



UNIVERSITAS AIRLANGGA  
Excellence with Morality

# **DATA DAN PUSTAKA (SIP107)**

Pertemuan 4: Membaca, Menyusun, dan Menginterpretasi Data yang  
Disajikan dalam Tabel





# Topik

- Pengantar
- *Tidy data* dan *Pivoting* (panjang [*long*] vs lebar [*wide*])
- Membuat tabel distribusi frekuensi
- Membuat tabel kontijensi (*cross tabulation*)
- Menghitung probabilitas, persentase kolom dan baris, proporsi, rasio, dan *rate* dari tabel kontijensi
- Visualisasi dan tata letak tabel





# Pengantar

- Tabel prinsipnya adalah sekumpulan data yang disusun dalam **baris** (*row*) dan **kolom** (*column*).
- **Baris** merupakan *record* yang dapat berisi informasi mengenai **unit analisis** (*case*), mulai dari usia, gender, hasil pengukuran, atau yang lainnya.
  - **Baris** dalam tabel biasanya bersifat *granular*, artinya mewakili informasi yang paling rendah levelnya dan paling detail.
- **Kolom** merupakan representasi dari variabel.
  - Pastikan tiap variabel terdiferensiasi dengan variabel lainnya.
  - e.g. Tabel yang baik memiliki kolom “penjualan” dan “keuntungan” secara terpisah, bukan menyatukannya dalam kolom “uang” karena “penjualan” dan “keuntungan” adalah dua variabel yang berbeda.





# Apa perbedaannya?

Hari	Jeruk	Apel	Buah Naga
Senin	10	12	15
Selasa	15	14	16
Rabu	15	11	19

Bulan	Jeruk	Apel	Buah Naga
Januari	453	637	234
Februari	456	674	245
Maret	543	654	236

**Tabel Hasil Panen A**

**dengan**

**Tabel Hasil Panen B**





# Unit Analisis

- Tabel A dan Tabel B merupakan dua tabel dengan unit analisis yang berbeda.
  - Tabel A adalah hasil panen **per hari**, sedangkan Tabel B adalah hasil panen **per bulan**.
  - Namun **variabelnya (kolom) sama**, yaitu jumlah buah jeruk, apel, dan buah naga yang dipanen dalam satuan waktu.
- Beberapa tips!
  - Berikan **nomor unik** (*unique identifier*) **pada setiap baris** (*case*) agar setiap baris mewakili unit analisis yang benar-benar unik. (e.g. kalau hanya ada nama hari, bagaimana membedakan jumlah panen pada hari Selasa minggu ini dengan minggu depan?)
  - **Urutkan data** dalam tabel **dengan aturan tertentu** yang **cocok dengan konteks** data. Misalnya, diurutkan secara alfabetis, berdasarkan kategori tertentu, atau dari data yang terkecil ke terendah.





# Mana yang Lebih Tepat?

ID	Hari	Jeruk	Apel	Buah Naga	Jumlah
001	Senin	10	12	15	37
002	Selasa	15	14	20	49
003	Rabu	15	11	19	45

**Diurutkan  
berdasarkan ID**

ID	Hari	Jeruk	Apel	Buah Naga	Jumlah
001	Senin	10	12	15	37
003	Rabu	15	11	19	45
002	Selasa	15	14	20	49

**Diurutkan  
berdasarkan jumlah**





## Prinsip *tidy data* (Wickham, 2014)

Dalam menyimpan data dalam bentuk tabel (set data/*dataset*), gunakan prinsip *tidy*

1. Satu kolom memuat satu variabel
2. Satu baris mewakili satu unit analisis (*case*)
3. Satu tabel mewakili satu *observational unit*

Dengan *tidy data*, maka..

1. Mudah membuat visualisasinya
2. Memudahkan eksplorasi data
3. Mudah “dibersihkan”
4. Tidak berantakan apabila dilakukan koreksi/penyesuaian

Contoh-contoh tabel/set data yang sesuai/melanggar prinsip *tidy* [[klik disini](#)].





