxplot mahasiswa TPB laki-laki cenderung simetris. Di kedua boxplot terlihat ada pencilan atas. Secara keseluruhan, Median pengeluaran per bulan mahasiswa TPB perempuan lebih besar dari kuartil atas pengeluaran per bulan mahasiswa TPB laki-laki. Jadi dapat disimpulkan bahwa pengeluaran per bulan mahasiswa TPB ada kaitannya dengan jenis kelamin.



## Contoh Histogram

Sebaran Pendapatan Tahun 1970 vs Tahun 2010 di negara berkembang dan di negara maju

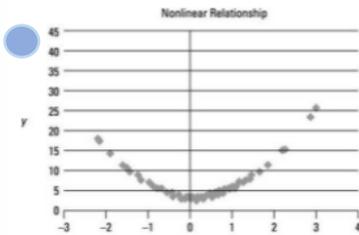


Pada negara maju (Western Europe and North America), income menjadi sedikit lebih meningkat pada tahun 2010 dibandingkan tahun 1970, tetapi secara persentase, income negara negara berkembang tampaknya lebih meningkat, terjadi pergeseran ke kanan pada tahun 2010 dibandingkan tahun 1970. Dari histogram di samping terlihat juga bahwa proporsi income negara berkembang yang berpenghasilan lebih dari \$ 16 per hari meningkat secara substansial.

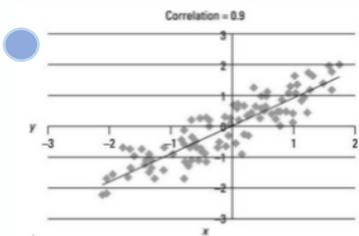


# Eksplorasi Hubungan

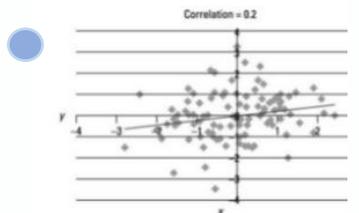
## Antar Peubah



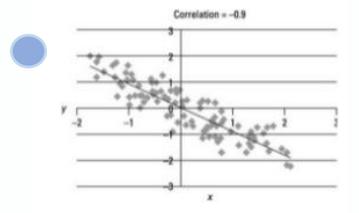
-Scatter plot (juga disebut diagram pencar) digunakan untuk melihat hubungan antara dua peubah. Sebagai contoh, scatter plot disamping menunjukkan diagram pencar untuk dua peubah yang memiliki hubungan nonlinier di antara keduanya. Setiap titik pada diagram pencar mewakili satu pasangan (X, Y). Karena diagram pencar bukan garis lurus, hubungan antara X dan Y adalah nonlinier. Perhatikan bahwa dimulai dengan nilai X yang paling negatif, dengan meningkatnya X, Y pada awalnya menurun; lalu ketika X terus meningkat, Y meningkat.



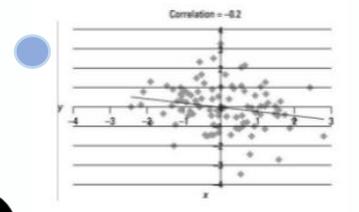
-Scatter plot menunjukkan hubungan dua peubah yang memiliki hubungan linier positif yang kuat. Scatter plot tersebut menunjukkan kecenderungan yang sangat kuat untuk X dan Y, semakin naik nilai X maka nilai Y juga akan semakin meningkat. Garis lurus adalah garis trend, dirancang sedekat mungkin dengan semua titik data. Garis tren memiliki kemiringan positif, yang menunjukkan hubungan positif antara X dan Y.



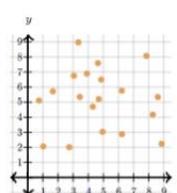
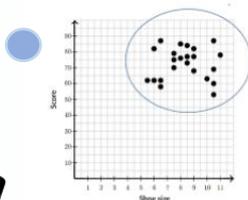
-Scatter plot berikutnya menunjukkan hubungan dua peubah yang memiliki hubungan linier positif lemah. Perhatikan bahwa titik-titik pada grafik lebih tersebar di sekitar garis trend daripada di gambar sebelumnya, karena hubungan yang lebih lemah antara X dan Y.



-Scatter plot menunjukkan hubungan dua peubah yang memiliki hubungan linier negatif yang kuat. Scatter plot tersebut menunjukkan kecenderungan yang sangat kuat untuk X dan Y dengan arah yang berlawanan, semakin naik nilai X maka nilai Y semakin menurun. Garis tren memiliki kemiringan negatif, yang menunjukkan hubungan negatif antara X dan Y.

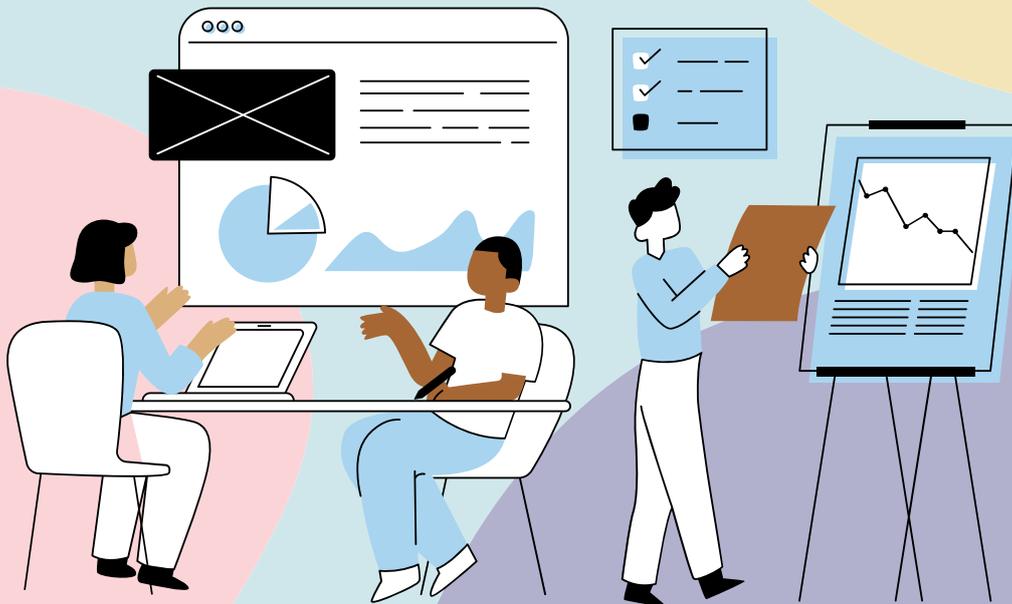


-Scatter plot berikutnya menunjukkan hubungan dua peubah yang memiliki hubungan linier negatif lemah. Perhatikan bahwa titik-titik pada grafik lebih tersebar di sekitar garis trend daripada di gambar sebelumnya, karena hubungan yang lebih lemah antara X dan Y.



Jika tidak ada hubungan yang jelas antara kedua peubah, maka dapat dikatakan tidak ada korelasi antara kedua peubah tersebut.





# STATISTIKA

RANTIK PEKAN 6

DATA BIVARIATE

### **Data Bivariate**

- **Peubah atau variable adalah setiap karakteristik dari objek dalam populasi yang akan dianalisis.**
- **Data adalah suatu nilai (dapat berupa kategorik atau kuantitatif) yang diukur untuk suatu peubah.**
- **Jika kita hanya memiliki satu peubah yang diukur, maka kita menyebutnya data univariate.**
- **Jika dua peubah diukur, maka kita menyebutnya data bivariate**
- **Tiga kombinasi data bivariate:**
  - 1. Kedua peubahnya kategorik**
  - 2. Satu peubah kategorik dan satu peubah lainnya numerik**
  - 3. Kedua peubahnya numerik**

#### **Contoh 1 (Dua Peubah Kategorik)**

- **Sebuah survei dilakukan untuk menyelidiki hubungan antara preferensi responden untuk mendapatkan informasi/berita melalui berbagai media informasi (televisi, radio, dan surat kabar) dan jenis kelamin.**
- **Media informasi adalah peubah kategorik dengan tiga kategori (televisi, radio, dan surat kabar)**
- **Jenis kelamin adalah peubah kategorik (laki-laki dan perempuan)**

#### **Contoh 2 (Satu Peubah Kategorik dan Satu Peubah Numerik)**

- **Misalkan diambil 10 contoh acak rumah tangga dari setiap enam kecamatan di Bogor yang berbeda dan dicatat tagihan listrik mereka pada bulan Juni 2021.**
- **Kecamatan atau lokasi adalah peubah kategorik dengan enam kategori (bogor barat, bogor timur, bogor tengah, bogor utara, bogor selatan dan tanah sereal)**
- **Besaran tagihan listrik setiap rumah tangga adalah peubah numerik**

#### **Contoh 3 (Kedua peubahnya numerik)**

- **Dalam sebuah penelitian ingin dilihat hubungan usia dengan tingkat ketakutan anak-anak ketika dirawat di rumah sakit. Skor tingkat ketakutan diukur dengan menggunakan Skala Ketakutan Medis Anak (Child Medical Fear Scale/CMFS).**
- **Usia dan skor tingkat ketakutan, keduanya merupakan peubah numerik**



### Hubungan dua peubah kategorik

- Untuk melihat hubungan antar dua peubah kategorik dapat menggunakan:
- Tabel frekuensi dua arah (two ways frequently table) atau disebut juga tabulasi silang (cross tables) atau disebut juga klasifikasi silang (cross-classification)
- Ukuran keeratan hubungan antar peubah kategorik, sering disebut sebagai ukuran asosiasi. Salah satu ukuran asosiasi adalah Chi-square

### Table Frekuensi Dua Arah atau Tabel Kontingensi

- Tabel Kontingensi adalah tabel yang menyajikan jumlah pengamatan untuk setiap kombinasi nilai untuk dua peubah kategorik
- Jika terdapat  $r$  kategori untuk peubah pertama (baris) and  $c$  kategori untuk peubah kedua (kolom), maka table ini disebut sebagai tabulasi silang  $r \times c$  atau table kontingensi  $r \times c$

### $r \times c$ Tabel Kontingensi

Kategori dari peubah A	Kategori dari peubah B				Totals
	1	2	...	C	
1	$O_{11}$	$O_{12}$	...	$O_{1c}$	$R_1$
2	$O_{21}$	$O_{22}$	...	$O_{2c}$	$R_2$
.	.	.	...	.	.
.	.	.	...	.	.
r	$O_{r1}$	$O_{r2}$	...	$O_{rc}$	$R_r$
Totals	$C_1$	$C_2$	...	$C_c$	n

### Contoh 1: Media Informasi Vs Jenis Kelamin

Dari hasil survey terhadap 1320 responden, diperoleh sebaran responden menurut jenis media informasi yang diakses dan jenis kelamin sebagai berikut:

	TV	Radio	Surat Kabar	Total Baris
Laki-laki	280	175	305	760
Perempuan	115	275	170	560
<b>Total Kolom</b>	<b>395</b>	<b>450</b>	<b>475</b>	<b>1320</b>

Frekuensi mutlak (absolute frequently)

Informasi yang dapat diperoleh dari table kontingensi:

- Frekuensi relatif bersama (joint relative frequency)
- Frekuensi relatif marginal (marginal relative frequency)
- Frekuensi relatif bersyarat (conditional relative frequency)

	TV	Radio	Surat Kabar	Total Baris
Laki-laki	280	175	305	760
Perempuan	115	275	170	560
<b>Total Kolom</b>	<b>395</b>	<b>450</b>	<b>475</b>	<b>1320</b>

Frekuensi relatif bersama memberikan persentase atau proporsi individu yang memiliki nilai khusus untuk satu variabel kategoris dan nilai spesifik untuk variabel kategoris lainnya.

	TV	Radio	Surat Kabar	Total Baris
Laki-laki	21,2%	13,3%	23,1%	
Perempuan	8,7%	20,8%	12,9%	
Total kolom				100%

Persentase responden mendapatkan informasi dari radio dan berjenis kelamin laki-laki didapatkan dari  $175/1320=13,3\%$

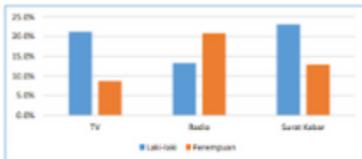
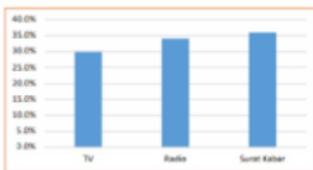
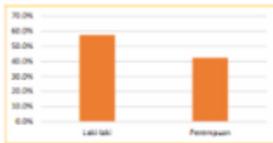
	TV	Radio	Surat Kabar	Total Baris
Laki-laki	280	175	305	760
Perempuan	115	275	170	560
<b>Total Kolom</b>	<b>395</b>	<b>450</b>	<b>475</b>	<b>1320</b>

Frekuensi relatif marginal memberikan persentase atau proporsi individu yang memiliki nilai khusus untuk satu variabel kategoris (baris) atau kolom)

	TV	Radio	Surat Kabar	Total Baris
Laki-laki				57,6%
Perempuan				42,4%
<b>Total kolom</b>	<b>29,9%</b>	<b>34,1%</b>	<b>36,0%</b>	<b>100%</b>

persentase responden yang mendapatkan informasi dari media TV didapatkan dari  $395/1320=29,9\%$

	TV	Radio	Surat Kabar	Total Baris
Laki-laki	21,2%	13,3%	23,1%	57,6%
Perempuan	8,7%	20,8%	12,9%	42,4%
Total Kolom	29,9%	34,1%	36,0%	100%



	TV	Radio	Surat Kabar	Total Baris
Laki-laki	280	175	305	760
Perempuan	115	275	170	560
Total Kolom	395	450	475	1320

Marginal and joint relative frequencies tidak dapat menunjukkan atau menjawab pertanyaan apakah ada hubungan antara kedua peubah media informasi dan jenis kelamin.

Frekuensi relatif bersyarat memberikan persentase atau proporsi individu yang memiliki nilai khusus untuk satu variabel kategoris di antara individu yang berbagi nilai yang sama dari variabel kategoris lain (kondisi).

	TV	Radio	Surat Kabar	Total Baris
Laki-laki			64,2%	
Perempuan			35,8%	
Total Kolom			100%	

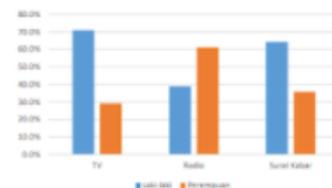
Contoh: Surat kabar sebagai syarat

Persentase media informasi surat kabar yang dibaca oleh responden perempuan didapatkan dari  $170/475=35,8\%$

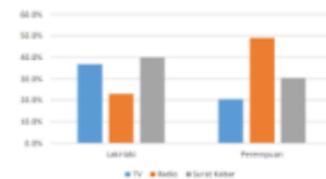
Sebaran frekuensi relative bersyarat

Bersyarat media informasi

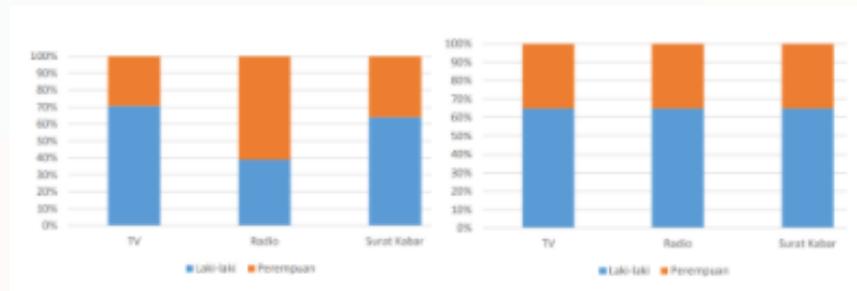
	TV	Radio	Surat Kabar	Total Baris
Laki-laki	70,9%	38,9%	64,2%	57,6%
Perempuan	29,1%	61,1%	35,8%	42,4%
Total Kolom	100%	100%	100%	100%



	TV	Radio	Surat Kabar	Total Baris
Laki-laki	36,8%	23,0%	40,1%	100,0%
Perempuan	20,5%	49,1%	30,4%	100,0%
Total Kolom	29,9%	34,1%	36,0%	100,0%



## Cara Mengidentifikasi Hubungan antara Dua Peubah Kategorik



Ada hubungan antara dua variabel jika mengetahui nilai satu variabel membantu kita memprediksi nilai dari yang lain. Media informasi dapat memprediksi dengan baik sebaran frekuensi relatif jenis kelamin, dari gambar terlihat jelas sebaran frekuensi relatif dari jenis kelamin berbeda-beda untuk setiap media informasi. Jika mengetahui nilai satu variabel tidak membantu kita memprediksi nilai yang lain, maka tidak ada hubungan antara variabel. Media informasi tidak dapat memprediksi dengan baik sebaran frekuensi relatif dari jenis kelamin, dalam artian sebaran frekuensi relatif dari jenis kelamin tidak tergantung dari jenis media.

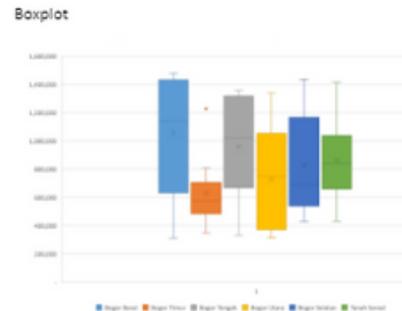
Untuk melihat hubungan antar dua peubah (1 peubah kategorik dan 1 peubah numerik)

- Ketika data bivariat dihasilkan dari satu peubah kategorik dan satu peubah numerik, nilai yang disajikan dipandang sebagai contoh terpisah.
- Setiap set data diidentifikasi oleh kategori/level dari peubah kategorik.
- Setiap set data tersebut dijelaskan menggunakan statistik ringkasan, dan hasilnya ditampilkan untuk perbandingan antar kategori/level dari peubah kategorik.
- Statistik yang biasa digunakan untuk perbandingan: ukuran pemusatan dan penyebaran, ringkasan 5 angka (min, maks, rata-rata, median dan modus).
- Grafik untuk perbandingan: boxplot.

