

**ANALISIS *NUMERICAL SUMMARY* PADA DATA TABULAR DAN DATA SPASIO-
TEMPORAL**

TUGAS TUTORIAL 02

MATA KULIAH ANALISIS DATA CUACA DAN IKLIM

Oleh:

Fardhan Indrayesa

12821046



PROGRAM STUDI METEOROLOGI

FAKULTASI ILMU DAN TEKNOLOGI KEBUMIHAN

INSTITUT TEKNOLOGI BANDUNG

2023

1. Tugas 1: Menghitung *Numerical Summary* Tabel A1

Tabel A1 adalah tabel dalam ekstensi file .txt dan berisi tabel yang memiliki kolom tanggal (*date*), presipitasi (*precip1* dan *precip2*), temperatur maksimum (*tmax1* dan *tmax2*), dan temperatur minimum (*tmin1* dan *tmin2*).

Numerical Summary yang digunakan:

- *Location*: *Mean*, *median*, *trimean*, *trimmed mean*
- *Spread*: Standar deviasi, *Interquartile Range* (IQR), *Median Absolute Deviation* (MAD)
- *Symmetry*: *Skewness*, Yule-Kendall

Tampilan data tabel A1:

| | precip1 | tmax1 | tmin1 | precip2 | tmax2 | tmin2 |
|---|---------|-------|-------|---------|-------|-------|
| 0 | 0.00 | 33 | 19 | 0.00 | 34 | 28 |
| 1 | 0.07 | 32 | 25 | 0.04 | 36 | 28 |
| 2 | 1.11 | 30 | 22 | 0.84 | 30 | 26 |
| 3 | 0.00 | 29 | -1 | 0.00 | 29 | 19 |
| 4 | 0.00 | 25 | 4 | 0.00 | 30 | 16 |

a. *Location*

1) *Mean*

| | Mean |
|---------|-----------|
| precip1 | 0.101613 |
| tmax1 | 29.870968 |
| tmin1 | 13.000000 |
| precip2 | 0.077419 |
| tmax2 | 31.774194 |
| tmin2 | 20.225806 |

Berdasarkan data di atas, wilayah 1 cenderung lebih basah dibandingkan di wilayah 2. Hal ini dapat dilihat pada data rata-rata di atas, bahwa presipitasi wilayah 1 lebih tinggi daripada wilayah 2. Selain itu, temperatur maksimum dan minimum di wilayah 1 cenderung lebih rendah dibandingkan temperatur maksimum dan minimum di wilayah 2. Oleh sebab itu, berdasarkan data di atas, dapat disimpulkan bahwa di wilayah 1 curah hujan cenderung lebih tinggi daripada wilayah 2.

Metode ini memberikan informasi yang cukup mudah dipahami, tetapi tidak dapat mengatasi data pencilon karena dalam metode ini semua nilai data dilibatkan dalam perhitungan.

2) Median

| Median | |
|---------|-------|
| precip1 | 0.00 |
| tmax1 | 30.00 |
| tmin1 | 19.00 |
| precip2 | 0.02 |
| tmax2 | 31.00 |
| tmin2 | 23.00 |

Berdasarkan data di atas, presipitasi wilayah 2 lebih tinggi sedikit daripada di wilayah 1, tetapi temperatur maksimum dan minimum di wilayah 1 lebih rendah daripada di wilayah 2. Hal ini dapat terjadi karena faktor presipitasi bukan hanya dari hujan saja, tetapi dapat juga disebabkan oleh kelembapan udara. Oleh sebab itu, presipitasi di wilayah 2 dapat lebih tinggi daripada di wilayah 1.

Metode ini dapat memberikan informasi tentang data tengah yang tidak dipengaruhi oleh data pencilan. Oleh karena itu, metode ini lebih *robust* dan *resistant* dibandingkan metode *mean*.

3) Trimean

| Trimean | |
|---------|----------|
| precip1 | 0.01250 |
| tmax1 | 29.75000 |
| tmin1 | 15.62500 |
| precip2 | 0.02375 |
| tmax2 | 31.37500 |
| tmin2 | 21.62500 |

Berdasarkan data di atas, data presipitasi di wilayah 1 lebih rendah daripada di wilayah 2, tetapi temperatur maksimum dan minimum di wilayah 2 lebih tinggi daripada di wilayah 1. Hal ini juga dapat terjadi disebabkan oleh faktor lain, yaitu seperti kelembapan dan tekanan udara. Selain itu, penyebab lain presipitasi di wilayah 2 lebih tinggi adalah, saat sebelum terjadi hujan, awan yang mengandung air hujan mengeluarkan panas untuk mengubah uap air menjadi air. Oleh karena itu, mungkin saja sebelum terjadinya hujan, temperatur di wilayah 2 dapat lebih tinggi daripada wilayah 1.