## ***Long vs wide***

1. *Long form* (**vertikal**) → tiap **variabel mendapat kolom sendiri**, paling sesuai dengan prinsip *tidy*. Pada banyak kesempatan, lebih disarankan untuk digunakan karena memudahkan data untuk diproses lebih lanjut. Sebagian perangkat lunak statistik (e.g. SPSS, jamovi, JASP, STATA, dsb. memproses data dalam bentuk *long form*, bukan *wide*.
2. *Wide form* (**horizontal**) → tiap **baris berisi data hasil observasi satu unit analisis** (*case*) dan biasanya mengandung pengukuran berulang. Model *wide form* sangat sesuai apabila peneliti ingin melihat tren (berdasarkan periode/waktu tertentu) sehingga sering digunakan untuk mengolah data dari studi *time series* atau *longitudinal*.





## Contoh *long form*

Nama Kucing	Waktu Pemberian Makanan	Jumlah Makanan (gr)
Lorenz	Senin	100
Lorenz	Selasa	120
Sale	Senin	123
Sale	Selasa	125
Wolfgang	Senin	132
Wolfgang	Selasa	130





## Contoh *wide form*

Nama Kucing	Senin	Selasa
Lorenz	100 gr	120 gr
Sale	123 gr	125 gr
Wolfgang	132 gr	130 gr





# Cara Penyusunan Tabel

Penyajian tabel ini bergantung dengan jenis informasi yang diinginkan. Terdapat banyak cara penyusunan tabel antara lain :

1. Penyusunan secara alfabetis
2. Penyusunan secara geografis
3. Penyusunan menurut besaran angka-angkanya
4. Penyusunan secara historis
5. Penyusunan atas dasar kelas-kelas yang lazim
6. Penyusunan secara progresif
7. Tabel distribusi frekuensi
8. Tabulasi silang





# Perhatikan!

## 1. Teknik penyusunan tabel

Pemilihan teknik penyusunan harus disesuaikan dengan informasi yang diinginkan dalam pembuatan tabel.

## 2. Judul tabel

Judul tabel harus jelas menggambarkan karakteristik data dalam tabel.

## 3. Judul kepala tabel

Judul kepala tabel disesuaikan dengan uraian di bawahnya.

## 4. Sumber data

Sumber data harus jelas dan lengkap diletakkan di bawah tabel.





# Perhatikan!

## 5. Persentase

Jika angka persentase dibutuhkan harus dihitung dengan perincian yang jelas.

## 6. Jumlah

Jika angka jumlah dalam tabel merupakan sesuatu yang penting harus ditonjolkan.

## 7. Unit pengukuran

Unit pengukuran data harus dicantumkan, e.g. tahun, Kg, Km, dsb.

## 8. Keterangan

Hal-hal lain yang perlu dicantumkan sebagai bentuk penjelasan. Pada umumnya berbentuk catatan kaki (*footnote*) di bawah tabel.





## Penyusunan Secara Alfabetis

Jumlah kecelakaan akibat kerja di 9 negara tahun 2005

Negara	Jumlah
Amerika Serikat	2160
Australia	2280
Belanda	780
Cina	940
India	2760
Inggris	960
Jepang	2150
Kanada	1960
Spanyol	410
Jumlah Total	14400

Sumber : data hipotetik

Jika data jumlah kecelakaan akibat kerja tersebut meliputi seluruh negara-negara di dunia, maka teknik penyusunan secara alfabetis ini memudahkan pembaca untuk mencari data yang diinginkan

Contoh lain tabel alfabetis

<https://febpwt.webhosting.rug.nl/Dmn/AggregateXs/PivotShow#>





# Penyusunan Secara Geografis

Region	Study	Rates (%)			
		Preeclampsia	Flare	Active disease	Live births
Africa	Alyet al	13	50.5	0.0	76.9
	Mbuli et al	20	36.1	83.3	57.1
	Whitelaw et al	26	12.8	-	59.0
	Dey et al	-	14.3	22.6	76.6
China	Ku et al	-	45.0	-	61.5
India	Aggarwal et al	13	11.1	67.7	46.2
	Chandran et al	8	35.3	47.1	45.5
	Gupta et al	3	29.4	26.7	86.7
South-east Asia	Tan et al	22	33.3	26.5	89.7
	Teh et al	24	13.3	7.1	72.9
	Phadungkiatwattana et al	21	5.8	5.6	85.2
	Sittiwangkul et al	0	27.3	12.5	82.4