Contoh 2 (Satu peubah kategorik dan satu peubah lainnya numerik)

- Misalkan diambil 10 contoh acak rumah tangga dari setiap enam kecamatan di Bogor yang berbeda dan dicatat tagihan listrik mereka pada bulan Juni 2021.

No	Bogor Barat	Bogor Timur	Bogor Tengah	Bogor Utara	Bogor Selatan	Tanah Sereal
1	658,882	1,228,388	958,196	388,571	476,779	648,358
2	1,429,951	571,026	1,333,844	828,865	1,081,645	1,413,845
3	1,478,629	607,659	883,835	880,629	559,023	1,204,869
4	1,129,486	529,237	331,348	1,110,411	1,435,815	673,959
5	1,329,952	809,461	317,583	313,724	746,735	790,439
6	1,188,627	491,497	717,807	330,054	429,637	918,519
7	553,485	346,934	1,090,390	1,337,854	1,424,550	976,752
8	1,083,222	672,316	1,312,065	672,642	398,550	431,347
9	1,480,790	803,462	1,359,137	396,543	677,742	662,972
10	311,803	586,027	1,088,257	1,033,270	632,239	887,783



### Hubungan dua peubah numerik

Untuk melihat hubungan antar dua peubah numerik dapat menggunakan:

- Scatter plot
- Ukuran keeratan hubungan antar peubah numerik, sering disebut sebagai ukuran korelasi. Beberapa ukuran korelasi seperti:
  - Koefisien Korelasi peringkat spearman
  - Koefisien Korelasi Pearson

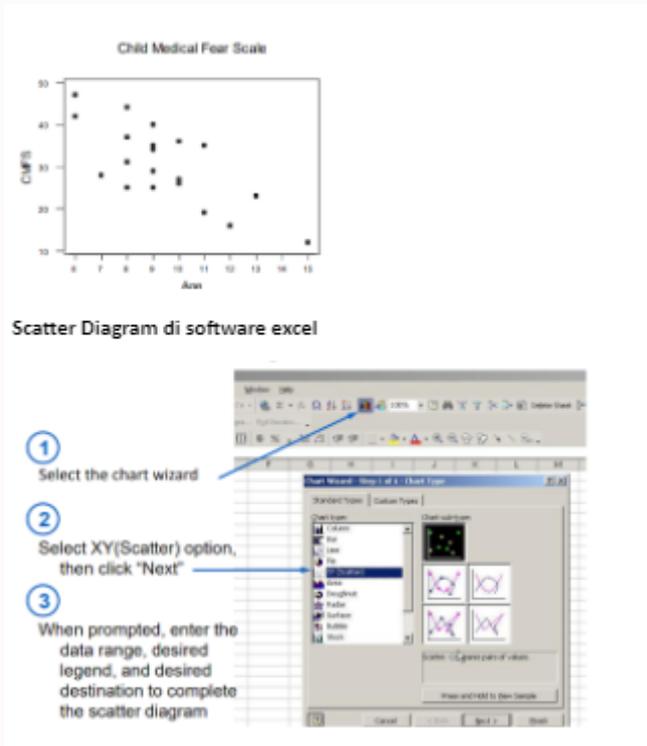
### Contoh 3 (Dua Peubah Numerik)

- Dalam sebuah penelitian ingin dilihat hubungan usia dengan tingkat ketakutan anak-anak ketika dirawat di rumah sakit. Skor tingkat ketakutan diukur dengan menggunakan Skala Ketakutan Medis Anak (Child Medical Fear Scale/CMFS).

Age (x)	8	9	9	10	11	9	8	9	8	11
CMFS (y)	31	25	40	27	35	29	25	34	44	19
Age (x)	7	6	6	8	9	12	15	13	10	10
CMFS (y)	28	47	42	37	35	16	12	23	26	36

### Scatter Diagrams

Scatter Diagrams digunakan untuk melihat hubungan pengamatan berpasangan yang diambil dari dua peubah numerik



### Koefisien Korelasi

- Digunakan untuk melihat kekuatan hubungan linier antar dua peubah
- Koefisien korelasi Spearman,  $r_s$ , adalah suatu nilai yang mengukur kekuatan hubungan linier antar dua peubah kategorik yang berskala ordinal
- Koefisien korelasi Pearson,  $r$ , adalah suatu nilai yang mengukur kekuatan hubungan linier antar dua peubah numerik (skala pengukuran interval atau rasio)

### Korelasi Peringkat Spearman

- Koefisien korelasi peringkat Spearman,  $r_s$ , merupakan ukuran atas kekuatan hubungan antar data yang telah tersusun sesuai peringkat data, bukan nilai aktual dari data tersebut.
- Nilai dari  $r_s$ , berkisar antara  $-1.0$  s.d  $+1.0$ , dimana:
- Nilai  $r_s$  mendekati  $+1.0$  mengindikasikan ada asosiasi/hubungan positif antar peringkat, artinya kenaikan (penurunan) satu peubah berasosiasi dengan kenaikan (penurunan) peubah lain
- Nilai  $r_s$  mendekati  $-1.0$  mengindikasikan ada asosiasi/hubungan negatif antar peringkat, artinya kenaikan (penurunan) satu peubah berasosiasi dengan penurunan (kenaikan) peubah lain
- Nilai  $r_s = 0$  mengindikasikan tidak ada asosiasi/hubungan monoton

Cara mencari nilai rs:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Di mana:

n= jumlah data yang diperingkatkan

$x_i$  = peringkat data ke - i pada peubah pertama

$y_i$  = peringkat data ke-i pada peubah pertama

$$d_i = x_i - y_i$$

### Contoh: Connor Investors

Connor Investors menyediakan layanan manajemen portofolio untuk kliennya. Dua dari analis Connor menilai sepuluh investasi dari risiko tinggi hingga rendah seperti yang ditunjukkan di bawah ini. Hitunglah koef korelasi peringkat Spearmen antara kedua penilaian analis tersebut.

Investment	A	B	C	D	E	F	G	H	I	J
Analyst#1	1	4	9	8	6	3	5	7	2	10
Analyst#2	1	5	6	2	9	7	3	10	4	8

Investment	A	B	C	D	E	F	G	H	I	J
Analyst#1	1	4	9	8	6	3	5	7	2	10
Analyst#2	1	5	6	2	9	7	3	10	4	8
$d_i$	0	-1	3	6	-3	-4	2	-3	-2	2
$(d_i)^2$	0	1	9	36	9	16	4	9	4	4

Korelasi Peringkat Spearman

$$\sum d_i^2 = 92$$

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Nilai kekuatan dari hubungan kedua analis, nilai rs mendekati +1.0 yang mengindikasikan ada asosiasi/hubungan positif antar peringkat, artinya kenaikan (penurunan) satu peubah berasosiasi dengan kenaikan (penurunan) peubah lain.

**Koefisien Korelasi Pearson**  
**Koefisien Korelasi untuk Populasi :**

$$\rho = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$$

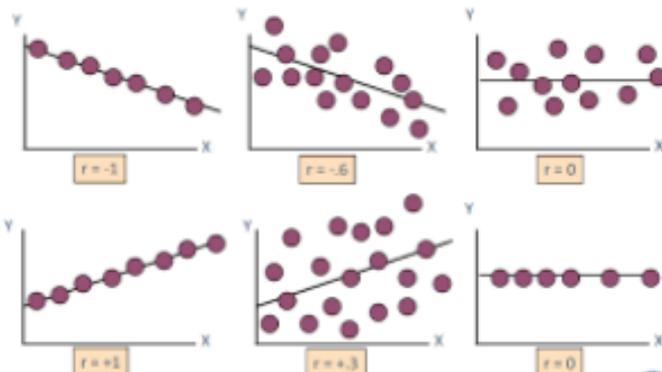
• Koefisien Korelasi untuk contoh :

$$r = \frac{\text{Cov}(x, y)}{s_x s_y}$$

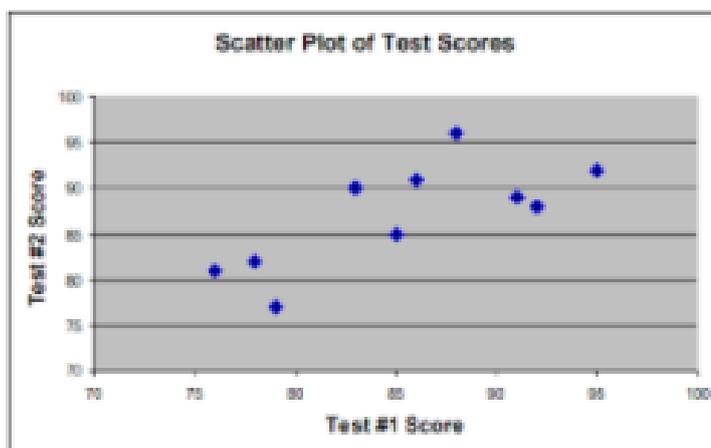
### Korelasi Perason

- Nilai dari  $r$ , berkisar antara  $-1.0$  s.d  $+1.0$ , dimana:
- Nilai  $r$  mendekati  $+1.0$  mengindikasikan ada hubungan linier positif antar kedua peubah numerik, artinya kenaikan (penurunan) satu peubah berkorelasi dengan kenaikan (penurunan) peubah lain
- Nilai  $r$  mendekati  $-1.0$  mengindikasikan ada hubungan linier negatif antar kedua peubah numerik, artinya kenaikan (penurunan) satu peubah berkorelasi dengan penurunan (kenaikan) peubah lain
- Nilai  $r = 0$  mengindikasikan tidak ada hubungan linier antar kedua peubah numerik tersebut

#### Scatter Plots of Data with Various Correlation Coefficients



#### Interpretasi Koefisien Korelasi Pearson



### Penggunaan Excel untuk Mencari Koefisien Korelasi

- Select Tools/Data Analysis
- Choose Correlation from the selection menu
- Click OK . . .

- Input data range and select appropriate options
- Click OK to get output

	A	B	C
1		Test #1 Score	Test #2 Score
2	Test #1 Score		
3	Test #2 Score	0.733243795	1
4			

$r = 0.733$

- Ada hubungan linier positif antara test score #1 dan test score #2
- Siswa dengan skor tinggi pada test pertama akan cenderung mendapatkan skor yang tinggi pula pada test kedua



# STATISTIKA

**RANTIK PEKAN 7**

**MODELLING  
ANALISIS REGRESI**

**Analisis regresi merupakan metode untuk mengembangkan sebuah model (persamaan) yang menjelaskan hubungan di antara beberapa variable. Fungsi analisis regresi, antara lain:**

1. **memprediksi nilai dari peubah tak bebas (dependent variable) berdasarkan nilai setidaknya satu peubah bebas (independent variable).**
2. **Menjelaskan dampak perubahan peubah bebas terhadap peubah tak bebas.**

**Keterangan:**

**Peubah tak bebas (Y) merupakan peubah yang ingin dijelaskan**

**Peubah bebas (X) merupakan peubah yang digunakan untuk menjelaskan peubah tak bebas**

**Model Regresi Linier**

a. **hubungan antara X dan Y dijelaskan dengan fungsi linier**

b. **Perubahan Y diasumsikan disebabkan dengan perubahan X**

**Model Regresi Linier Sederhana**

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Diagram illustrating the components of the simple linear regression equation:

- Dependent variable:**  $Y_i$
- Population Y intercept:**  $\beta_0$
- Population Slope Coefficient:**  $\beta_1$
- Independent Variable:**  $X_i$
- Random Error term:**  $\varepsilon_i$
- Linear component:**  $\beta_0 + \beta_1 X_i$
- Random Error component:**  $\varepsilon_i$

Persamaan regresi linier sederhana memberikan dugaan bagi garis regresi populasi

$$\hat{y}_i = b_0 + b_1 X_i$$

Diagram illustrating the components of the estimated regression equation:

- Estimated (or predicted) y value for observation i:**  $\hat{y}_i$
- Estimate of the regression intercept:**  $b_0$
- Estimate of the regression slope:**  $b_1$
- Value of x for observation i:**  $X_i$

The individual random error terms  $e_i$  have a mean of zero

$$e_i = (y_i - \hat{y}_i) = y_i - (b_0 + b_1 X_i)$$

## Penduga Kuadrat Terkecil (Least Squares Estimators (LSE))

- $b_0$  dan  $b_1$  adalah dugaan bagi parameter regresi  $\beta_0$  dan  $\beta_1$  yang didapat **salah satunya** dengan cara meminimumkan jumlah kuadrat galat (JKG). Galat/sisaan = selisih antara  $y$  dan  $\hat{y} \rightarrow$  Metode Kuadrat Terkecil (MKT) :

$$\begin{aligned}
 \min \text{JKG} &= \min \sum e_i^2 \\
 &= \min \sum (y_i - \hat{y}_i)^2 \\
 &= \min \sum [y_i - (b_0 + b_1 x_i)]^2
 \end{aligned}$$

Teknik kalkulus digunakan untuk mendapatkan nilai  $b_0$  dan  $b_1$  sedemikian hingga meminimumkan JKG

- Penduga bagi koefisien kemiringan garis  $\beta_1$  ialah:

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{S_{XY}}{S_{XX}} = r_{xy} \frac{s_Y}{s_X}$$

(Note:  $S_{XY}$  is labeled as the numerator,  $S_{XX}$  as the denominator, and the fraction  $\frac{S_{XY}}{S_{XX}}$  is labeled as the Coefficient of Correlation Pearson.)

- Penduga bagi intersep  $\beta_0$  ialah:

$$b_0 = \bar{y} - b_1 \bar{x}$$

- Garis regresi selalu melalui titik  $\bar{x}, \bar{y}$

## Estimator Kotak

- $b_0$  dan  $B_1$  diperoleh dengan mencari nilai  $b_0$  dan  $B_1$  bahwa meminimalkan jumlah selisih kuadrat antara  $y$  dan:  $\hat{y}$

$$\begin{aligned}
 \min \text{SSE} &= \min \sum_{i=1}^n e_i^2 \\
 &= \min \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\
 &= \min \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)]^2
 \end{aligned}$$

Kalkulus diferensial digunakan untuk mendapatkan penduga koefisien  $b_0$  dan  $B_1$  yang meminimalkan SSE

- Penaksir koefisien kemiringan adalah

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = r_{xy} \frac{S_y}{S_x}$$

- Dan konstanta atau perpotongan  $y$  adalah

$$b_0 = \bar{y} - b_1 \bar{x}$$

- Garis regresi selalu melalui mean  $x, y$

**NASKAH ASLI**

**Soal-Soal – Tutor Sebaya 58 Sesi UTS**

**STATISTIKA**

**PENDIDIKAN KOMPETENSI UMUM (PKU)**

**IPB UNIVERSITY**

**SOAL-SOAL**

**7 FEBRUARI 2022**

**Instagram : @tutorsebaya\_ppku**

**YouTube : Tutor Sebaya 58**



## SOAL STATISTIKA SESI UTS

1. Rata-rata nilai 1000 orang adalah 75,63. Annisa mengambil 700 orang sebagai perwakilan dari 1000 orang tersebut. Rata-rata nilai 700 orang tersebut adalah 83,25. 1000 orang tersebut; 75,63; 700 orang; dan 83,25 berturut-turut merupakan...
- Parameter, statistik, populasi, sampel
  - Statistik, parameter, sampel, populasi
  - Populasi, parameter, sampel, statistik
  - Sampel, statistik, populasi, parameter

Jawaban : c. Populasi, parameter, sampel, statistik

Pembahasan : Populasi merupakan keseluruhan objek (1000 orang)

Parameter merupakan karakteristik dari populasi (75,63)

Sampel merupakan amatan yang diambil untuk mewakili objek (700 orang)

Statistik merupakan karakteristik dari sampel (83,25)

2. Yang tidak termasuk ke dalam ukuran pemusatan adalah...

- Range
- Median
- Quartil
- Mean

Jawaban: a. range

Pembahasan: Range tidak termasuk ukuran pemusatan tetapi termasuk salah satu ukuran penyebaran.

3. Modus merupakan nilai tengah suatu data yang sudah diurutkan. (Benar/Salah)

Jawaban: salah

Pembahasan: Modus merupakan nilai yang paling sering muncul.

4. Yang merupakan ciri dari modus adalah...

- Menunjukkan nilai tengah dari suatu data
- Tidak dipengaruhi oleh nilai pencilan
- Hanya dapat digunakan pada data numerik
- Selalu ada satu modus di setiap data

Jawaban: b. tidak dipengaruhi oleh nilai pencilan.

Pembahasan: Modus menampilkan nilai yang paling sering muncul pada suatu data, sehingga nilai pencilan tidak mempengaruhi.

5. Jika data  $>$  PDB atau data  $<$  PDA maka data dikatakan memiliki Pencilan. (Benar/Salah)

Jawaban: salah

Pembahasan: Suatu data memiliki pencilan jika data yang paling kecil  $<$  PDB dan data yang paling besar bernilai  $>$  PDA.