Metode ini memberikan informasi yang lebih akurat dibandingkan dengan metode-metode sebelumnya. Metode ini menggunakan bobot yang lebih besar pada median dan kuartil atas dan bawah, sehingga tidak memperhitungkan data pencilan.

4) *Trimmed Mean*

| Trimmed Mean | |
|--------------|---------|
| precip1 | 0.0356 |
| tmax1 | 29.7200 |
| tmin1 | 14.1200 |
| precip2 | 0.0344 |
| tmax2 | 31.8000 |
| tmin2 | 20.7200 |

Berdasarkan data di atas, data presipitasi lebih besar di wilayah 1 daripada di wilayah 2. Metode ini memotong data sebesar proporsi yang diberikan *user*, sehingga dapat menghindari data pencilan dan hanya mengambil data yang tepat. Oleh karena itu, rata-rata presipitasi di kedua wilayah lebih terpusat di satu nilai, sehingga kedua wilayah memiliki data presipitasi yang saling mendekati.

b. *Spread*

1) Standar Deviasi

| Standar Deviasi | |
|-----------------|-----------|
| precip1 | 0.242887 |
| tmax1 | 7.714670 |
| tmin1 | 13.618615 |
| precip2 | 0.167650 |
| tmax2 | 7.864306 |
| tmin2 | 8.807988 |

Berdasarkan data di atas, nilai standar deviasi terbesar adalah variabel temperatur minimum di wilayah 1. Nilai tersebut didapat karena pada data temperatur minimum di wilayah 1 nilainya sangat beragam. Selain itu, pada data presipitasi di wilayah 2, standar deviasinya paling kecil karena tingkat keragaman nilai data tidak terlalu jauh dengan nilai rata-ratanya.

Metode ini tidak *robust* dan tidak *resistant* karena semua data dilibatkan dalam perhitungan, sehingga menyebabkan data pencilan dapat mempengaruhi berubahnya nilai standar deviasi pada suatu data.

2) *Interquartile Range (IQR)*

| IQR | |
|---------|--------|
| precip1 | 0.050 |
| tmax1 | 7.000 |
| tmin1 | 22.500 |
| precip2 | 0.055 |
| tmax2 | 8.500 |
| tmin2 | 12.500 |

Berdasarkan data di atas, variabel dengan IQR terbesar adalah temperatur minimum wilayah 1. Nilai ini diperoleh karena sebagian besar dari sekitar 25% (kuartil bawah) data nilainya berada cukup jauh dari nilai 75% (kuartil atas) data. Metode ini dapat digunakan untuk mencari sebaran data berdasarkan nilai data yang berada 25% dan 75% dari data. Nilai pencilan data harus berada di bawah 25% data atau di atas 75% data agar dapat memperoleh hasil sebaran data yang tepat.

3) *Median Absolute Deviation (MAD)*

| MAD | |
|---------|------|
| precip1 | 0.00 |
| tmax1 | 4.00 |
| tmin1 | 8.00 |
| precip2 | 0.02 |
| tmax2 | 4.00 |
| tmin2 | 5.00 |

Data di atas diperoleh dari median suatu data yang telah dimutlakkan dan dikurangkan dengan median data asli. Nilai MAD terbesar berada di variabel temperatur minimum wilayah 1 dan nilai MAD terkecil berada di variabel presipitasi wilayah 1. Metode ini cukup *robust* dan *resistant* karena menghasilkan nilai sebaran yang relatif lebih kecil dibandingkan metode lainnya.

c. *Symmetry*

1) *Skewness*

| Skewness | |
|----------|-----------|
| precip1 | 3.016924 |
| tmax1 | 0.342515 |
| tmin1 | -0.615762 |
| precip2 | 3.438342 |
| tmax2 | -0.097653 |
| tmin2 | -0.420895 |

Data di atas adalah nilai kemiringan suatu data atau nilai kecondongan data ke nilai terbesar atau terkecil. *Skewness* positif menandakan bahwa median data lebih kecil dari rata-ratanya, sedangkan untuk *skewness* negatif sebaliknya, median data lebih besar dari rata-ratanya. *Skewness* yang mendekati nilai 0 disebut sebagai data yang hampir simetris. Variabel yang memiliki nilai *skewness* paling jauh dengan 0 adalah presipitasi di wilayah 2 dan temperatur minimum di wilayah 1. Data presipitasi di wilayah 2 memiliki median yang lebih kecil dari nilai rata-ratanya dan datanya cenderung paling banyak berada di nilai terkecil. Data temperatur minimum di wilayah 1 memiliki median yang lebih besar dari nilai rata-ratanya dan datanya cenderung paling banyak berada di nilai terbesar.