[https://www.sjkdt.org/viewimage.asp?img=SaudiJKidneyDisTranspl\\_2020\\_31\\_1\\_10\\_279928\\_t5.jpg](https://www.sjkdt.org/viewimage.asp?img=SaudiJKidneyDisTranspl_2020_31_1_10_279928_t5.jpg)





# Penyusunan menurut besaran angkanya

## By 2100, five of the world's 10 largest countries are projected to be in Africa

*Countries with largest population, in millions*

1950		2020		2100	
China	554	China	1,439	India	1,450
India	376	India	1,380	China	1,065
U.S.	159	U.S.	331	Nigeria	733
Russia	103	Indonesia	274	U.S.	434
Japan	83	Pakistan	221	Pakistan	403
Germany	70	Brazil	213	D.R. Congo	362
Indonesia	70	Nigeria	206	Indonesia	321
Brazil	54	Bangladesh	165	Ethiopia	294
UK	51	Russia	146	Tanzania	286
Italy	47	Mexico	129	Egypt	225

Note: Countries are based on current borders. In this data source, China does not include Hong Kong, Macau or Taiwan.

Source: United Nations Department of Economic and Social Affairs, Population Division, "World Population Prospects 2019."

PEW RESEARCH CENTER





# Penyusunan secara historis

Year	Vehicular population (in '000)	Persons killed (D)	Population (million) (P)	Year	Vehicular population (in '000)	Persons killed (D)	Population (millions) (P)
1971	1865	15,000	560.2675	1989	16,920	50,700	832.535
1972	2045	16,100	573.1299	1990	19,152	54,100	849.515
1973	2109	17,600	586.2198	1991	21,374	56,400	866.53
1974	2327	17,300	599.6427	1992	23,507	57,200	882.821
1975	2472	16,900	613.459	1993	25,299	60,700	899.329
1976	2700	17,800	627.6324	1994	26,464	64,000	915.697
1977	3260	20,100	642.1336	1995	30,125	70,600	932.18
1978	3614	21,800	656.9406	1996	33,786	74,600	948.7589
1979	4059	22,600	672.0209	1997	37,332	77,000	965.4282
1980	4521	24,600	687.332	1998	41,368	79,900	982.1825
1981	5391	28,400	702.8212	1999	44,875	82,000	999.016
1982	6055	30,700	718.4256	2000	48,857	78,900	1015.923
1983	6973	32,800	734.072	2001	54,991	80,900	1032.473
1984	7949	35,100	749.6769	2002	58,924	84,600	1048.641
1985	9170	39,200	765.147	2003	67,007	85,900	1064.399
1986	10,577	40,000	781.893	2004	72,718	92,600	1079.721
1987	12,618	44,400	798.68	2005	81,501	94,900	1094.583
1988	14,818	46,600	815.59	2006	89,618	105,700	1109.811

Source: Ministry of Road Transport & Highways, Government of India ([www.morth.nic.in](http://www.morth.nic.in)).

<https://www.semanticscholar.org/paper/A-comparative-study-on-application-of-time-series-Jha-Sinha/1d2e15c7db73604cf3c0159735bf24222bb60d15/figure/0>





# Penyusunan atas dasar kelas lazim

**Tabel 4.4** Suhu yang Tepat untuk Beberapa Jenis Buah dan Sayur

Bahan	Suhu optimum (°C)
<u>Buah-buahan :</u>	
Alpukat	7.5
Anggur	7.5
Apel	1 – 2
Jeruk	2 – 3
Mangga	10
Nenas	10 – 30
Pepaya	7.5
Pisang	13.5
<u>Sayur-sayuran :</u>	
Buncis	7.5 – 10
Kentang	4,5
Ketimun	7,5
Kol	0
Terung	7 – 10
Tomat hijau	13
Tomat matang	10
Wortel	0 – 1,5