6. Cara mendeteksi nilai ekstrem bisa menggunakan eksplorasi data sebagai berikut, kecuali...

- a. Scatter plot
- b. Boxplot
- c. Histogram
- d. Steam and leaf

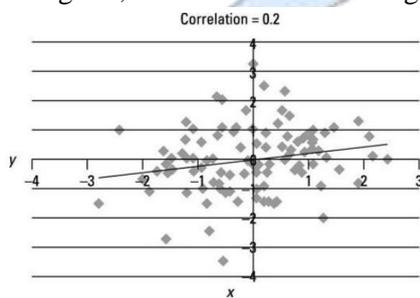
Jawaban: d. steam and leaf

Pembahasan: Eksplorasi data yang bisa digunakan untuk menentukan adanya data pencilan atau nilai ekstrem adalah scatter plot, boxplot, dan histogram.

7. Eksplorasi perbandingan antar grup dapat dilakukan dengan menggunakan Boxplot dan Histogram serta Statistik lima serangkai. (Benar/Salah)

Jawaban: benar

Pembahasan: Untuk eksplorasi perbandingan antar grup bisa menggunakan boxplot, histogram, dan statistik lima serangkai.



8. Scatter plot disamping ini memiliki hubungan linier yang...

- a. Positif kuat
- b. Negatif lemah
- c. Positif lemah
- d. Negatif kuat

Jawaban: c. Positif lemah

Pembahasan: Titik-titik pada grafik lebih tersebar (Agak jauh dari garis) di sekitar garis tren. Dan garis tren ini memiliki kemiringan garis yang positif, ditandai dengan jika nilai X naik, maka nilai Y juga akan naik. Serta nilai correlation yang didapat menunjukkan nilai yang positif.

9. Berdasarkan hasil survey pada 4000 orang, hasil yang diperoleh adalah kebanyakan mahasiswa IPB University menyukai perkuliahan luring tetapi menginginkan ujian secara daring. Kalimat tersebut termasuk statistika inferensia (Benar/Salah).

Jawaban : benar

Pembahasan : statistika inferensia mengambil hasil/kesimpulan/keputusan di bagian akhir.

10. Pada ujian kalkulus, soal yang diujikan terlalu sulit bagi sebagian besar mahasiswa. Sehingga ketika pembagian hasil nilai ujian, data yang dihasilkan akan menjulur ke kanan. Data tersebut sebaiknya tidak digunakan karena tidak simetris. (Benar/Salah)

Jawaban : Salah

Pembahasan : Data yang baik bukan berarti harus simetris. Pola data yang menjulur tersebut akan memudahkan interpretasi.

11. Hari Jumat adalah hari yang melelahkan bagi Bagus karena jadwal kuliahnya penuh dari pagi sampai sore. Setelah kelas sore, Bagus merasa lapar, namun di meja makan hanya ada nasi dan ayam semur. Karena Bagus adalah seorang vegetarian, ia memutuskan pergi ke kebun untuk memetik kangkung yang ditanam kakaknya secara hidroponik. Agar tidak mubazir, Bagus mengukur tinggi setiap tanaman kangkung untuk mengetahui yang mana saja yang akan ia petik. Peubah tinggi kangkung tersebut diukur dengan skala pengukuran ordinal (Benar/Salah).

Jawaban : Salah

Pembahasan : Tinggi tanaman diukur dengan skala pengukuran rasio, karena ada perbandingan antar pengukuran dan terdapat nol mutlak.

12. Rizky ingin mengetahui lama waktu teman sekelasnya mengerjakan laprak fisika. Dari 119 orang, ia mewawancarai 85 orang karena belum mengenal teman lainnya. Hasil yang Rizky dapatkan adalah sebagai berikut.

Waktu (menit)	Frekuensi
30	1
45	5
60	10
90	12
120	16
150	19
180	17
210	5

Pernyataan yang salah dari data tersebut adalah...

- Modusnya adalah 150 menit
- Mediannya adalah 120 menit
- Rangennya adalah 5
- Jumlah responden yang mengerjakan laprak 45 menit dan 90 menit lebih sedikit daripada yang mengerjakan selama 150 menit

Jawab : c. Rangennya adalah 5

Pembahasan :

a. Modusnya adalah 150 menit → benar, karena frekuensi paling banyak ada di 150 menit, yaitu sebanyak 19 orang

b. Mediannya adalah 120 menit → benar,  $Median = X_{(n+1)/2} = X_{(85+1)/2} = X_{43}$

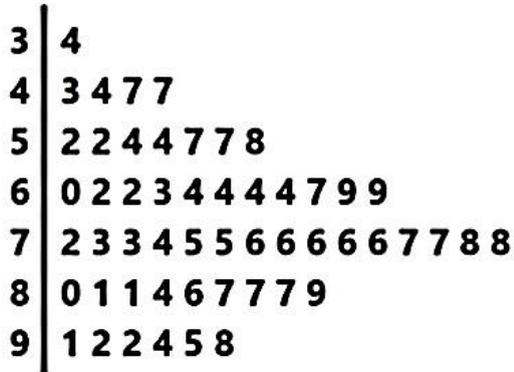
Maka, median terletak pada data ke-43, yaitu 120 menit.

c. Rangennya adalah 5 → salah,  $R = \max - \min = 210 - 30 = 180$

d. Jumlah responden yang mengerjakan laprak 45 menit dan 90 menit lebih sedikit daripada yang mengerjakan selama 150 menit → benar, yang mengerjakan laprak 45 menit + 90 menit

= 5 + 12 = 17 orang. Sedangkan yang mengerjakan selama 150 menit adalah 19 orang. Maka, yang mengerjakan selama 45 menit dan 90 menit lebih sedikit dibandingkan yang mengerjakan selama 150 menit.

13. Waktu pelayanan pada 53 kendaraan roda dua saat servis berkala di suatu bengkel, disajikan dalam diagram berikut.



Median dari data waktu servis di atas adalah ...

- a. 75 menit
- b. 73 menit
- c. 74 menit
- d. 77 menit

Pembahasan :

Jumlah data ( $n$ ) = 53. Berdasarkan jumlah  $n$ , gunakan metode interpolasi dalam menentukan posisi kuartil (dalam soal ini kuartil 2 yang memiliki makna serupa dengan median ).

$$\begin{aligned}
 n_{Q_2} &= \frac{1}{2} (n + 1) \\
 &= \frac{1}{2} (53 + 1) \\
 &= \frac{54}{2} = 27
 \end{aligned}$$

Data yang tertera di atas disajikan dalam bentuk diagram dahan dan daun dan telah disusun secara berurutan. Median berada pada bagian paling tengah dari urutan data. Untuk mempermudah pencarian median, gunakan akumulasi data yang ada dengan cara sebagai berikut :

Dahan	Daun	Jumlah data per dahan	Data kumulatif
3	4	1	1
4	3 4 7 7	4	5
5	2 2 4 4 7 7 8	7	12
6	0 2 2 3 4 4 4 4 7 9 9	11	23

7	2 3 3 4 5 5 6 6 6 6 6 7 7 8 8	15	38
8	0 1 1 4 6 7 7 7 9	9	47
9	1 2 2 4 5 8	6	53

Berdasarkan tabel, nomor urut data ke-27 berada pada dahan ke-7. Untuk memastikan bahwa median tersebut berada di dahan tersebut, buatlah tabel seperti berikut (penomoran dimulai dari angka 24 karena nomor urut 1-23 sudah digunakan oleh urutan data yang lebih kecil) :

Dahan Daun Urutan

7	
2	24
3	25
3	26
4	27
5	28
5	29
6	30
6	31
6	32
6	33
6	34
7	35
7	36
8	37
8	38

Berdasarkan hasil tersebut, median dari waktu servis adalah 74 menit.

14 . Hasil dari IQR adalah median

a . Benar

b.Salah

Pembahasan : IQR (*Interquartile Range*) adalah selisih dari Kuartil ketiga dengan kuartil pertama, sedangkan median merupakan nilai tengah dari kumpulan data yang telah diurut.

15. Tabel kontingensi merupakan tabel yang menyajikan jumlah pengamatan untuk setiap kombinasi nilai peubah kategorik dan numerik. (Benar/Salah)

Jawaban : Salah

Pembahasan : Tabel kontingensi merupakan tabel yang menyajikan jumlah pengamatan untuk setiap kombinasi nilai dua peubah kategorik.

16. Hari ini adalah hari yang melelahkan bagi Adis. Setelah kuliah seharian, Adis rebahan sejenak sambil membuka WA. Seseorang mengirimkan pesan kepada Adis bahwa orang tersebut menyukainya. Namun Adis adalah orang yang selektif, sehingga ia menguji kemampuan orang

tersebut lewat soal berikut. Rata-rata nilai suatu kelas adalah 55. Lisna adalah anak baru dikelas itu, sehingga ketika ditambah nilai Lisna yaitu 75, rata-ratanya menjadi 57. Banyaknya mahasiswa di kelas tersebut setelah kedatangan Lisna adalah... orang

- a. 8
- b. 9
- c. 10
- d. 11

Jawab : c. 10

Pembahasan :  $57 = (55n + 75)/(n+1)$

$$57n + 57 = 55n + 75$$

$$2n = 18$$

$$n = 9 \text{ (sebelum Lisna bergabung)}$$

Maka setelah Lisna bergabung jumlahnya menjadi  $9 + 1 = 10$  orang

17. Taopiq adalah mahasiswa ramah, tampan dan suka membantu orang lain. Beberapa mahasiswi memiliki rasa agak senang, senang, dan sangat senang kepada Taopiq. Tingkat kesenangan tersebut merupakan peubah...

- a. Diskret
- b. Nominal
- c. Ordinal
- d. Interval

Jawaban : c. Ordinal

Pembahasan : Tingkat kesenangan merupakan peubah kategorik dan dapat diurutkan, sehingga merupakan skala pengukuran ordinal.

18. Dalam analisis regresi, jika peubah bebas diukur dalam satuan cm, maka satuan peubah tak bebas...

- a. Harus dalam cm
- b. Harus dalam %
- c. Tidak boleh cm
- d. Bebas apa saja

Jawaban : d. Bebas apa saja

Pembahasan : Satuan peubah bebas dan peubah tak bebas dalam analisis regresi tidak ada aturan khusus (harus sama atau harus berbeda), atau kata lainnya bebas.

19. Pernyataan berikut yang benar mengenai regresi linear sederhana, kecuali...

- a. Hubungan antara peubah x dan y ditunjukkan dalam bentuk fungsi linear
- b. Perubahan pada y diduga karena adanya perubahan pada x
- c. Ada lebih dari satu variabel bebas
- d. Berguna untuk menduga garis regresi populasi

Jawaban : c. Ada lebih dari satu variabel bebas

Pembahasan : Dalam regresi linier sederhana, variabel bebas yang diamati hanya satu.

20. Koefisien korelasi Spearman ( $r_s$ ) adalah suatu nilai yang mengukur kekuatan hubungan linier antar dua peubah kategorik yang berskala ordinal.

- a. benar

b. salah

Pembahasan : Korelasi Spearman memang digunakan dalam pengukuran kekuatan hubungan linier antar dua peubah kategorik ordinal, dengan rincian sebagai berikut :

- Koefisien korelasi peringkat Spearman ( $r_s$ ) merupakan ukuran atas kekuatan hubungan antar data yang telah tersusun sesuai peringkat data, bukan nilai aktual dari data tersebut.
- Nilai dari ( $r_s$ ) berkisar antara -1.0 s.d. +1.0, dimana :
- Nilai ( $r_s$ ) mendekati +1.0 mengindikasikan ada asosiasi/hubungan positif antar peringkat, artinya kenaikan (penurunan) satu peubah berasosiasi dengan kenaikan (penurunan) peubah lain
- Nilai ( $r_s$ ) mendekati -1.0[ mengindikasikan ada asosiasi/hubungan negatif antar peringkat, artinya kenaikan (penurunan) satu peubah berasosiasi dengan penurunan (kenaikan) peubah lain
- Nilai ( $r_s$ ) = 0 mengindikasikan tidak ada asosiasi/hubungan monoton
- Rumus Korelasi Spearman :

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Keterangan:

$n$  = jumlah data yang di peringkat

$x_i$  = peringkat data ke-i pada peubah pertama

$y_i$  = peringkat data ke-I pada peubah kedua

$d_i = x_i - y_i$



# **STATISTIKA**

## **SESI PRA UAS**



IPB University  
— Bogor Indonesia —

THE DOCTOR  
Dormitory Creator



# STATISTIKA DAN ANALISIS DATA

Teknik Pengumpulan Data :  
Metode Survei

PEKAN - 9



IPB University  
— Bogor Indonesia —



# Teknik Pengumpulan Data

## METODE SURVEI



### Teknik Pengumpulan Data

#### Observasi (Survei - Sensus)



Data sudah tersedia di alam, tinggal dikumpulkan

Banyak diterapkan pada bidang sosial ekonomi/ non IPA

#### Administratif



Data yang diperoleh dari hasil pencatatan administrasi suatu institusi

Misal :

Data transaksi perbankan  
Data transaksi supermarket

#### Percobaan/ Experiment



Data dibangkitkan dengan memberikan perlakuan terhadap suatu objek

Sering diterapkan pada penelitian pertanian/ IPA/ psikologi



### Daftar Istilah

**Populasi** : Himpunan seluruh objek kajian yang perlu ditentukan berikut isi, unit cakupan, waktu.  
*Misal* : Semua Mahasiswa IPB Tingkat 1 Tahun 2019

**Sampel** : Himpunan bagian dari Populasi yang diamati. Sampel yang baik adalah yang cukup dan mewakili

#### Unit/Elemen

**Unit Observasi** : Unit dimana informasinya diperoleh baik secara langsung maupun melalui responden

**Unit Sampling** : unit yang dijadikan dasar penarikan sampel

**Unit Analisis** : unit yang digunakan dalam tabulasi data

**Kerangka elemen** : Daftar seluruh anggota populasi

**Parameter** : Angka Ringkasan yang menunjukkan karakteristik dari populasi  
*Misal* : Rata-rata Indeks Prestasi Mahasiswa IPB Tingkat 1 tahun 2019

**Statistik** : angka ringkasan yang menunjukkan karakteristik dari sampel  
*Misal* : Rata-rata Indeks Prestasi sebagian Mahasiswa IPB Tingkat 1 tahun 2019

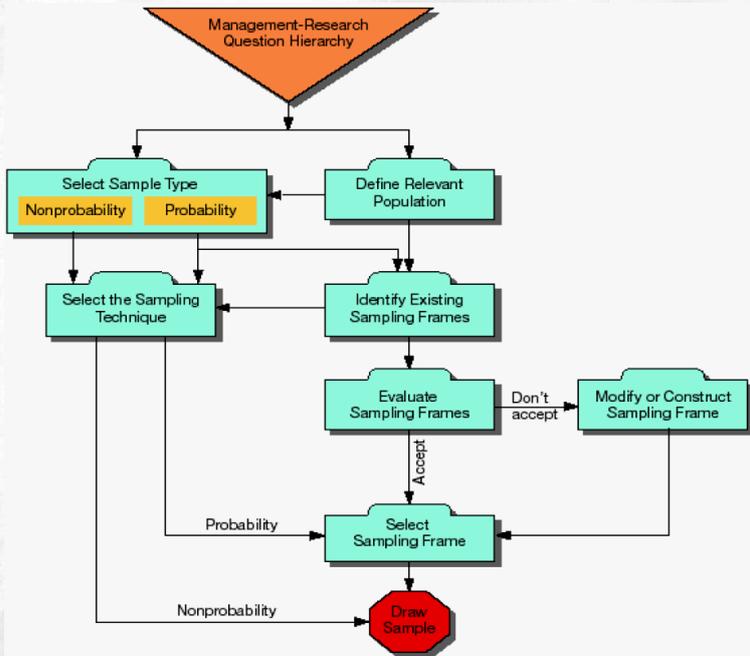
**Sampling** : Proses pengambilan sebagian populasi untuk diamati

## Penarikan Contoh yang Sah

### Mengapa Perlu Sampling Sah ?

**Validitas** menyiratkan sejauh mana instrumen penelitian mengukur, apa yang ingin diukur. **Reliabilitas** mengacu pada sejauh mana alat penilaian menghasilkan hasil yang konsisten, ketika pengukuran berulang dilakukan.

### Tahapan dalam mengambil sampel



## Metode Survei

### Pengertian Survei

Metode pengumpulan data terhadap sampel objek dari populasi untuk memperoleh informasi untuk menjawab tujuan penelitian. Survei biasanya dilakukan dengan wawancara dan dengan menggunakan kuesioner/angket dan pedoman wawancara.