Metode ini dapat dipengaruhi oleh nilai pencilan karena dalam formulanya terdapat perhitungan standar deviasi yang melibatkan semua data.

2) Yule-Kendall

| Yule-Kendall | |
|--------------|-----------|
| precip1 | 1.000000 |
| tmax1 | -0.142857 |
| tmin1 | -0.600000 |
| precip2 | 0.272727 |
| tmax2 | 0.176471 |
| tmin2 | -0.440000 |

Nilai YK menjelaskan kemiringan atau kecondongan suatu data ke nilai terbesar atau terkecil. Arti nilai positif dan negatif YK sama seperti *skewness* dan nilai 0 menunjukkan bahwa data simetris. Nilai YK diperoleh berdasarkan selisih dari selisih kuartil atas dan bawah terhadap median dan hasilnya dibagi dengan IQR yang menyebabkan data pencilan yang berada 25% atau 75% dari data tidak dilibatkan dalam perhitungan. Variabel nilai YK yang paling jauh dengan 0 adalah data presipitasi di wilayah 1 dan data temperatur minimum di wilayah 1. Artinya, terdapat banyak data presipitasi yang bernilai kecil dan data temperatur minimum yang bernilai besar. Metode ini lebih *robust* dan *resistant* dibandingkan metode *skewness*.

2. Tugas 2: Mempetakan *Numerical Summary* Data Spasio-Temporal Setiap Musim







Data yang digunakan adalah data presipitasi per bulan (mm/bulan) tiga dimensi (*time*, *lat*, *lon*) dengan dimensi *time* dimulai dari tanggal 1979-02-01 sampai 2020-11-01, dimensi *lat* dimulai dari 14.95°S sampai 14.95°N (langkah 0.1), dan dimensi *lon* dimulai dari 90.05°E sampai 149.95°E (langkah 0.1). *Numerical Summary* yang digunakan adalah *trimean* (*location*) dan *skewness* (*symmetry*).

Tampilan data:

```
xarray.DataArray 'precipitation' (time: 502, lat: 300, lon: 600)
```

```
0.4375 0.25 0.25 0.25 0.3125 0.3125 ... 59.38 61.0 58.0 64.06 65.44
```

▼ Coordinates:

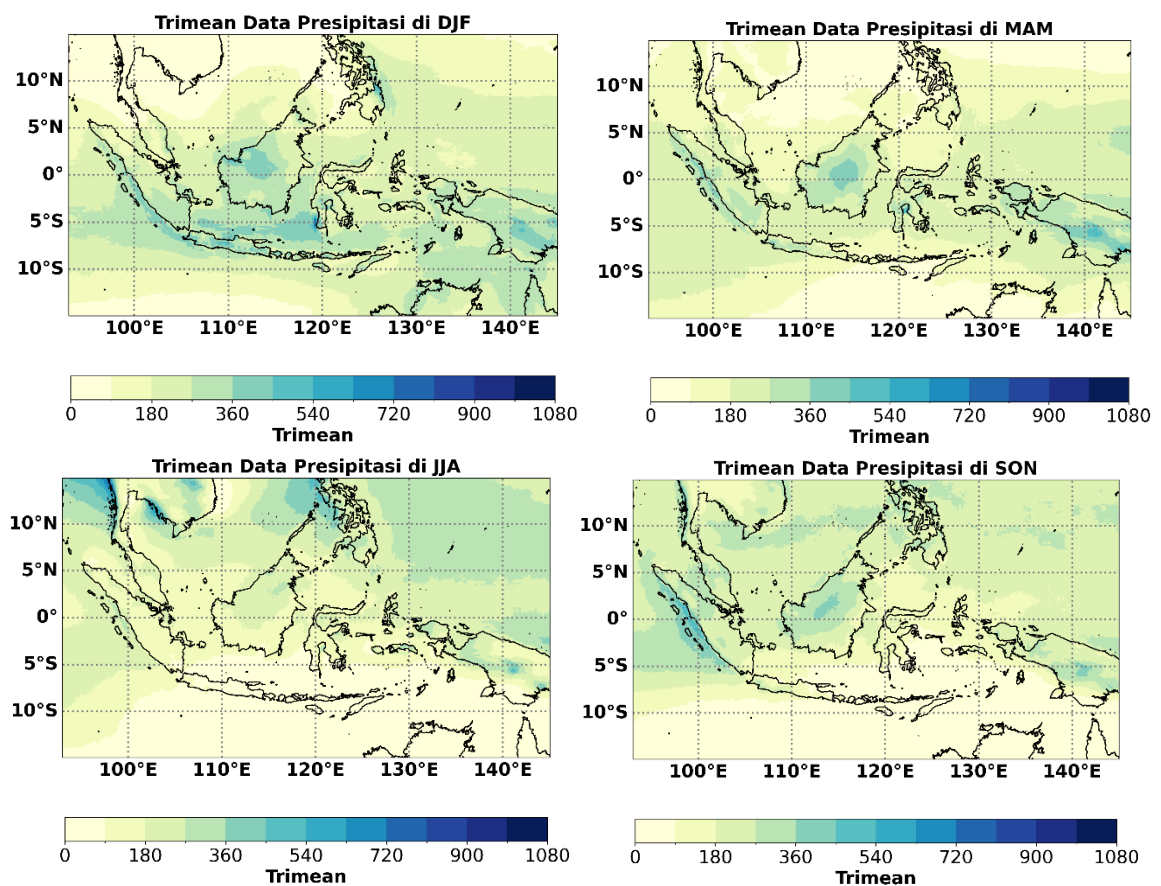
| | | | | |
|------|--------|----------------|-------------------------------------|---|
| time | (time) | datetime64[ns] | 1979-02-01 ... 2020-11-01 |   |
| lon | (lon) | float32 | 90.05 90.15 90.25 ... 149.9 149.9 |   |
| lat | (lat) | float32 | 14.95 14.85 14.75 ... -14.85 -14.95 |   |