# Penyusunan secara progresif

**World population (mid-year)**  
(million)

	1995	2000	2005	2010	2017 (*)	Annual average growth rate 1995-2017 (%)
<b>World</b>	5 735	6 127	6 520	6 930	7 550	1.3
<b>Asia</b>	3 475	3 714	3 945	4 170	4 504	1.2
<b>Africa</b>	720	814	920	1 044	1 256	2.6
<b>Europe</b>	728	726	729	735	742	0.1
<b>Caribbean, Central and South America</b>	487	527	564	600	646	1.3
<b>Northern America</b>	296	314	329	344	361	0.9
<b>Oceania</b>	29	31	33	36	41	1.5
<b>China</b>	1 228	1 270	1 306	1 341	1 410	0.6
<b>India</b>	961	1 053	1 144	1 231	1 339	1.5
<b>EU-28 (*)</b>	482	487	495	503	512	0.3
<b>United States</b>	266	283	296	310	324	0.9
<b>Japan</b>	124	126	127	127	127	0.1

(\*) 2017: UNPD estimate, medium fertility scenario

(\*) EU-28: population as of 1 January

Source: Eurostat (online data code: demo\_pjan) and United Nations Population Division





# Tabel distribusi frekuensi

Tabel distribusi frekuensi adalah tabel yang memuat data frekuensi dari tiap variabel sehingga setiap *case* dalam tabel memuat frekuensi (atau kejadian - *occurrence*) dari nilai tertentu.

Tabel distribusi frekuensi dapat memudahkan peneliti untuk memvisualisasikan data yang tersaji dalam tabel menjadi, misalnya, *line graph*, *bar chart*, histogram, dan lain-lain.

Berikut adalah contoh cara membuat tabel distribusi frekuensi.





# Tabel Kontijensi

- Tabel distribusi frekuensi cocok digunakan untuk menyajikan data dari satu variabel (univariat).
- Namun seringkali peneliti berkebutan dengan data lebih dari satu variabel (multivariat).
- Dalam kasus dua variabel (bivariat), maka tabel kontijensi adalah teknik terbaik dalam menyajikan, bahkan untuk memproses data dalam kasus tertentu.
- Tabel kontijensi merupakan tabel (dalam format matriks) yang menyajikan distribusi frekuensi dari dua variabel.
- Tabel kontijensi dapat membantu peneliti untuk mengestimasi kemungkinan adanya keterkaitan antara dua variabel dan interaksi antara (kedua) variabel tersebut.
- Peneliti juga dapat menghitung probabilitas kondisional.





# Informasi dalam tabel kontijensi

Umumnya, tabel kontijensi berisi:

1. **Kolom dan baris** mewakili variabel yang berbeda. Misalnya, variabel “menyukai/tidak menyukai K-Pop” sebagai kolom dan “gender” sebagai baris.
  - a. Oleh karena itu, jenis **variabel yang dapat disajikan** dalam tabel kontijensi adalah variabel nominal (kategorikal -- ingat materi minggu lalu tentang jenis variabel!)
2. Jumlah kolom dan baris **mengikuti jumlah kategori** di tiap variabel. Misalnya, variabel “gender” (untuk diletakkan sebagai baris) memiliki dua kategori (laki-laki dan perempuan), sehingga akan ada dua baris.





# Informasi dalam tabel kontijensi

4. Berisi *nets* atau *netts*, yang artinya **subtotal** pada bagian baris dan kolom.
5. Memuat informasi (seluruhnya atau sebagian); persentase, persentase kolom, dan persentase baris.
6. Jumlah sampel ( $n$ )

Berikut ini adalah contoh tabel kontijensi dengan persentase baris dan kolom.