### Apa Bedanya Dengan Observasi ?

Observasi merupakan pengamatan yang dilakukan dengan datang langsung dan mencatat segala hal-hal yang dianggap perlu, yang dilakukan oleh objek dari penelitian itu. Dengan demikian, Survei berbeda dengan observasi. **Survei dilakukan dengan berinteraksi langsung dengan objek penelitian melalui wawancara, sedangkan observasi tidak.**

### Tahapan Survei

1. Menentukan Masalah Penelitian
2. Membuat Desain Survei
3. Mengembangkan Instrumen Survei
4. Menentukan Sampel
5. Melakukan pretest
6. Mengumpulkan data
7. Pengolahan dan analisis data (coding, editing, inputting)
8. Manajemen Data
9. Interpretasi hasil analisis
10. Membuat kesimpulan serta rekomendasi

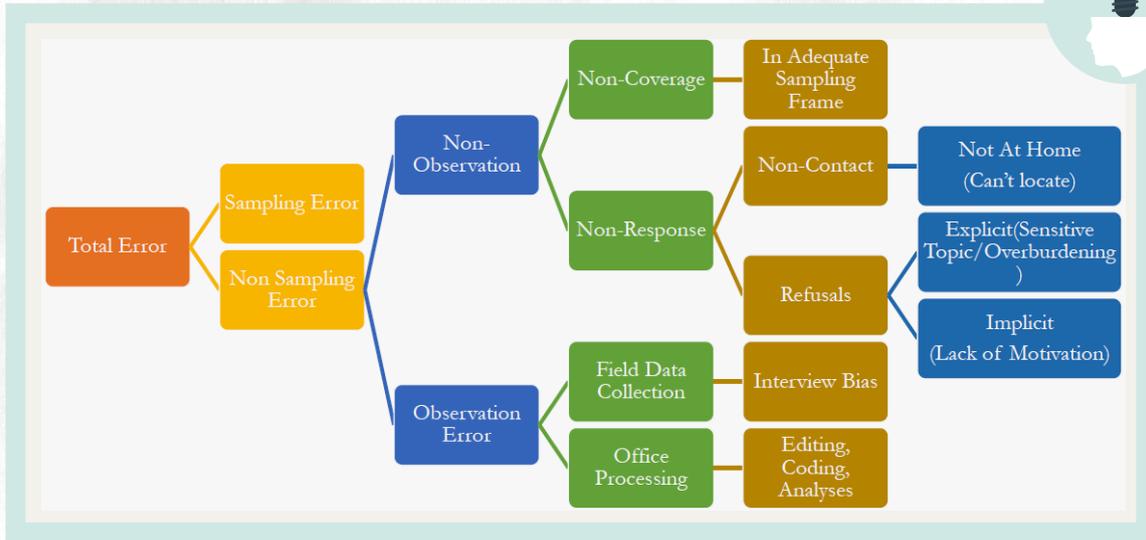
### Instrumen Survei

**Kuesioner** adalah daftar sekumpulan pertanyaan terstruktur yang dirancang secara sengaja untuk memperoleh informasi dari responden, yang disesuaikan dengan tujuan penelitian.

Informasi yang didapat dari kuesioner **HARUS**:

1. Valid : mengukur yang seharusnya diukur
2. Reliable : mengukur dengan konsisten
3. Unbiased : tidak bias dalam mengukur. Misal: salah jawab, atau mengkira-kira jawaban responden (ex: jumlah pupuk yang dipakai setahun yang lalu)
4. Discriminating : dapat membedakan karakteristik antar responden dengan baik

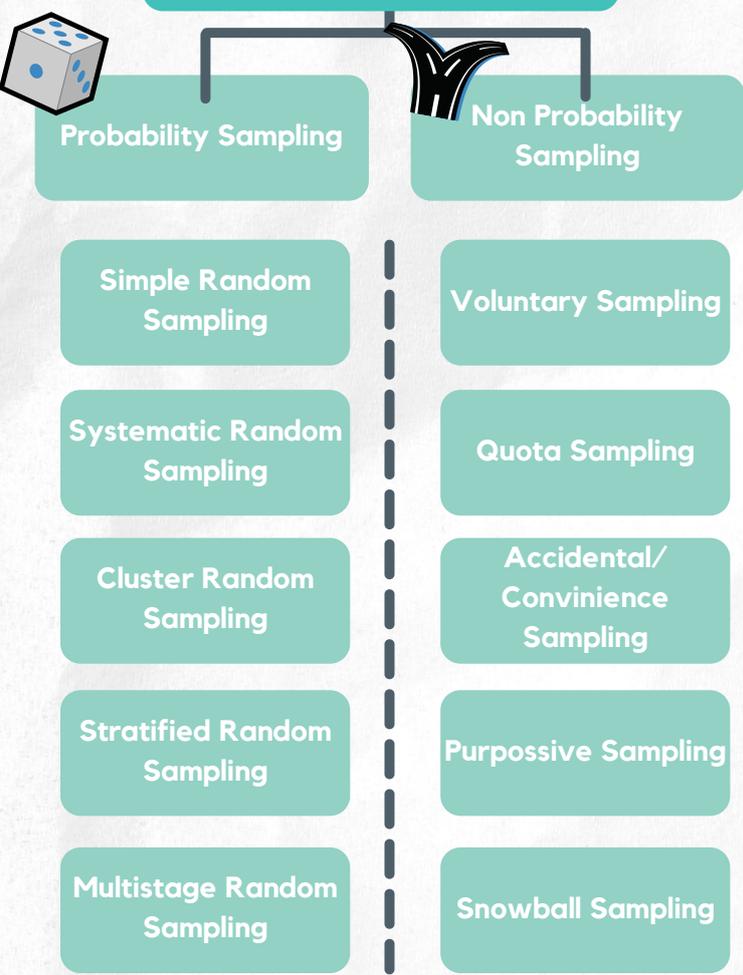
## Sumber Kesalahan Survei



## Metode Penarikan Contoh / Sampling methods

## Penarikan Contoh Berpeluang (Probability Sampling)

### Sampling Methods/ Sampling Techniques



### 1. Simple Random Sampling (Penarikan Contoh Acak Sederhana)

Proses pengacakan dilakukan langsung tanpa batasan apapun pada seluruh anggota populasi yang terdaftar dalam kerangka sampel.

- +** Cara Pengambilan sampel dan teknik estimasi parameternya sederhana
- Hanya cocok untuk populasi yang relative homogen
- Hanya cocok untuk cakupan survei yang tidak terlalu luas, karena membutuhkan kerangka sampel sampai level elemen
- Biaya tinggi untuk populasi yang besar

### 2. Systematic Random Sampling (Penarikan Contoh Acak Sistematis)

Proses pengacakan dilakukan tidak pada seluruh kerangka percontohan namun hanya pada pemilihan contoh pertama (pada rentang sampling fraction). Pemilihan contoh berikutnya dilakukan dengan lompatan sistematis sejauh sampling fraction dari contoh sebelumnya.



## 2. Systematic Random Sampling (Penarikan Contoh Acak Sistematis)

Dalam pengambilan sampel sistematis, terdapat dua pilihan untuk pengumpulan data yaitu:

- Memilih sampel dari daftar objek survei dan kemudian melakukan pendataan pada subjek yang dipilih, atau
- Langsung memilih sampel dari populasi target di lapangan, mendata setiap anggota ke-k

Langkah-langkahnya:

1. Tentukan ukuran contohnya ( $n$ )
2. Hitung intervalnya ( $k=N/n$  dan bulatkan ( $N =$  Populasi))
3. Pilih secara acak objek pertama dari  $k$  objek pertama
4. Pilih objek berikutnya setiap interval  $k$



- Penarikan sampel lebih mudah dan cepat
- Dengan pengurutan tertentu, sampel akan lebih representatif
- Dengan pengurutan yang tepat, presisi lebih tinggi dari simple random sampling



- Pengurutan unit yang kurang tepat akan memperbesar varians
- Tidak cocok diterapkan untuk populasi dengan variasi periodik
- Penduga varian yang unbiased sulit diperoleh dari sampel sistematis tunggal

## 3. Stratified Random Sampling (Penarikan Contoh Acak Berlapis)

Percontohan untuk menyiasati kondisi populasi yang beragam dengan membuat lapisan /subpopulasi/strata yang kondisinya seragam. Pemilihan contoh selanjutnya dilakukan pada tiap subpopulasi/ lapisan/ strata tersebut. Basis pembuatan lapisan adalah faktor(-faktor) yang terkait dengan substansi permasalahan yang diangkat.



- Meningkatkan efisiensi estimasi karakteristik populasi
- Masing-masing strata bisa dianggap sebagai populasi tersendiri sehingga bisa diterapkan desain sampling yang berbeda
- Estimasi bisa dilakukan untuk pengujian sampai level strata



- Menambah Error jika masing-masing strata dipilih pada tingkat yang berbeda
- Memiliki biaya yang lebih tinggi jika strata perlu dibuat terlebih dahulu

## 4. Clustered Random Sampling (Penarikan Contoh Acak Bergerombol)

Percontohan untuk menyiasati kondisi populasi yang beragam dengan membuat subpopulasi /gerombol yang kondisinya heterogen.

- Basis penggerombolan biasanya adalah kedekatan wilayah.
- Pemilihan contoh diawali dengan memilih cluster, dan seluruh anggota populasi pada cluster terpilih akan diambil.



- Pengumpulan data lebih murah dan mudah karena berdekatan, bila dibandingkan menyebar ke seluruh wilayah
- Tidak memerlukan syarat penyediaan kerangka sampel yang memuat seluruh unit sampel



- Tidak bisa digunakan untuk estimasi pada level strata
- Sering tidak efisien secara statistik jika dibandingkan karena bila anggota cluster cenderung homogen

## 5. Multistage Random Sampling (Penarikan Contoh Acak Bertahap)

Penarikan contoh yang dilakukan bertahap sebelum pemilihan elemen utama.



- Mengurangi biaya jika pada tahap pertama dapat dihasilkan data yang cukup untuk membuat strata atau cluster dari populasi



- Menambah biaya jika digunakan sembarangan
- Meningkatkan error akibat sampling.

## Stratified vs Clustered (Strata vs Bergerombol)

### Strata

- Populasi terbagi menjadi **beberapa strata**
- Di dalam strata cenderung **homogen**
- Antar strata cenderung **heterogen**
- Elemen yang dipilih diambil dari setiap strata

### Gerombol

- Populasi terbagi menjadi **banyak gerombol**
- Di dalam strata gerombol **heterogen**
- Antar gerombol cenderung **homogen**
- Elemen yang dipilih adalah gerombol

## Penarikan Contoh Tidak Berpeluang (Non Probability Sampling)

### 1. Non-Probability Sampling (Penarikan Contoh Tidak Berpeluang)

Sampel dipilih dengan metode non-random. Pemilihan sampel sangat tergantung pada kebijaksanaan atau pertimbangan dari peneliti.

- Dapat digunakan tanpa kerangka sampel
- Tidak dapat melakukan generalisasi hasil Survei ke populasi.
- Tidak mungkin untuk mengukur tingkat ketelitian data dari sampelnya

### 2. Purposive Sampling (Penarikan Contoh Bertujuan)

Teknik penarikan contoh bertujuan merupakan salah satu penarikan contoh tidak berpeluang berdasarkan karakteristik yang telah ditentukan sebelumnya dari objek yang akan dijadikan sampel, sesuai dengan tujuan yang telah ditetapkan. Dalam teknik ini, peneliti harus terlebih dahulu menetapkan tujuan sebelum mengambil sampel. Penetapan tujuan dilakukan agar sampel yang dikumpulkan memenuhi standar yang dibutuhkan untuk mendukung penelitian

(Sumber :  
<https://www.statmat.net/teknik-purposive-sampling/>)

### 3. Quota Sampling (Penarikan Contoh Kuota)

Teknik sampel ini adalah bentuk dari sampel distratifikasikan secara proposional, namun tidak dipilih secara acak melainkan secara kebetulan saja.

### 4. Accidental/ Convenience Sampling (Penarikan Contoh Kebetulan)

Dalam memilih sampel, peneliti tidak mempunyai pertimbangan lain kecuali berdasarkan kemudahan saja. Seseorang diambil sebagai sampel karena kebetulan orang tadi ada di situ atau kebetulan dia mengenal orang tersebut.

Jenis sampel ini sangat baik jika dimanfaatkan untuk penelitian penjajagan, yang kemudian diikuti oleh penelitian lanjutan yang sampelnya diambil secara acak (random). Beberapa kasus penelitian yang menggunakan jenis sampel ini, hasilnya ternyata kurang objektif.

### 5. Snowball Sampling (Penarikan Contoh Bola Salju)

Cara ini banyak dipakai ketika peneliti tidak banyak tahu tentang populasi penelitiannya. Dia hanya tahu satu atau dua orang yang berdasarkan penilaiannya bisa dijadikan sampel. Karena peneliti menginginkan lebih banyak lagi, lalu dia minta kepada sampel pertama untuk menunjukan orang lain yang kira-kira bisa dijadikan sampel.



**IPB University**  
— Bogor Indonesia —



## **6. Voluntary Sampling (Penarikan Contoh Sukarela)**

Voluntary Sampling (Sampel sukarela) adalah salah satu jenis utama metode pengambilan sampel non-probabilitas. Sampel sukarela terdiri dari orang-orang yang memilih sendiri ke dalam survei. Seringkali, orang-orang ini sangat tertarik dengan topik utama survei.

(Sumber :

<https://penelitianilmiah.com/jenis-non-probability-sampling/>)

**TUTOR  
SEBAYA 58**



**PB10**

# **PERANCANGAN PERCOBAAN**

**STATISTIKA DAN  
ANALISIS DATA**



## Pendahuluan

- Percobaan merupakan salah satu metode pengumpulan data (data collection) dalam suatu penelitian.
- DATA yang diperoleh merupakan RESPON dari objek percobaan yang telah diberikan suatu PERLAKUAN.
- Percobaan perlu PERANCANGAN agar diperoleh data yang valid, sehingga kesimpulan yang diperoleh akurat (tidak berbias) dan presisi (tingkat ketelitian) tinggi.

Perbandingan metode pengumpulan data percobaan:

Komponen	Percobaan	Observasi/Survey/Administrasi
Data	Belum tersedia, dibangkitkan berupa respon	Sudah tersedia, hanya perlu proses pengumpulan
Objek penelitian	<ul style="list-style-type: none"> <li>• Kondisi objek harus serba sama</li> <li>• Diberikan tindakan berupa perlakuan</li> </ul>	<ul style="list-style-type: none"> <li>• Kondisi objek beragam</li> <li>• Tidak diberikan tindakan apa-apa, mengamati apa adanya</li> </ul>
Lingkungan	Ada pengendalian kondisi lingkungan	Pengendalian lingkungan secara langsung tidak ada

## Rancangan Percobaan

Suatu uji atau sederetan uji dimana suatu proses atau sistem mengakibatkan terjadinya perubahan yang cukup berarti dari variabel input, yang dapat diamati melalui respon yang muncul.

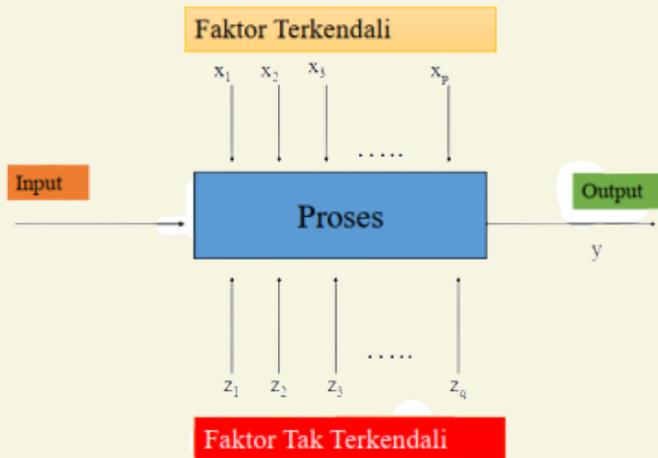
Perencanaan suatu percobaan untuk memperoleh informasi yang relevan dengan tujuan dari penelitian. perancangan percobaan meliputi beberapa hal, yaitu:

- Penentuan uji atau sederetan uji, disebut sebagai perlakuan.
- Proses menata percobaan di 'lapangan'.
- Proses pengukuran respon, (pengaruh dari uji yang diterapkan).

## Tujuan Perancangan

1. Memilih peubah terkendali (X) yang paling berpengaruh terhadap respon (Y).
2. Memilih gugus peubah X yang paling mendekati nilai harapan Y
3. Memilih gugus peubah X yang menyebabkan keragaman respon paling kecil
4. Memilih gugus peubah X yang mengakibatkan pengaruh peubah tak terkendali paling kecil

## Skema Proses



### Contoh

1. Input: Tanaman komoditi tertentu
2. Faktor Terkendali: Pola budidaya, varietas, jarak tanam, pupuk, dll
3. Faktor Tak Terkendali: curah hujan, kelembaban, suhu, intensitas penyinaran, dll.
4. Output: Tinggi tanaman, luas daun, komponen produksi, dll.

## PRINSIP DASAR PERANCANGAN PERCOBAAN

### Pengacakan (Randomization)

Setiap unit percobaan memiliki peluang yang sama untuk diberikan suatu perlakuan. Hal ini dimaksudkan untuk:

- Menghindari galat sistematis
- Meningkatkan validitas kesimpulan (pemenuhan asumsi kebebasan)
- Caranya dengan metode lotere, tabel bilangan acak, *software computer*

No	1	2	3	4	5	6	7	8	9
Perlakuan	A	B	C	A	B	C	A	B	C
Bilangan acak	0,586	0,157	0,456	0,987	0,933	0,412	0,032	0,906	0,042
Peringkat	4	7	5	1	2	6	9	3	8
No Unit percobaan	4	7	5	1	2	6	9	3	8

Penempatan perlakuan dalam petak percobaan

A	B	B
(1)	(2)	(3)
A	C	C
(4)	(5)	(6)
B	C	A
(7)	(8)	(9)

Catatan: cara pengacakan sangat tergantung pada jenis rancangan percobaan yang digunakan

### Ulangan (Replication)

Proses penerapan suatu perlakuan terhadap beberapa unit percobaan yang kondisinya relatif sama. Ulangan perlakuan (replication) tidak sama dengan ulangan pengukuran (repeatability).