► Indexes: (3)

▼ Attributes:

units : mm month-1

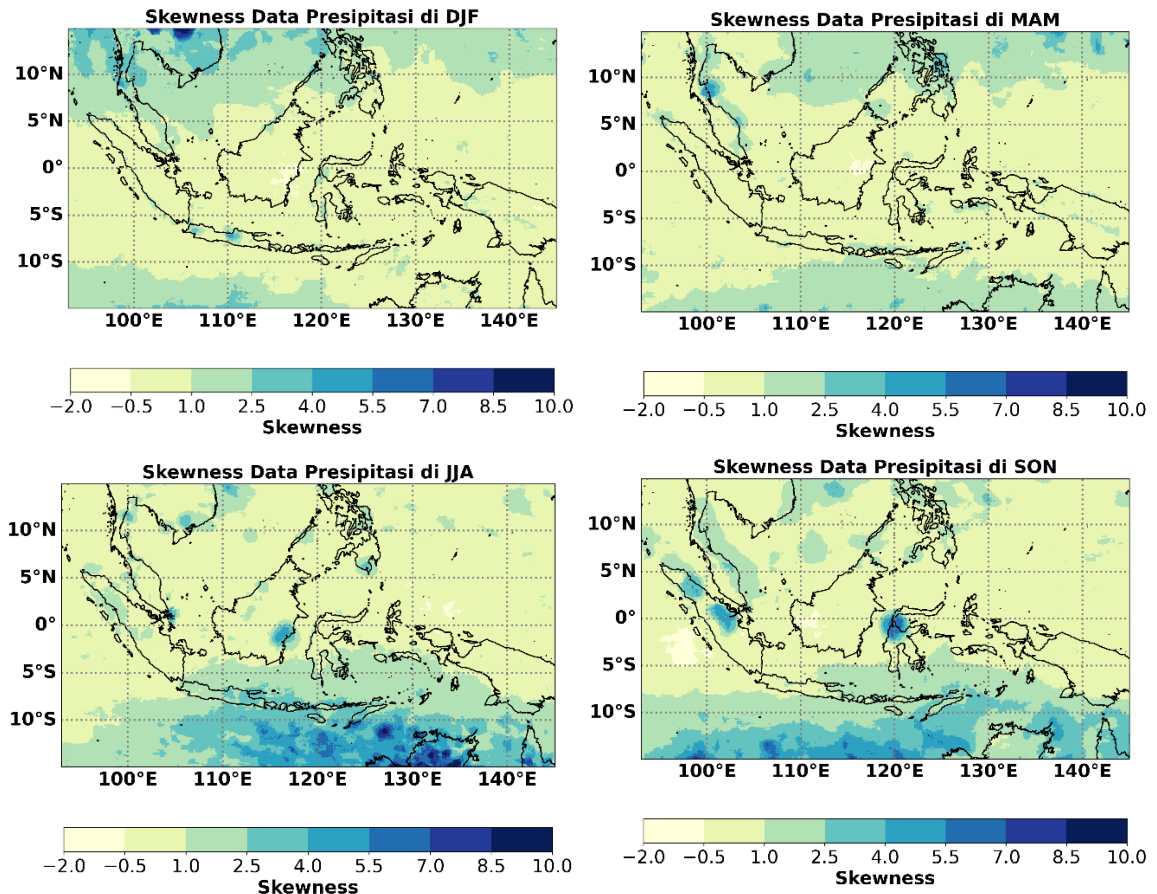
a. *Location: Trimean*



Peta di atas menunjukkan data presipitasi di wilayah Indonesia. Pada musim DJF, terlihat bahwa Indonesia memiliki rata-rata presipitasi yang tinggi dan tersebar di seluruh wilayah. Selain itu, pada musim ini, angin monsun barat dari Asia bertiup ke arah tenggara hingga Indonesia, yang menyebabkan musim hujan di Indonesia. Pada musim MAM pun terlihat bahwa wilayah Indonesia rata-rata masih memiliki preipitasi yang tinggi. Musim MAM adalah musim peralihan, sehingga rata-rata presipitasi masih cenderung tinggi. Beralih ke musim JJA, Indonesia sudah mulai memiliki rata-rata presipitasi yang rendah dan menyebabkan musim kemarau di Indonesia. Pada musim ini, angin monsun bertiup

dari arah Australia ke Indonesia yang menyebabkan musim kemarau di Indonesia. Musim peralihan SON sudah mulai terjadi presipitasi di beberapa wilayah, menandakan Indonesia akan terjadi musim hujan kembali pada musim selanjutnya. Metode yang digunakan untuk mengetahui kepusatan data adalah trimean karena metode ini dapat menghindari nilai data yang terdapat suatu pencilan.

b. *Symmetry: Skewness*



Peta di atas menunjukkan nilai sebaran presipitasi di wilayah Indonesia. Semakin jauh nilai sebaran dari 0, maka presipitasi dapat terjadi dengan nilai sangat kecil atau sangat besar dengan frekuensi yang tinggi di suatu wilayah. Berdasarkan peta di atas, dapat dilihat wilayah mana yang paling sering terjadi hujan dengan intensitas besar atau kecil dengan frekuensi yang tinggi. Pada musim DJF, terlihat bahwa wilayah Indonesia memiliki frekuensi yang cukup besar untuk presipitasi di atas rata-rata, sedangkan wilayah di sebelah utara dan selatannya hanya memiliki presipitasi di bawah rata-rata. Selain itu, di musim ini juga ITCZ cenderung berada di ekuator yang menyebabkan