# Menghitung berbagai bentuk probabilitas

Tabel kontijensi memungkinkan peneliti untuk menghitung berbagai bentuk probabilitas, misalnya:

## 1. Rasio

- a. Rasio adalah besaran relatif atau perbandingan dari dua nilai.
- b. Dapat dihitung dengan membagi satu variabel skala interval atau rasio dengan variabel lainnya, namun pembilang dan penyebut tidak harus berhubungan.
- c. Oleh karena itu, kita dapat membandingkan rasio laki-laki dengan perempuan dan rasio laki-laki dengan penggemar K-Pop.
- d. **Contoh rasio** → “Rasio antara remaja putri yang menyukai K-Pop dengan yang tidak adalah 5 : 1”





# Menghitung berbagai bentuk probabilitas

## 2. Proporsi

- a. Proporsi adalah perbandingan sebagian dengan keseluruhan.
- b. Proporsi adalah jenis rasio di mana pembilang termasuk dalam penyebut.
- c. Proporsi dapat dinyatakan sebagai desimal, pecahan, atau persentase.
- d. **Contoh proporsi** → “35% penduduk Indonesia berusia 18-30 tahun mengalami buta aksara fungsional.”





# Menghitung berbagai bentuk probabilitas

## 3. *Rate*

- a. *Rate* merupakan salah satu bentuk pengukuran frekuensi yang menggambarkan suatu kejadian di suatu populasi dalam satu periode waktu tertentu.
- b. Dalam Ilmu Epidemiologi, *rate* merupakan ukuran yang sangat informatif karena *rate* memberikan informasi berapa banyak orang yang menderita penyakit pada satu kelompok demografi (subpopulasi) tertentu.
- c. Oleh karena itu, *rate* sering digunakan untuk menggambarkan tingkat risiko (dalam menderita penyakit).
- d. **Contoh *rate*** → “perempuan lebih berisiko menjadi korban *begal* karena pada tahun 2020 saja, ada sekitar 300 perempuan korban *begal* per 1000 perempuan di Indonesia.”

Berikut ini adalah contoh penghitungan dan interpretasi dari proporsi, rasio, dan *rate*.





# Korelasi dan tabel kontijensi

Tabel kontijensi juga dapat memungkinkan peneliti mengukur **korelasi antara dua variabel** (nominal) dengan berbagai teknik, misalnya:

1. *Odds ratio* (OR)
2. *Phi coefficient*
3. *Cramer's V*
4. Koefisien *lambda*
5. *Tetrachoric* dan *polychoric correlation*

Namun kelimanya tidak dicakup dalam mata kuliah ini.





# Latihan Mandiri

1. Silakan kerjakan tugas yang ada di spreadsheet ini secara mandiri, diluar jam perkuliahan, sebagai latihan mandiri.
2. Tugas **tidak perlu dikumpulkan** ke dosen pengampu.
3. Tugas **tidak dinilai** oleh dosen pengampu.





# Visualisasi dan tata letak tabel

Berikut ini adalah beberapa tips mengatur tata letak tabel:

1. Perhatikan gaya penulisan ilmiah yang diikuti
  - a. Misalnya, tata letak tabel yang mengikuti gaya American Psychological Association (APA) mungkin berbeda dengan American Medical Association (AMA).
2. Kapan saatnya menggunakan garis vertikal sebagai batas antar kolom?
  - a. Umumnya, tabel disajikan dengan garis horizontal yang tegas (tanpa garis vertikal). Itupun hanya untuk baris pertama sebagai pemisah antara nilai dengan label kolom.
  - b. Anda ***hanya*** disarankan untuk menggunakan model *spreadsheet* (yaitu tabel dengan garis vertikal dan horizontal yang tegas) ketika menyajikan tabel dengan informasi yang padat.





# Visualisasi dan tata letak tabel

3. Gunakan petunjuk visual
  - a. Anda dapat menggunakan warna latar (*shading*) yang berbeda untuk masing-masing baris (atau sel) agar memudahkan pembaca untuk memeriksa tabel.
4. Gunakan *tabular numeral* ketika menyajikan informasi berupa angka agar informasi lebih mudah dibaca
  - a. Gunakan *font* seperti Courier, Courier New, Lucida Console, Monaco, dst.

Proportional	Tabular
390,209,000	390,209,000
112,371,000	112,371,000