Pengulangan ditujukan untuk:

- Untuk menduga galat percobaan
- Untuk menduga *standard error* rata-rata perlakuan
- Untuk meningkatkan presisi kesimpulan

#### Berapa jumlah ulangan?

1. Minimal 3
2. Minimal db-galat 15
3. Gunakan formula yang ada

$$r \geq 2(Z_{\alpha/2} + Z_{\beta})^2 \left( \frac{\sigma}{\delta} \right)^2$$

Hal-hal yang mempengaruhi ulangan yaitu:

1. **Tingkat kecermatan** yang diinginkan → semakin kecil perbedaan dari Hipotesis, semakin besar ulangan yang diperlukan
2. **Keragaman materi** → makin beragam makin banyak ulangan diperlukan.
3. Umumnya sangat ditentukan oleh **biaya** tersedia dan **waktu**.
4. Ulangan **tidak akan memperbaiki galat** akibat dari kesalahan teknis

### Pengendalian Lingkungan (Local Control) -----

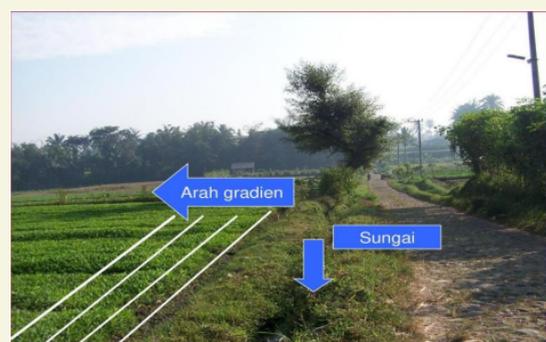
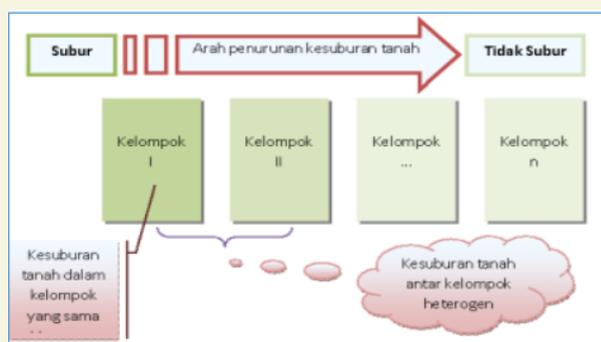
Proses mengendalikan kondisi unit percobaan agar kondisinya relatif sama (homogen). Pengendalian lingkungan diperlukan pada saat kondisi unit percobaan kondisinya tidak sama (heterogen).

Hal ini ditujukan untuk:

- Untuk menjamin bahwa setiap perlakuan dicobakan pada kondisi yang serba sama
- Untuk meningkatkan presisi kesimpulan

Pengendalian lingkungan dapat dilakukan melalui **strategi pengelompokan** (blocking).

- Pengelompokan dimaksudkan mengumpulkan unit-unit percobaan yang relatif sama menjadi satu kelompok dan memisahkan unit-unit percobaan yang berbeda dalam kelompok yang lain.
- Pengelompokan berhasil jika keragaman dalam kelompok kecil (minimum) dan keragaman antar kelompok besar (maximum).
- Untuk mencapai pengelompokan optimum dapat dilakukan dengan membuat arah kelompok yang tegak lurus dengan arah keragaman unit percobaan.
- Syarat kelompok : tidak berinteraksi dengan perlakuan.



## Istilah

- **Perlakuan:** suatu prosedur yang akan diterapkan terhadap unit percobaan.
- **Unit Percobaan:** unit terkecil dalam suatu percobaan yang diterapkan suatu perlakuan
- **Unit Pengamatan:** unit terkecil dalam suatu percobaan tempat dilakukannya pengukuran respon perlakuan
- **Faktor:** peubah bebas/kovariat penyusun perlakuan, yang diduga berpengaruh terhadap respon
- **Taraf:** nilai-nilai dari peubah bebas, yang dicobakan dalam suatu percobaan. Dapat dibedakan menjadi taraf kualitatif dan kuantitatif

## KLASIFIKASI PERANCANGAN PERCOBAAN

RANCANGAN PERLAKUAN	RANCANGAN LINGKUNGAN	RANCANGAN PENGUKURAN
Rancangan yang terkait dengan apa yang akan diuji pengaruhnya.	Rancangan yang terkait dengan bagaimana pengendalian lingkungan yang dilakukan dan bagaimana mengalokasikan perlakuan ke unit percobaan.	Rancangan yang terkait dengan bagaimana mengukur respon perlakuan.
FAKTOR TUNGGAL	Rancangan Acak Lengkap (RAL)	Mengukur respon kualitatif: Tingkat kematangan buah, dll.
MULTI FAKTOR (FAKTORIAL, TERSARANG, SPLIT PLOT, SPLIT BLOK, DLL)	Rancangan Acak Kelompok Lengkap (RAKL)  Rancangan Bujur Sangkar Latin (RBSL)	Mengukur respon kualitatif: Ukuran panjang, berat, suhu, luas, dll.

## Jenis-Jenis Rancangan Percobaan

Rancangan Perlakuan	Rancangan Lingkungan	Penamaan Rancangan
Faktor Tunggal Perlakuan yang dicobakan merupakan taraf-teraf dari satu faktor tertentu.	Rancangan Acak Lengkap (RAL) • Kondisi unit percobaan diasumsikan serbasama (homogen)	Faktor Tunggal - RAL
	Rancangan Acak Kelompok Lengkap (RAKL) • Kondisi unit percobaan tidak homogen. Sumber ketidakhomogenan unit percobaan berasal dari satu arah. Pengendalian ketidakhomogenan dapat dilakukan dengan pengelompokan	Faktor Tunggal - RAKL
	Rancangan Bujur Sangkar Latin (RBSL) • Kondisi unit percobaan tidak homogen. Sumber ketidakhomogenan unit percobaan berasal dari dua arah. • Pengendalian ketidakhomogenan dapat dilakukan dengan pengelompokan dua arah (blok baris dan blok lajur) • Jumlah perlakuan=jumlah blok baris=jumlah blok lajur	Faktor Tunggal - RBSL

## Jenis-Jenis Rancangan Percobaan

Rancangan Perlakuan	Rancangan Lingkungan	Penamaan Rancangan
<b>Faktorial</b>  Perlakuan yang dicobakan merupakan kombinasi taraf-taraf dari dua factor atau lebih dari dua factor	<b>Rancangan Acak Lengkap (RAL)</b> • (idem)	Faktorial – RAL
	<b>Rancangan Acak Kelompok Lengkap (RAKL)</b> • (idem)	Faktorial– RAKL
	<b>Rancangan Bujur Sangkar Latin (RBSL)</b> • (idem)	Faktorial – RBSL
<b>Tersarang</b>  Perlakuan yang dicobakan merupakan kombinasi taraf-taraf dari dua factor atau lebih dari dua factor, tetapi taraf suatu factor tidak sama pada berbagai taraf factor yang lain	<b>Rancangan Acak Lengkap (RAL)</b> • (idem)	Tersarang - RAL
	<b>Rancangan Acak Kelompok Lengkap (RAKL)</b> • (idem)	Tersarang - RAKL
	<b>Rancangan Bujur Sangkar Latin (RBSL)</b> • (idem)	Tersarang - RBSL
Rancangan Perlakuan	Rancangan Lingkungan	Penamaan Rancangan
<b>Split-plot</b>  <ul style="list-style-type: none"> <li>• Perlakuan yang dicobakan merupakan kombinasi taraf-taraf dari dua factor atau lebih dari dua factor.</li> <li>• Tingkat kepentingan antar factor tidak setara (Factor yang satu lebih dipentingkan dibanding yang lainnya).</li> <li>• Secara teknis taraf-taraf dari factor tertentu membutuhkan petak yang lebih besar dibandingkan dengan yang lainnya</li> </ul>	<b>Rancangan Acak Lengkap (RAL)</b> • (idem)	Split plot – RAL
	<b>Rancangan Acak Kelompok Lengkap (RAKL)</b> • (idem)	Split plot – RAKL
	<b>Rancangan Bujur Sangkar Latin (RBSL)</b> • (idem)	Split plot – RBSL
<b>Split-blok</b>  <ul style="list-style-type: none"> <li>• Perlakuan yang dicobakan merupakan kombinasi taraf-taraf dari dua factor atau lebih dari dua factor.</li> <li>• Pengaruh Interaksi lebih diutamakan dibandingkan dengan pengaruh utama</li> <li>• Secara teknis taraf-taraf dari kedua factor membutuhkan petak yang besar</li> </ul>		
	<b>Rancangan Acak Kelompok Lengkap (RAKL)</b> • (idem)	Split blok – RAKL

Jenis rancangan lain, seperti:

- Rancangan faktor tunggal dalam rancangan acak kelompok tidak lengkap
- Rancangan faktorial dalam rancangan acak kelompok tidak lengkap
- Rancangan fraksional Faktorial
- Rancangan Lattice
- Rancangan augmented
- Rancangan Percobaan dengan pengamatan berulang (*repeated measurement*)
- Uji coba multilokasi (*Multi locations Trials*)

## FAKTOR TUNGGAL DALAM RAL

### KARAKTERISTIK RANCANGAN

- Perlakuan yang dicobakan merupakan taraf-teraf dari **satu faktor** tertentu. Misal faktor yang ingin dikaji pengaruhnya adalah **Varietas**. Perlakuan yang dicobakan adalah Var1, Var2 dan Var3 (taraf-teraf dari varietas).
- Faktor-faktor diluar perlakuan dikondisikan serbasama
- Kondisi unit percobaan diasumsikan serbasama (**homogen**)

Ilustrasi: Pengaruh dosis pupuk N terhadap produksi jagung. Pupuk N dibedakan menurut dosisnya yaitu 0 kg/ha, 100 kg/ha, dan 200 kg/ha. Seluruh faktor diluar dosis pupuk N dikondisikan serba sama.

- Faktor: Dosis pupuk N
- Perlakuan: taraf-teraf yang dicobakan (0, 100, 200)

## FAKTOR TUNGGAL DALAM RAKL

### KARAKTERISTIK RANCANGAN

- Perlakuan yang dicobakan merupakan taraf-teraf dari satu faktor tertentu.
- Faktor-faktor diluar perlakuan dikondisikan serbasama
- Kondisi unit percobaan **tidak homogen**. Sumber ketidakhomogenan unit percobaan berasal dari **satu arah**. Pengendalian ketidakhomogenan dapat dilakukan dengan pengelompokan

Pemberian perlakuan terhadap unit percobaan dilakukan secara acak **pada setiap kelompok**, dengan batasan bahwa setiap perlakuan **muncul sekali pada setiap kelompok**. Contoh: suatu percobaan dengan enam buah perlakuan (P1, P2, P3, P4, P5, P6) dan setiap perlakuan diulang dalam tiga kelompok atau blok.

Dengan demikian unit percobaan yang melibatkan sebanyak 6 unit pada setiap blok sehingga secara keseluruhan dibutuhkan  $3 \times 6 = 18$  unit percobaan. Pengacakan perlakuan dilakukan pada masing-masing blok percobaan. Setiap perlakuan hanya muncul sekali pada setiap blok.

P1	P3	P2	P4	P6	P5
----	----	----	----	----	----

→ BLOK 1

P3	P5	P6	P4	P1	P2
----	----	----	----	----	----

→ BLOK 2

P1	P5	P3	P4	P2	P6
----	----	----	----	----	----

→ BLOK 3

## FAKTOR TUNGGAL DALAM RBSL

### KARAKTERISTIK RANCANGAN

- Perlakuan yang dicobakan merupakan taraf- taraf dari satu faktor tertentu.
- Faktor-faktor diluar perlakuan dikondisikan serbasama
- Kondisi unit percobaan **tidak homogen**.

Sumber ketidakhomogenan unit percobaan berasal dari **dua arah**. Pengendalian ketidakhomogenan dapat dilakukan dengan pengelompokan dua arah (blok baris dan blok lajur)

**Ilustrasi:** Penerapan perlakuan terhadap unit percobaan dilakukan secara acak, dengan memperhatikan batasan bahwa setiap perlakuan hanya muncul sekali pada arah baris dan hanya muncul sekali pada arah lajur.

- **Kasus:** Suatu penelitian melibatkan 4 perlakuan (A, B, C, D), dimana penempatan perlakuan diacak berdasarkan posisi baris dan lajur. Dengan demikian diperlukan empat posisi baris dan empat posisi lajur. Oleh karena posisi perlakuan tersarang pada posisi baris dan lajur maka banyak unit percobaan yang diperlukan adalah 4x4 unit percobaan.
- **Pengacakan Perlakuan:** Salah satu cara untuk mendapatkan penempatan perlakuan yang tepat maka dapat diambil tiga langkah utama sebagai berikut: (i) Tempatkan perlakuan pada arah diagonal secara acak, (ii) acaklah penempatan baris dan (iii) acaklah penempatan lajur.

#### Penempatan perlakuan searah diagonal

No. baris	1	A	C	D	B
2	B	A	C	D	
3	D	B	A	C	
4	C	D	B	A	
No. lajur	1	2	3	4	

#### Pengacakan penempatan baris

No. baris	3	D	B	A	C
2	B	A	C	D	
4	C	D	B	A	
1	A	C	D	B	
No. lajur	1	2	3	4	

#### Pengacakan penempatan lajur

No. baris	3	B	C	D	A
2	A	D	B	C	
4	D	A	C	B	
1	C	B	A	D	
No. lajur	2	4	1	3	

## Bagan Percobaan

Tata letak perlakuan-perlakuan pada seluruh objek percobaan di lapangan. Bagan percobaan berbeda-beda antar jenis rancangan percobaan. Adapun tahapan yang dilakukan untuk membangun bagan percobaan adalah sebagai berikut:

- Beri nomor perlakuan-perlakuan yang akan dicobakan ( $1, 2, 3, \dots, t$ )
- Siapkan unit percobaan yang akan digunakan (Misal: Banyaknya perlakuan  $t$  dan ulangan  $r$  maka siapkan sebanyak  $tr$  unit percobaan)
- Beri nomor unit percobaan (Misal:  $1, 2, \dots, tr$ )
- Acak perlakuan untuk setiap unit percobaan. Metode pengacakan dapat dilakukan dengan metode lotre, tabel bilangan acak, bangkitkan bilangan acak dengan menggunakan software, dll

Ilustrasi: Penerapan perlakuan terhadap unit percobaan dilakukan secara acak terhadap seluruh unit percobaan

Contoh: suatu percobaan melibatkan enam buah perlakuan (P1, P2, P3, P4, P5, P6) dan setiap perlakuan diulang sebanyak tiga kali. Dengan demikian unit percobaan yang dilibatkan sebanyak  $3 \times 6 = 18$  unit percobaan. Pengacakan perlakuan dilakukan langsung terhadap 18 unit percobaan.



Tata letak perlakuan  
di Lapangan

Dibangkitkan dengan Excel: rand()

No	Perlakuan	Ulangan	Bilangan Acak	Peringkat	No Unit Percobaan
1	P1	1	0,049	16	16
2	P2	1	0,151	14	14
3	P3	1	0,068	15	15
4	P4	1	0,023	17	17
5	P5	1	0,630	6	6
6	P6	1	0,778	3	3
7	P1	2	0,012	18	18
8	P2	2	0,694	5	5
9	P3	2	0,423	13	13
10	P4	2	0,801	2	2
11	P5	2	0,593	7	7
12	P6	2	0,498	10	10
13	P1	3	0,459	12	12
14	P2	3	0,547	9	9
15	P3	3	0,993	1	1
16	P4	3	0,464	11	11
17	P5	3	0,562	8	8
18	P6	3	0,706	4	4

Up1	Up2	Up3	Up4	Up5	Up6
P3	P4	P6	P6	P2	P5
Up7	Up8	Up9	Up10	Up11	Up12
P5	P5	P2	P6	P4	P1
Up13	Up14	Up15	Up16	Up17	Up18
P3	P2	P3	P1	P4	P1

## Model Linier Aditif

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \text{ atau } Y_{ij} = \mu_i + \varepsilon_{ij}$$

dimana:  $i=1, 2, \dots, t$  dan  $j=1, 2, \dots, r$

$Y_{ij}$  = Pengamatan pada perlakuan ke- $i$  dan ulangan ke- $j$

$\mu$  = Rataan umum

$\mu_i$  = Rata-rata perlakuan ke- $i$

$\tau_i$  = Pengaruh perlakuan ke- $i$   
=  $\mu_i - \mu$

$\varepsilon_{ij}$  = Galat percobaan pada perlakuan ke- $i$  ulangan ke- $j$

### HIPOTESIS:

$$H_0: \tau_1 = \dots = \tau_t = 0$$

(perlakuan tidak berpengaruh terhadap respon yang diamati)

$$H_1: \text{Paling sedikit ada satu } i \text{ dimana } \tau_i \neq 0$$

atau ,

$$H_0: \mu_1 = \dots = \mu_t = \mu$$

(semua perlakuan memberikan respon yang sama)

$$H_1: \text{Paling sedikit ada sepasang perlakuan } (i, i') \text{ dimana } \mu_i \neq \mu_{i'}$$

## Tabulasi Data Respon

Perlakuan	Ulangan				Ragam	Total
	1	2	....	r		
					$(S_i^2)$	$Y_i = \sum_{j=1}^r Y_{ij}$
Perlakuan 1	$Y_{11}$	$Y_{12}$	....	$Y_{1r}$	$S_1^2$	$Y_1$
Perlakuan 2	$Y_{21}$	$Y_{22}$	....	$Y_{2r}$	$S_2^2$	$Y_2$
....	....	....	....	....	....	....
Perlakuan t	$Y_{t1}$	$Y_{t2}$	....	$Y_{tr}$	$S_t^2$	$Y_t$

$$Y_i = \sum_{j=1}^r Y_{ij}$$

$$\bar{y}_i = \frac{\sum_{j=1}^r Y_{ij}}{r} = \frac{Y_i}{r}$$

$$S_i^2 = \sum_{j=1}^r (Y_{ij} - \bar{y}_i)^2 / (r - 1)$$

## Pendekatan Eksplorasi

Secara eksplorasi, beberapa langkah analisis yang dapat kita lakukan untuk data respon dari suatu percobaan, antara lain:

- Hitung ukuran peringkasan untuk setiap perlakuan: rata-rata dan ragam.