wilayah Indonesia cenderung memiliki curah hujan tinggi dengan frekuensi yang sering. Angin monsun dari Asia juga dapat terjadi di musim ini, sehingga menyebabkan wilayah Indonesia cenderung memiliki intensitas hujan dan frekuensi yang tinggi. Musim MAM adalah musim peralihan ke musim kemarau di Indonesia. Pada musim ini, wilayah Indonesia masih banyak terjadi hujan dengan intensitas dan frekuensi yang tinggi. Selanjutnya, pada musim JJA, bagian Selatan Indonesia sebagian besar sudah mulai jarang terjadi curah hujan dengan intensitas yang

tinggi. Hal ini juga diperkuat bahwa pada musim ini, angin monsun dari Australia bertiup, sehingga menyebabkan beberapa wilayah di Indonesia mengalami musim kemarau. Pada musim peralihan SON, bagian Selatan Indonesia sudah mulai terjadi curah hujan dengan intensitas yang tinggi. Metode yang digunakan untuk mengetahui kesimetrisan data adalah *skewness* karena metode ini melibatkan semua data dalam perhitungan, sehingga hasil yang diperoleh berasal dari data asli. Tapi, metode ini tidak *robust* dan tidak *resistant* karena dapat saja terdapat data pencilan yang akan sangat berpengaruh pada hasil yang diperoleh.

Kesimpulan:

Ke-*robust*-an dan ke-*resistant*-an hasil *numerical summary* yang diperoleh bergantung berdasarkan metode mana yang diambil. Setiap metode memiliki kelebihan dan kekurangan masing-masing. Metode yang sangat *robust* dan *resistant* dapat saja tidak melibatkan seluruh data, sehingga hasil yang diperoleh tidak murni dan cenderung kurang tepat. Metode yang kurang *robust* dan *resistant* dapat saja melibatkan seluruh data, tetapi apabila terdapat suatu data pencilan, hasil yang diperoleh akan sangat jauh dari hasil yang seharusnya. Jadi, dalam memilih suatu metode dalam *numerical summary*, dapat diperhatikan kelebihan dan kelemahan dari masing-masing metode, sehingga akan mendapatkan hasil yang optimal.