Perlakuan	$P_1$	$P_2$	$P_3$	....	$P_t$
Jumlah respon pada perlakuan ke-i	$Y_1 = \sum_{j=1}^r Y_{1j}$	$Y_2 = \sum_{j=1}^r Y_{2j}$	$Y_3 = \sum_{j=1}^r Y_{3j}$	....	$Y_t = \sum_{j=1}^r Y_{tj}$
Rata-rata respon perlakuan ke-i	$\bar{Y}_1 = Y_1/r$	$\bar{Y}_2 = Y_2/r$	$\bar{Y}_3 = Y_3/r$	....	$\bar{Y}_t = Y_t/r$
Ragam galat percobaan pada perlakuan ke-i	$S_1^2$	$S_2^2$	$S_3^2$		$S_t^2$

- Buat sajian grafis dari ukuran peringkasan di atas dengan menggunakan diagram batang
- Untuk melihat pola sebaran data untuk setiap perlakuan dapat juga disajikan dalam bentuk box-plot

*Dari kedua langkah ini, kita dapat melihat kecenderungan perlakuan-perlakuan mana yang memberikan respon tinggi dan yang rendah, serta juga dapat dilihat perbandingan keragaman antar perlakuan.*

### Ilustrasi

Seorang mahasiswa SPs-IPB melakukan percobaan untuk melihat pengaruh dosis pupuk N terhadap produksi jagung. Dosis pupuk N dibedakan 0 kg/ha, 100 kg/ha, dan 200 kg/ha dengan ulangan 4 kali. Unit percobaan adalah petak lahan berukuran 1m x 4m. Percobaan dilaksanakan di kebun percobaan IPB-Dramaga, kondisi lahan percobaan secara umum relatif sama (homogen). Seluruh faktor diluar dosis pupuk N seperti jenis varietas, kesehatan benih, pengolahan lahan, jarak tanam, periode dan frekuensi penyiangan/pemupukan/penyemprotan hama, umur panen dan faktor lainnya dikondisikan serba sama. Respon yang diamati adalah produksi jagung per petak (kg/petak). Jika dibuatkan ringkasan struktur percobaan diatas, sebagai berikut:

- Faktor : Dosis pupuk N
- Perlakuan : 0 kg/ha, 100 kg/ha, 200 kg/ha
- Ulangan : 4 kali
- Respon : Produksi jagung per petak (kg/petak)
- Rancangan : Faktor tunggal-RAL

## Data respon produksi jagung diperoleh sebagai berikut:

### Tata letak/Bagan/Denah percobaan

$N_0^{(1)}$	$N_{100}^{(1)}$	$N_0^{(3)}$	$N_{200}^{(1)}$
$N_{200}^{(3)}$	$N_0^{(2)}$	$N_{200}^{(4)}$	$N_{100}^{(3)}$
$N_{100}^{(2)}$	$N_{200}^{(2)}$	$N_{100}^{(4)}$	$N_0^{(4)}$

Petak percobaan:  
1m x 4m



### Tabulasi data respon: Produksi Jagung (kg/petak)

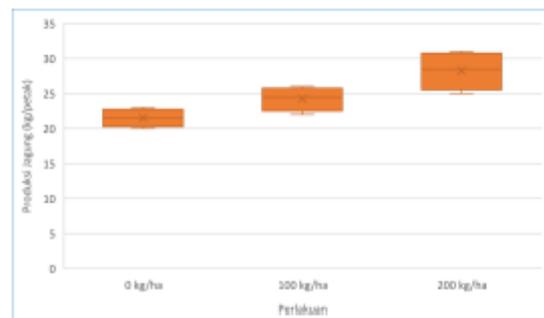
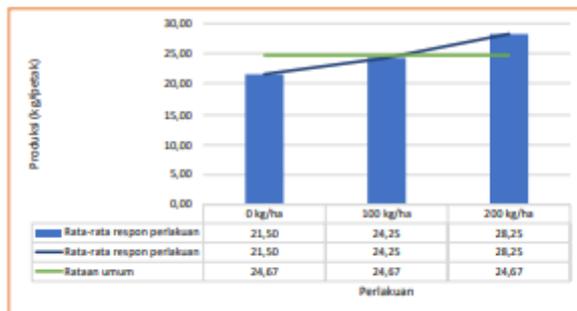
Perlakuan	Ulangan			
	1	2	3	4
0 kg/ha	20	22	21	23
100 kg/ha	25	24	26	22
200 kg/ha	30	31	25	27



Setelah masa panen, produksi jagung dicatat untuk setiap unit percobaan dan ditabulasikan seperti tabel disamping

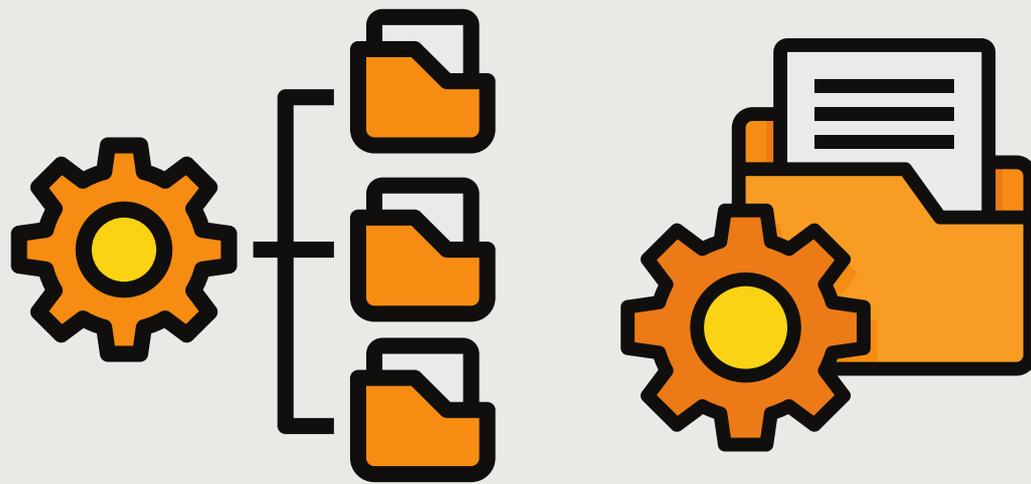
## Hasil Analisis Eksplorasi

Perlakuan	0 kg/ha	100 kg/ha	200 kg/ha	Keseluruhan
Total respon	86	97	113	296
Rata-rata respon perlakuan	21,50	24,25	28,25	24,67
Ragam galat	1,67	2,92	7,58	11,70



Apa yang dapat anda bisa simpulkan terkait analisis eksplorasi tersebut?

- Bagaimana rata-rata respon antar perlakuan? Mana perlakuan yang memberikan respon paling tinggi?
- Bagaimana dengan keragaman galat antar perlakuan? Perlakuan mana yang menyumbangkan keragaman galat terbesar?



# Manajemen Data STA-PB11



**IPB University**  
— Bogor Indonesia —



# Manajemen Data

## MANAJEMEN

Dapat diartikan, sebagai proses pengaturan atau pengelolaan di mana manajemen ini berfungsi untuk merencanakan, mengorganisasikan, mengarahkan, mengkoordinasikan, dan mengawasi. Hal tersebut mesti dijalankan dengan baik dalam rangka mencapai tujuan suatu organisasi/ institusi/ perusahaan.

Manajemen Data adalah proses merencanakan, mengolah, menyimpan, mengatur, dan memelihara pusat data yang dimiliki oleh sebuah organisasi/institusi. Manajemen data yang efektif dapat **meningkatkan efisiensi pekerjaan** dan dapat memberikan **informasi** yang dibutuhkan **secara cepat** untuk membantu proses **perancangan** dan **pengambilan** keputusan organisasi.

## Proses Manajemen Data



- Pada tahap perencanaan, disusun rancangan pengelolaan datanya. Misalnya, bagaimana menerjemahkan variabel penelitian ke dalam kuesioner (atau alat pengumpul data lainnya), bagaimana proses coding dan entry data-nya, bagaimana proses cleaning datanya nanti, bagaimana menyimpan datanya, dan lain-lain. Intinya, apa yang akan dilakukan di proses selanjutnya itu direncanakan terlebih dahulu dengan baik.
- Tahap pengolahan data dilakukan setelah data terkumpul, data tersebut meliputi kegiatan editing, coding, entry data, dan cleaning.

**Editing**, dimaksudkan untuk meneliti kembali apakah isian dalam kuesioner sudah cukup baik dan dapat segera diproses lebih lanjut. Misal memeriksa **apakah jawabannya sudah lengkap atau belum? Sudah seragam satuan pengukurannya atau belum? Sudah sesuai dengan yang seharusnya atau belum?** dan lain-lain.

**Coding**, artinya memberikan code pada nilai variabel. Misalnya, laki-laki di kode 1 perempuan di kode 0, atau pendapat yang setuju di kode 4, kurang setuju di kode 3, tidak setuju di kode 2, sangat tidak setuju di kode 1.

**Proses entri data**, menginput data pada form entry data yang digunakan. Apakah ada aplikasi tertentu untuk mengentri data atau hanya dientri dalam excel biasa, atau yang lainnya.

**Cleaning data**, dilakukan untuk memeriksa kembali apakah data yang dientri tadi sudah benar-benar siap dianalisis atau belum. Misalnya, melihat Apakah ada duplikasi data atau data yang sama terinput dua kali, dan lain-lain.

- Proses **penyimpanan data**, biasanya data-data yang telah diolah yang sudah rapi Itu disimpan dalam suatu *platform* yang sering dikenal dengan nama database. Database ini berisi kumpulan data yang terorganisir sehingga dapat diakses, diperbaharui, dan dikelola.
- Dalam **proses mengatur**, adalah proses bagaimana kita melakukan filter data sesuai dengan kebutuhan kita, proses melakukan agregasi, melakukan penggabungan, dll.
- Pada **proses pemeliharaan**, dilakukan perawatan secara berkala terhadap *database* yang kita miliki agar tetap terjaga kualitas dan keamanan data yang tersimpan. Di perusahaan, biasanya pengelolaan data ditangani oleh tim IT dan manajemen data, sedangkan analisis data ditangani oleh *data scientist* dan *data analytics*.

## Pentingnya Manajemen Data

- **Struktur data tidak terstandarisasi**  
Solusi: Sebaiknya rencana manajemen data sudah dibuat secara detail sejak awal. Fokus membangun sistem struktur dasar terlebih dahulu untuk memudahkan dalam menemukan, mengakses, dan menganalisis data di kemudian hari.
- **Entri Data tidak akurat dan terpercaya, serta inkonsisten**  
Solusi: Otomatisasi, lebih diandalkan karena menggunakan system database sehingga pengelolaan data dapat dirancang sebaik mungkin, untuk meminimalkan error yang rentan terjadi bila dilakukan secara manual.  
→ Otomatisasi adalah penggantian tenaga manusia dengan tenaga mesin yang secara otomatis melakukan dan mengatur pekerjaan sehingga tidak memerlukan lagi pengawasan manusia
- **Data tersimpan dalam sumber yang berbeda**
- **Kesulitan mencari data dalam waktu singkat**



## Proses Entri dan Coding Data

### Data Percobaan

Berikut ini adalah data yang dicatat dari suatu percobaan yang meneliti pengaruh lama penyinaran (P) dan kekuatan arus air (K) terhadap kenaikan berat badan ikan gurami

Lama Penynr.	Kektn. arus	K e l o m p o k				T o t a l
		I	II	III	IV	
p <sub>1</sub>	k <sub>1</sub>	56	45	43	46	190
	k <sub>2</sub>	60	50	45	48	203
	k <sub>3</sub>	66	57	50	50	223
p <sub>2</sub>	k <sub>1</sub>	65	61	60	63	249
	k <sub>2</sub>	60	58	56	60	234
	k <sub>3</sub>	53	53	48	55	209
p <sub>3</sub>	k <sub>1</sub>	60	61	50	53	224
	k <sub>2</sub>	62	68	67	60	257
	k <sub>3</sub>	73	77	77	65	292
T o t a l		555	530	496	500	2081



TIM DEPT STK-IPB UNIVERSITY

	A	B	C	D	E
1	No	Lama Penyinaran	Kekuatan arus	Kelompok	Respon
2	1	1	1	1	56
3	2	1	2	1	60
4	3	1	3	1	66
5	4	1	1	2	45
6	5	1	2	2	50
7	6	1	3	2	57
8	7	1	1	3	43
9	8	1	2	3	45
10	9	1	3	3	50
11	10	1	1	4	46
12	11	1	2	4	48
13	12	1	3	4	50
14	13	2	1	1	65
15	14	2	2	1	60
16	15	2	3	1	53
17	16	2	1	2	61
18	17	2	2	2	58
19	18	2	3	2	53
20	19	2	1	3	60
21	20	2	2	3	56
22	21	2	3	3	48
23	22	2	1	4	63
24	23	2	2	4	60
25	24	2	3	4	55
26	25	3	1	1	60
27	26	3	2	1	62
28	27	3	3	1	73
29	28	3	1	2	61
30	29	3	2	2	68
31	30	3	3	2	77
32	31	3	1	3	50
33	32	3	2	3	67
34	33	3	3	3	77
35	34	3	1	4	53
36	35	3	2	4	60
37	36	3	3	4	65

### Pemfilteran Data (Data filtering)

- Pemfilteran data adalah proses memilih sebagian kecil dari kumpulan data dan menggunakan subset (bagian) tersebut untuk mengeksplorasi atau menganalisis pola data.
- Pemfilteran biasanya (tetapi tidak selalu) bersifat sementara - kumpulan data lengkap tetap disimpan, tetapi hanya sebagian yang digunakan untuk penghitungan.
- Pemfilteran data biasanya digunakan untuk :
  - Lihat hasil untuk jangka waktu tertentu.
  - Hitung hasil untuk kelompok/group data tertentu.
  - Mendeteksi pengamatan yang "aneh" atau "buruk" dari analisis.
  - "Train" dan memvalidasi model statistik yang dibangun untuk suatu analisis
- Ada banyak alasan mengapa memfilter data
- Pada Big Data melakukan pemfilteran data adalah hal yang umum dilakukan
- Pemfilteran data sebenarnya dapat dilakukan pada satu peubah atau beberapa peubah dalam database.

# Pemfilteran Data pada Satu Peubah

**Pemfilteran Data pada satu peubah di Excel**

1. Arahkan kursor pada baris 1 atau baris yang berisi nama peubah pada data Anda
2. Pilih "filter" pada menu

userID	smoker	dress_preference	transport	marital_status	interest	activity	color	weight	budget	height	age
U0001	FALSE	informal	on foot	single	variety	student	black	69	medium	1.77	21
U0002	FALSE	informal	public	single	technology	student	red	40	low	1.87	20
U0003	FALSE	formal	public	single	none	student	blue	60	low	1.69	21
U0004	FALSE	no preference	public	single	none	student	black	65	medium	1.69	18
U0005	TRUE	no preference	car owner	single	variety	student	blue	75	medium	1.8	21
U0006	FALSE	informal	public	single	variety	student	purple	60	low	1.59	21
U0007	FALSE	formal	public	single	technology	student	green	68	low	1.77	21
U0008	FALSE	formal	on foot	single	variety	student	green	75	medium		
U0009	FALSE	no preference	car owner	married	technology	student	green	40	medium		
U0010	FALSE	no preference	public	single	variety	student	purple	68	medium		
U0011	FALSE	formal	public	single	technology	student	red	98	medium		
U0012	FALSE	no preference	public	widow	retro	profession	blue	80	medium		
U0013	TRUE	informal	public	single	technology	student	black	87	medium		
U0014	FALSE	informal	on foot	single	eco-friendly	student	green	70	medium		
U0015	FALSE	formal	public	single	retro	profession	orange	79	medium		
U0016	TRUE	formal	public	single	technology	student	blue	65	low		
U0017	FALSE	no preference	public	single	variety	student	purple	59	medium		
U0018	FALSE	no preference	public	single	technology	profession	blue	40	medium		
U0019	FALSE	no preference	car owner	single	none	student	purple	108	medium		
U0020	FALSE	formal	car owner	single	variety	student	purple	46	medium		
U0021	FALSE	informal	car owner	single	none	student	blue	103	low	1.79	22
U0022	FALSE	no preference	public	single	none	student	red	65	medium	1.79	20
U0023	FALSE	formal	public	single	technology	student	red	49	medium	1.55	21
U0024	TRUE	no preference	public	single	retro	student	purple	60	low	1.65	21
U0025	FALSE	no preference	public	single	variety	student	blue	77	medium	1.63	21
U0026	TRUE	no preference	public	single	variety	student	white	58	low	1.6	20
U0027	FALSE	formal	on foot	single	eco-friendly	student	black	64	medium	1.75	19
U0028	FALSE	formal	public	single	variety	student	blue	40	medium	1.75	21
U0029	FALSE	no preference	on foot	single	none	student	yellow	54	low	1.6	19

**Pemfilteran Data pada satu peubah di Excel**

Misalkan ingin melakukan filter untuk responden yang status pernikahannya adalah menikah (married)

**Hasil yang diperoleh**

userID	smoker	dress_preference	transport	marital_status	interest	activity	color	weight	budget	height	age
U0009	FALSE	no preference	car owner	married	technology	student	green	40	medium		
U0010	FALSE	no preference	public	single	variety	student	purple	68	medium		
U0011	FALSE	formal	public	single	technology	student	red	98	medium		
U0012	FALSE	no preference	public	widow	retro	profession	blue	80	medium		
U0013	TRUE	informal	public	single	technology	student	black	87	medium		
U0014	FALSE	informal	on foot	single	eco-friendly	student	green	70	medium		
U0015	FALSE	formal	public	single	retro	profession	orange	79	medium		
U0016	TRUE	formal	public	single	technology	student	blue	65	low		
U0017	FALSE	no preference	public	single	variety	student	purple	59	medium		
U0018	FALSE	no preference	public	single	technology	profession	blue	40	medium		
U0019	FALSE	no preference	car owner	single	none	student	purple	108	medium		
U0020	FALSE	formal	car owner	single	variety	student	purple	46	medium		
U0021	FALSE	informal	car owner	single	none	student	blue	103	low	1.79	22
U0022	FALSE	no preference	public	single	none	student	red	65	medium	1.79	20
U0023	FALSE	formal	public	single	technology	student	red	49	medium	1.55	21
U0024	TRUE	no preference	public	single	retro	student	purple	60	low	1.65	21
U0025	FALSE	no preference	public	single	variety	student	blue	77	medium	1.63	21
U0026	TRUE	no preference	public	single	variety	student	white	58	low	1.6	20
U0027	FALSE	formal	on foot	single	eco-friendly	student	black	64	medium	1.75	19
U0028	FALSE	formal	public	single	variety	student	blue	40	medium	1.75	21
U0029	FALSE	no preference	on foot	single	none	student	yellow	54	low	1.6	19

Contoh :

- Misalkan koordinator MK STA111 ingin mengetahui mahasiswa yang tidak mengikuti UTS
- Maka pemfilteran data akan dilakukan pada peubah "Nilai UTS" dengan memilih sel yang "#N/A"

Hasil filter mahasiswa yang tidak mengikuti UTS, koordinator harus melakukan pengecekan lebih lanjut apakah mahasiswa tersebut mengalami kendala saat UTS berlangsung atau karena alasan lainnya

## Pemfilteran Data pada Dua Peubah

Pemfilteran Data pada dua peubah atau lebih di Excel

Misalkan ingin melakukan filter untuk responden yang belum menikah (single)

Multiple filters

Hasil

Kemudian melakukan filter untuk responden yang belum menikah (single) DAN menggunakan alat transportasi mobil sendiri

Contoh :

- Misalkan koordinator MK STA111 ingin mengetahui hasil UTS untuk Kelas Paralel ST01 yang berasal dari Fakultas Kedokteran Hewan
- Maka pemfilteran data akan dilakukan pada peubah “Kelas Paralel” dan “Kode Fakultas

Data ini adalah data hasil UTS STA111 PPKU

Misalnya ingin diketahui hasil UTS kelas ST01 yang berasal dari FKH

Data ini adalah data hasil UTS PPKU

Misalnya ingin diketahui hasil UTS kelas ST01 yang berasal dari FKH

UTS ST01 FKH	
Mean	69.93
Standard Error	1.77
Median	71.00
Mode	74.00
Standard Deviation	11.33
Minimum	45.00
Maximum	88.00
Count	41.00





RANTIK PEKAN 10

# MANAJEMEN DATA (2)

## AGREGASI DATA

- Misalkan kita memiliki sekumpulan data dan perlu mengaturnya untuk melakukan suatu analisis data.
- Dari mana kita akan memulainya? Kita dapat membuat tabel ringkasan dengan cara mengagregatkan data tersebut.
- Untuk melakukan agregat, ada beberapa hal yang harus ada:
  1. Data yang ingin dilakukan agregat
  2. Peubah yang akan dikelompokkan dalam data
  3. Perhitungan yang akan dilakukan pada kelompok data tersebut

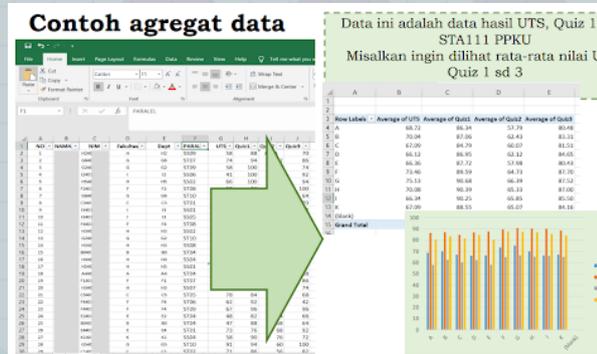
### Agregat Data dengan Excel

### Agregat Data dengan Excel

Count of marital_status	Column Labels			
4	Row Labels	married	single	Grand Total
5	user name	3	26	30
6	user hair	5	8	13
7	public	5	8	13
8	(blank)			
9	Grand Total	8	22	30

### CONTOH AGREGAT DATA :

- Misalkan ingin diketahui rata-rata nilai UTS, Quiz 1 sd 3 secara agregat per fakultas, fakultas mana yang rata-rata UTS nya paling tinggi ?
- Selain menampilkan tabel ringkasan dari rata-rata nilai UTS, Quiz 1 sd 3 secara agregat per fakultas, akan ditampilkan juga grafik dari masing-masing nilai rata-rata tersebut per fakultas



### ANOMALY DATABASE

- Anomali adalah proses pada basis data yang memberikan efek samping yang tidak diharapkan (misalkan menyebabkan ketidakkonsistenan data atau membuat sesuatu data menjadi hilang Ketika data lain dihapus)
- Umumnya ini dihilangkan dengan proses normalisasi yang dilakukan dengan pemisahan / penggabungan tabel.
- Ada tiga tipe anomali pada database:
  1. Anomali Penyisipan (Insertion anomaly) terjadi apabila pada saat penambahan hendak dilakukan, ternyata ada elemen data yang masih kosong, dan elemen data tersebut justru menjadi kunci.
  2. Anomali Peremajaan (Update anomaly) terjadi apabila ada perubahan pada sejumlah data yang mubazir pada suatu tabel tetapi tidak seluruhnya diubah. Ini terjadi jika terjadi redundansi data dan pembaruan parsial. Dengan kata lain, pembaruan database yang benar membutuhkan tindakan lain seperti penambahan, penghapusan, atau keduanya.
  3. Anomali Penghapusan (Deletion Anomaly) terjadi apabila suatu baris (record) yang tidak terpakai dihapus, dan sebagai akibat nya ada data lainnya yang hilang.

## PENGGABUNGAN DATA (MERGING DATA)

- Penggabungan data (merging data) adalah proses menggabungkan dua atau lebih kumpulan data menjadi satu kumpulan data
- Paling sering, proses ini diperlukan ketika Anda memiliki data mentah yang disimpan dalam beberapa file, lembar kerja, atau tabel data, yang ingin Anda analisis semuanya sekaligus.
- Ada dua contoh umum di mana seorang analis data perlu menggabungkan kasus baru menjadi file data utama :
  1. pengumpulan data dilakukan dalam studi longitudinal - data dikumpulkan selama periode waktu tertentu dan menganalisisnya per periode waktu.
  2. pengumpulan data dilakukan pada sebelum dan sesudah suatu proyek.
- Ada beberapa situasi yang mungkin memerlukan penggabungan data ke dalam file yang sudah ada,
  1. Adanya kasus baru,
  2. Adanya peubah baru (atau adanya kasus dan peubah baru)

## MENGGABUNGAN DATA BARU (MERGING IN NEW CASES)

- Menggabungkan/menambahkan data menurut baris (yaitu menambahkan baris data baru ke setiap kolom)
- Diasumsikan peubah yang digunakan dalam kedua file yang akan digabungkan adalah sama

## MENGGABUNGAN PEUBAH BARU (MERGING IN NEW VARIABLES)

- Berlawanan dengan penggabungan data baru, menggabungkan peubah baru memerlukan ID untuk setiap kasus di dua file agar sama, tetapi nama peubah harus berbeda.

## MENGGABUNGAN DATA MENGGUNAKAN LOOK-UPS

- Misalnya, suatu file data dimana salah satu peubahnya adalah kode pos responden, dan Anda ingin melampirkan/menambahkan beberapa data demografis ke data tersebut sebagai contoh data rata-rata pendapatan di setiap kode pos.



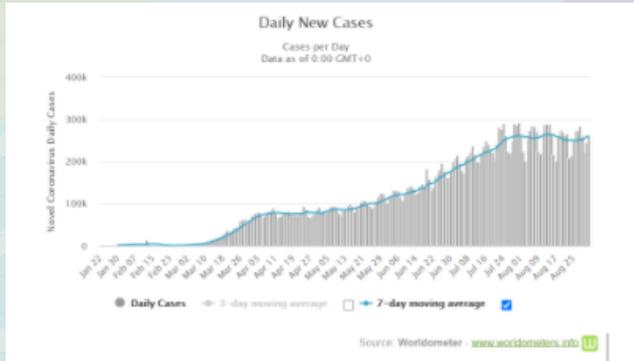
RANTIK PEKAN 13

# VISUALISASI DATA

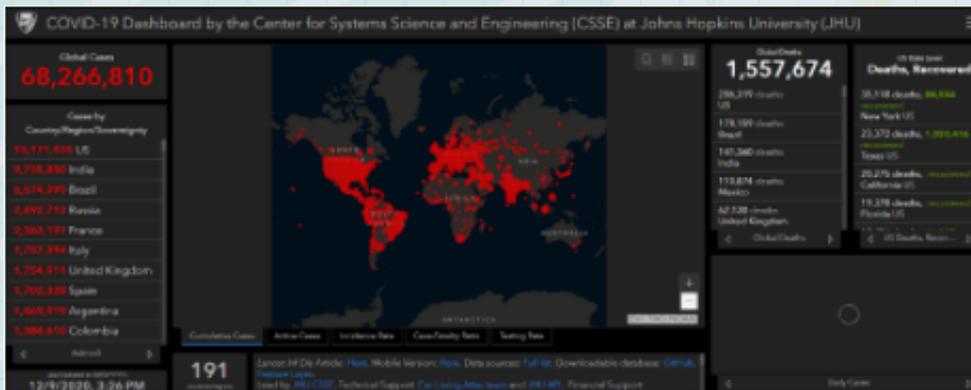
Pengertian : Penyajian informasi dalam bentuk gambar atau grafik yang diperoleh dari data melalui proses analitika.

## Bentuk Visualisasi Data

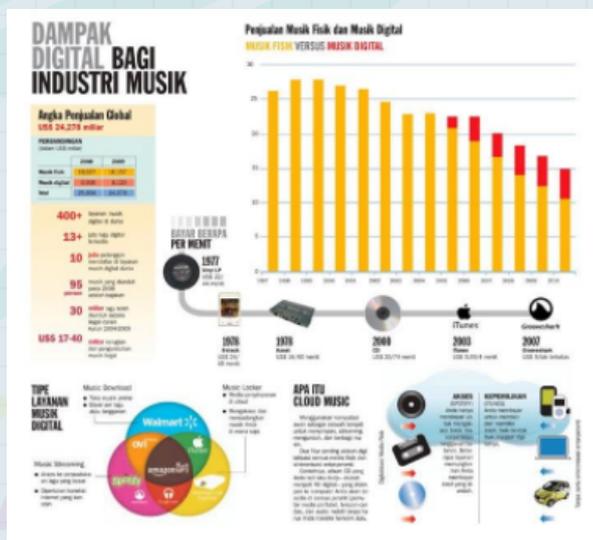
### 1. Diagram



### 2. Dashboard



### 3. Infografis



## Peranan Visualisasi dalam Proses Analisis Data

Visualisasi dalam analisis data memiliki peran sebagai berikut :

1. Selama proses analisis data
  - Memahami karakteristik perilaku data yang dimiliki (eksplorasi)
  - Mengarahkan pendalaman-pendalaman analisis data (elaborasi)
  - Membangkitkan ide baru terkait analisis data
2. Setelah proses analisis data
  - Meringkas informasi pada data besar sehingga mudah melihat pola pada data dan memahami cerita di belakangnya
  - Memudahkan menyampaikan informasi sehingga lebih cepat menyusun aksi
  - Membantu pembuat keputusan memahami apa yang sedang atau telah terjadi
  - Mengarahkan audiens untuk fokus pada hal yang dituju untuk diperhatikan
  - Menyibak informasi yang tidak terlihat sebelumnya sehingga memudahkan menuliskan narasi dan laporan

## JENIS VISUALISASI DATA

### 1. Visualisasi Besaran dan Perbandingannya

Contoh :

- BAR Chart (Diagram Batang)
- LOLLIPOP Chart

### 2. Visualisasi Proporsi

Contoh :

- Pie Chart
- DONUT Chart
- TREE-MAP Chart

### 3. Visualisasi Sebaran Data dan Perbandingannya

Contoh :

- HISTOGRAM
- DENSITY PLOT
- Boxplot
- Violin Plot

### 4. Visualisasi Hubungan Antar Variabel

Contoh :

- SCATTER PLOT (Plot tebaran) + TREND LINE
- Bubble Chart

### 5. Visualisasi Data Deret Waktu dan Trend

Contoh :

- TIME SERIES PLOT Line chart
- STACKED AREA PLOT
- STACKED-BAR CHART

### 6. Visualisasi Geospasial

Contoh :

- THEMATIC MAP (PETA TEMATIK - CHOROPLETH)

### 7. Visualisasi Aliran dan Jaringan

Contoh :

- NETWORK DIAGRAM (Diagram Jaringan)

### 8. Visualisasi Lainnya

Contoh :

WORD CLOUD (Awan kata)

## UNSUR VISUAL UTAMA PADA GRAFIK

### 1. POSISI

Dalam suatu diagram, posisi dapat digunakan merepresentasikan nilai amatan yang berbeda dengan cara menempatkan amatan tersebut pada lokasi yang berbeda pada ruang yang ada di diagram. Wujud dari posisi antara lain:

- Beberapa batang dengan letak bersebelahan pada bar chart dimana batang dari satu kategori diposisikan berdampingan dengan batang dari kategori lainnya
- Letak titik-titik amatan pada plot tebaran (scatter plot) yang letaknya tergantung pada nilai dari variabel yang digambarkan.

Unsur posisi ini dapat digunakan untuk memetakan data numerik maupun kategorik.

Untuk peubah numerik dan kategorik ordinal :

- Amatan dengan nilai yang lebih besar digambarkan pada posisi yang lebih ke kanan pada sumbu mendatar dan lebih ke atas pada sumbu tegak.
- Hanya pada situasi tertentu saja diposisikan sebaliknya.

Khusus untuk data kategorik yang bersifat nominal, urutan posisi dapat diberikan sembarang oleh pembuat diagram, atau menggunakan urutan yang dipilih secara subjektif.

### 2. SIMBOL (BENTUK SIMBOL)

Pada suatu grafik beberapa amatan bisa jadi diberikan simbol yang berbeda sesuai dengan karakteristik yang dinyatakan pada datanya. Titik-titik amatan pada plot tebaran bisa diatur berbeda tergantung pada variabel lain yang menggambarkan karakteristik ketiga.

- Misal: plot tebaran antara usia dan tingkat pendapatan beberapa orang responden digambarkan dalam plot tebaran, dan titik untuk responden laki-laki diberikan simbol yang berbeda dengan titik untuk responden perempuan.
- Karena simbol ini berwujud bentuk yang berbeda-beda tanpa memiliki urutan, maka unsur simbol cocok digunakan pada data dengan tipe kategorik nominal.

### 3. GARIS

- Unsur garis dapat digunakan untuk menghubungkan satu titik amatan dengan amatan yang lain.
- Garis dapat berguna dalam memperlihatkan trend perubahan pada line chart atau plot deret waktu.
- Bentuknya yang ke arah kanan atas menunjukkan peningkatan sedangkan yang ke arah kanan bawah menunjukkan penurunan.
- Garis juga dapat digunakan pada peta dan diagram jaringan/network yang dapat mengindikasikan adanya hubungan antara dua titik, dengan definisi hubungan tertentu seperti keberadaan jalur transportasi serta indikasi perdagangan barang.

#### 4. UKURAN

- Unsur ini menempel pada unsur lain
  - ukuran besar kecilnya titik pada plot tebaran
  - ukuran juring pada pie-chart
  - ukuran ketinggian barang pada bar chart
  - ketebalan garis.
- Ukuran umumnya merepresentasikan wujud dari data numerik dan kategorik ordinal.
- Nilai data yang semakin besar dapat digambarkan dalam ukuran yang juga lebih besar sehingga dengan adanya ukuran yang berbeda, pembaca grafik dapat dengan mudah melihat perbedaan nilai data.

#### 5. WARNA

- Sebuah warna dapat mewakili suatu nilai data.
- Pada data yang bisa terurut dengan makna tertentu (yaitu data ordinal dan numerik) dapat digunakan data dengan gradasi untuk membedakan nilai yang kecil dan besar.
- data yang bernilai lebih besar digambarkan dalam wujud warna yang semakin pekat atau gelap, sedangkan data yang kecil digambarkan dalam warna yang terang.
- Sementara itu pada data yang bersifat nominal, warna hanya digunakan untuk membedakan sifat perbedaan saja. Dengan adanya perbedaan warna kita dapat memperoleh dengan mudah perbedaan karakteristik antar kategori yang ada pada data



RANTIK PEKAN 14

# DATA STORYTELLING

Data Storytelling mengkomunikasikan informasi yang diperoleh dari data atau dari hasil analisis data, sehingga dapat mempercepat pengambilan keputusan dan perencanaan tindakan yang diperlukan.

Data Storytelling terdiri dari :

1. Data and Analysis : descriptive, predictive, prescriptive
2. Visualization : diagram/chart yang sesuai, informasi yang ingin disampaikan

- Gestalt laws in visual perception
- Gestalt Law of Continuation : mata kita cenderung mengikuti pola yang alurnya lebih mulus (smooth)
- Gestalt Law of Proximity : objek yang jaraknya dekat cenderung diasosiasikan sebagai satu kesatuan
- Gestalt Law of Similarity : objek yang terlihat sama cenderung diasosiasikan sebagai satu kesatuan
- Gestalt Law of Enclosure : objek yang terlihat sama cenderung diasosiasikan sebagai satu kesatuan

3. Narrative : alur narasi penyampaian, gaya penyampaian  
Mempersiapkan slide presentasi

Fase pengerjaan presentasi :

Phase 1 : Penentuan Masalah dan Analisis Data

- serangkaian diskusi dan kajian literatur
- pengumpulan data
- analisis data

Phase 2 : Penyusunan Bahan Presentasi

- pembuatan slide (teks, diagram, gambar)
- skenario presentasi
- kalimat pembuka dan penutup

Tiga prinsip penyusunan slide presentasi :

- visualisasi konten
- gunakan unsur secara konsisten (font type, warna)
- fokuskan perhatian audiens ke informasi dan pesan yang ingin disampaikan

Empat tipe slide presentasi :

- Slide Teks
- Slide Diagram (visualisasi data)
- Slide Gambar
- Slide Struktur Presentasi

Phase 3 : Pelaksanaan Presentasi di Hadapan Audiens

- berdiri di hadapan audiens
- menarik perhatian audience
- meyakinkan audiens

Cara menyampaikan presentasi dengan baik :

- Tampililah dengan percaya diri Berdiri tegak, senyum, dan santai
- Tatap mata audiens jangan membaca catatan
- Kendalikan kecepatan bicara
- Atur volume suara
- Bicaralah dengan jelas agar lebih mudah didengar oleh audiens
- Tarik perhatian dan partisipasi audiens visualisasi berikan pertanyaan aktifitas humor
- Luangkan waktu untuk berlatih

Target Audience

- latar belakang audiens
- tujuan: informasi? persuasi?

**NASKAH ASLI**

**Pembahasan Try Out – Tutor Sebaya 58 Sesi UAS**

**STATISTIKA DAN ANALISIS DATA**

**PENDIDIKAN KOMPETENSI UMUM (PKU)**

**IPB UNIVERSITY**

**SOAL-SOAL**

**25 NOVEMBER 2021**

Instagram : @tutorsebaya\_ppku

YouTube : Tutor Sebaya 58

## Pembahasan

1. Survei dilakukan dengan pengamatan secara langsung dan mencatat hal yang dilakukan oleh objek penelitian. (Benar/Salah)

Pembahasan :

Survei merupakan metode pengumpulan data terhadap sampel objek dari populasi untuk memperoleh informasi untuk menjawab tujuan penelitian. Survei biasanya dilakukan dengan wawancara dan dengan menggunakan kuesioner/angket dan pedoman wawancara. Di soal merupakan pengertian dari observasi.

2. Kuesioner harus berisi informasi yang bersifat, kecuali...

- a. Valid

- b. Bias**

- c. Reliable

- d. Discriminating

Pembahasan : Sifat informasi yang didapat dari kuesioner harus valid, reliable, unbiased, dan discriminating.

3. Kekurangan menggunakan simple random sampling adalah hanya cocok untuk populasi yang relatif homogen. (Benar/Salah)

Pembahasan : Simple random sampling hanya cocok untuk populasi yang relatif homogen dan cakupan yang tidak terlalu luas.

4. Dibawah ini yang bukan merupakan probability sampling adalah...

- a. Multistage sampling

- b. Stratified sampling

- c. Systematic sampling

- d. Snowball sampling**

Pembahasan : Yang termasuk probability adalah simple sampling, systematic sampling, cluster sampling, stratified sampling, dan multistage sampling. Sedangkan snowball termasuk kedalam non probability sampling.

5. Snowball sampling sering digunakan ketika responden merupakan pecandu narkoba, kelompok LGBT+, atau kelompok-kelompok sosial lain yang eksklusif(tertutup).  
(Benar/Salah)

Pembahasan :

Snowball sampling dipakai ketika peneliti tidak banyak tahu tentang populasi penelitiannya. Dia hanya tahu satu atau dua orang yang berdasarkan penilaiannya bisa dijadikan sampel. Karena peneliti menginginkan lebih banyak lagi, lalu dia minta kepada sampel pertama untuk menunjukkan orang lain yang kira-kira bisa dijadikan sampel.

6. Data dalam observasi belum tersedia, data harus dibangkitkan melalui respon.  
(Benar/Salah)

Pembahasan : Data dalam observasi sudah tersedia, hanya perlu proses pengumpulan. Sedangkan data yang belum tersedia dan harus dibangkitkan merupakan data percobaan.

7. Yang bukan tujuan dari pengulangan adalah...

- a. Untuk meningkatkan presisi kesimpulan
- b. Untuk meningkatkan validitas kesimpulan
- c. Untuk menduga galat percobaan
- d. Untuk menduga standard error rata-rata perlakuan

Pembahasan: Untuk meningkatkan validitas kesimpulan merupakan tujuan dari pengulangan.

8. Pengendalian lingkungan dapat dilakukan melalui strategi pengelompokan (blocking). Pengelompokan berhasil jika keragaman dalam kelompok besar (maximum) dan keragaman antar kelompok kecil (minimum). (Benar/Salah)

Pembahasan : Pengelompokan berhasil jika keragaman dalam kelompok kecil (minimum) dan keragaman antar kelompok besar (maximum).

9. Pengaruh dosis pupuk N terhadap produksi jagung. Pupuk N dibedakan menurut dosisnya yaitu 0 kg/ha, 100 kg/ha, dan 200 kg/ha. Seluruh faktor diluar dosis pupuk N dikondisikan serba sama. Rancangan percobaan yang tepat digunakan untuk percobaan diatas adalah...

- a. Faktor tunggal - RAKL
- b. Tersarang - RBSL
- c. Faktor tunggal - RAL
- d. Faktorial – RAKL

Pembahasan: perlakuan yang dicobakan merupakan taraf-teraf dari satu faktor tertentu. Dan kondisi unit percobaan diasumsikan serbasama (homogen).

10. Berikut ini yang benar mengenai rancangan bujur sangkar latin adalah...

- a. Kondisi unit percobaan homogen
- b. Sumber ketidakhomogenan unit percobaan berasal dari satu arah
- c. Pengendalian ketidakhomogenan dapat dilakukan dengan pengelompokan dua arah
- d. Semua jawaban salah

Pembahasan: Kondisi unit percobaan tidak homogen. Sumber ketidakhomogenan unit percobaan berasal dari dua arah. Pengendalian ketidakhomogenan dapat dilakukan dengan pengelompokan dua arah (blok baris dan blok lajur). Jumlah perlakuan = jumlah blok baris = jumlah blok lajur.

11. Tahapan dalam menyisihkan pengamatan duplikat pada data merupakan tahapan coding.

(Benar/Salah)

Pembahasan : Tahap menyisihkan pengamatan duplikat pada data merupakan tahapan data cleaning

12. Data cleaning digunakan untuk meningkatkan kualitas dan validitas data yang dimiliki.

(Benar/Salah)

13. Berikut ini adalah daftar permasalahan dalam manajemen data, kecuali...

- a. Struktur Data tidak terstandarisasi
- b. Entri data yang akurat
- c. Data tersimpan dalam sumber yang berbeda
- d. Kesulitan mencari Data dalam waktu Singkat

Pembahasan : Masalah yang dihadapi dalam manajemen data adalah sebagai berikut.

1. Struktur Data tidak terstandarisasi
2. Entri data yang tidak akurat
3. Data tersimpan dalam sumber yang berbeda
4. Kesulitan mencari Data dalam waktu Singkat

14. Dengan ukuran data yang sangat besar lebih baik menggunakan sistem database.

(Benar/Salah)

Sistem database digunakan untuk ukuran data yang besar (*big data*) agar tidak tercecer dan lebih mudah untuk dianalisis.

15. Diketahui data1 terdapat 6 pengamatan dengan peubah X1, X2, dan X3. Serta diketahui data2 yang berisi 5 pengamatan dengan peubah X4 dan X5. Jika kedua data tersebut digabungkan maka banyaknya pengamatan dan peubah pada data gabungannya adalah...

- a. 11 pengamatan dengan 5 peubah
- b. 10 pengamatan dengan 3 peubah
- c. 11 pengamatan dengan 3 peubah
- d. 10 pengamatan dengan 5 peubah

Pembahasan :

Data 1 berisi :

- 6 pengamatan
- peubah X1, X2, dan X3 = 3 peubah

Data 2 berisi :

- 5 pengamatan
- Peubah X4 dan X5 = 2 peubah

❖ Pengamatan total = Pengamatan data 1 + Pengamatan data 2 = 6+5 = 11 pengamatan

❖ Total peubah = Peubah data 1 + Peubah data 2 = 3+2 = 5 peubah

16. Berikut adalah contoh data yang sudah diurutkan dengan opsi Ascending:

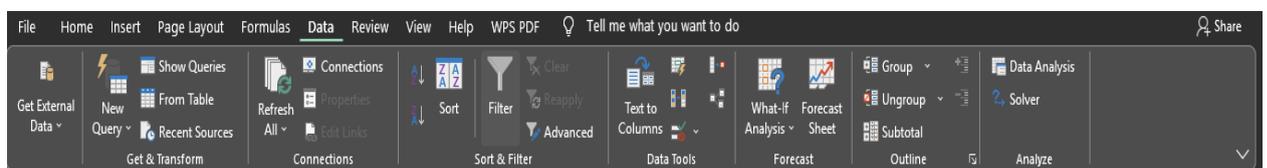
75,82,67,72,59,88,70,51. (Benar/Salah)

Pembahasan : Data yang diatur secara ascending adalah data yang diatur dari data terkecil ke data terbesar. Namun data yang diberikan tidak berurut sesuai dengan opsi ascending (tersusun secara acak).

17. Opsi Sort dan Filter hadir di bawah tab Microsoft Excel mana?

- a. Review tab
- b. Insert tab
- c. Page layout tab
- d. Data tab

Pembahasan : Buka aplikasi Microsoft Excel, kemudian pilih Data tab. Terlihat bahwa ada fitur “Sort and Filter” di bagian Data tab.



18. Descending adalah jenis pengurutan data dari...

- a. besar ke kecil
- b. kecil ke besar
- c. atas ke bawah
- d. kiri ke kanan

Pembahasan : Pengertian dari descending adalah pengurutan data terbesar ke data yang terkecil

19. Pivot Table bukanlah fitur dalam Microsoft Excel yang digunakan dalam agregasi data.

(Benar/Salah)

Pembahasan : Menu di excel yang digunakan untuk melakukan agregasi data adalah pivot table.

20. Suatu tabel data terdiri dari nomor, NIM (Nomor Induk Mahasiswa), nama, kelas paralel, dan nilai kuis. Kemudian ingin disisipkan data tentang nilai UTS setelah kelas paralel, langkah-langkahnya dapat dilakukan sebagai berikut:

- a. posisi kursor pada nilai kuis kemudian Home-> Insert->Insert Sheet Columns
- b. posisi kursor pada nilai kuis kemudian Home->Insert->Insert Cells
- c. posisi kursor pada kelas paralel kemudian Home-> Insert->Insert Sheet Columns
- d. posisi kursor bebas kemudian Home-> Insert->Insert Sheet Columns

Pembahasan : Untuk menyisipkan data di antara gabungan data yang menurun, posisi kursor berada pada data yang akan diposisikan setelah data yang bersangkutan. Setelah itu, tekan "Insert" bagian bawah yang berada di bagian home dan pilih opsi "Insert Sheet Columns" untuk membuat kolom kosong. Setelah itu, sisipkan data yang bersangkutan pada kolom kosong tersebut.

21. Fungsi pemfilteran data digunakan untuk, kecuali...

- a. Menghitung hasil kelompok data tertentu
- b. Mendeteksi pengamatan yang aneh dari analisis
- c. Train dan memvalidasi model statistik yang dibangun untuk suatu analisis
- d. Mengurangi risiko kehilangan informasi penting

Pembahasan : mengurangi risiko kehilangan informasi penting merupakan alasan dari pentingnya manajemen data. Opsi yang lain adalah fungsi dari pemfilteran.

22. Pengacakan bertujuan mendapatkan galat sistematis. (Benar/Salah)

Pembahasan : pengacakan (*randomization*) bertujuan untuk menghindari galat sistematis, bukan mendapatkan galat sistematis.

23. Misalkan pada suatu percobaan faktor tunggal dalam rancangan acak lengkap dimana  $r$  adalah banyaknya ulangan dan  $t$  adalah banyaknya perlakuan. Jika terdapat 3 perlakuan, maka minimal banyaknya ulangan adalah ...

- a. 2
- b. 3
- c. 4
- d. 6

Pembahasan :  $t(r-1) \geq 15 \rightarrow 3(r-1) \geq 15 \rightarrow (r-1) \geq 5 \rightarrow r \geq 6$

Jadi banyaknya ulangan yang dibutuhkan untuk masing-masing perlakuan minimal adalah 6.

24. Visualisasi data hanya berperan setelah proses analisis data selesai dilakukan. (Benar/Salah)

Pembahasan : Visualisasi data berperan selama dan setelah proses analisis data selesai dilakukan.

25. Berikut ini merupakan peranan visualisasi data selama proses analisis data, kecuali...

- a. Meringkas informasi data besar
- b. membangkitkan ide baru analisis data
- c. Eksplorasi data
- d. Elaborasi data

Pembahasan : Meringkas informasi pada data besar bertujuan untuk mudah melihat pola pada data dan memahami cerita di belakangnya. Langkah tersebut dilakukan setelah analisis data.

26. Unsur posisi hanya dapat digunakan untuk memetakan data numerik. (Benar/Salah)

Pembahasan : Unsur posisi dapat digunakan untuk memetakan data numerik maupun kategorik.

27. Dalam data numerik, semakin terang warna maka nilainya semakin kecil (Benar/Salah)

Pembahasan : Pada data yang bisa terurut dengan makna tertentu (yaitu data ordinal dan numerik), dapat digunakan data dengan gradasi untuk membedakan nilai yang kecil dan besar. Data yang bernilai lebih besar digambarkan dalam wujud warna yang semakin pekat atau gelap, sedangkan data yang kecil digambarkan dalam warna yang terang.

28. Yang bukan termasuk visualization dalam data storytelling adalah...

- a. Alur narasi penyampaian
- b. Diagram yang sesuai
- c. Informasi yang ingin disampaikan
- d. Gestalt laws in visual perception

Pembahasan : alur narasi penyampaian merupakan bagian dari Narrative dalam data storytelling.

29. Berikut ini yang dikerjakan ketika menyusun bahan presentasi, kecuali...

- a. Menyiapkan skenario presentasi
- b. Menyiapkan salam pembuka dan penutup
- c. Berdiskusi dan kajian literatur
- d. Membuat slide presentasi

Pembahasan : serangkaian diskusi dan kajian literatur dilakukan pada saat penentuan masalah dan analisis data

30. Dalam pembuatan slide teks, hal-hal yang perlu diperhatikan, kecuali...

- a. Menggunakan teks berupa poin-poin utama saja
- b. Membuang bagian footer yang tidak perlu
- c. Membuat teks dengan jumlah kata yang relatif banyak
- d. Membuat headline yang menarik

Pembahasan : dalam pembuatan slide teks, usahakan teks hanya berisi poin-poin penting dan jumlah kata yang tidak terlalu banyak, agar audiens tidak cenderung hanya membaca slide dan mengabaikan pembicara.

# DAFTAR PUSTAKA

STATISTIKA DAN ANALISIS DATA

## Pekan 1-14

MATERI PERKULIAHAN IPB

Agresti A, Franklin C, Kingenberg B. 2018. Statistics: the art and science of learning from data. Pearson - Harlow, England.

Moore DS, McCabe GP, Craig BA. 2014. Introduction to the Practice of Statistics. WH Freeman and Company - New York, USA.

Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying E. Y. 2011. Probability and Statistics for Engineers and Scientists

David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, James J. Cochran. 2017. Statistics for Business & Economics., Revised South Western Educational Publishing Boston, USA

# Modul Pembelajaran Angkatan 59

## Kunjungi Kami



Asrama PPKU IPB



@asramappku  
@TPBIPB



@tutorsebaya\_ppku  
@asramappku  
@ppkuipb  
@seniorresidentipb



@Tutor Sebaya PKU  
@Official DPKU (PPKU) IPB



<http://ppku.ipb.ac.id/>



# Biodata Penyusunan



Nama : Handito Rahman  
Fakultas : FATETA  
Jurusan : Teknik Sipil dan Lingkungan  
Tanggal lahir : 10 Juni 2003  
Domisili : Kota Bima  
Hobi : Membaca  
Motto hidup : Semua kejadian pasti ada artinya

Nama : Jesika Br Girsang  
Fakultas : FMIPA  
Jurusan : Fisika  
Tanggal lahir: 21 Oktober 2002  
Domisili : Kabupaten Simalungun  
Hobi : Rebahan :v  
Motto hidup : Just do your best



Nama : Muhammad Zhafran Iffat Taqy Arif  
Fakultas : FATETA  
Jurusan : Teknik Sipil dan Lingkungan  
Tanggal lahir: 31 Mei 2003  
Domisili : Kabupaten Maros  
Hobi : Bermain Game, Membaca, mendengarkan musik, berolahraga  
Motto hidup : Jangan menyerah



# Biodata Penyusunan



Nama : Putri Qoonitah Dewi  
Fakultas : FMIPA  
Jurusan : Matematika  
Tanggal lahir : 18 Januari 2003  
Domisili : Kota Depok  
Hobi : Nonton film/drakor dan baca novel  
Motto hidup : Jalanin aja dulu

Nama : Taffyana Izza Alifa  
Fakultas : FAHUTAN  
Jurusan : Konservasi Sumberdaya Hutan dan Ekowisata  
Tanggal lahir : 2 Juli 2003  
Domisili : Bekasi  
Hobi : Doodling, nonton, baca  
Motto hidup : Have courage and be kind, never give up!

