Pengenalan Statistika

Secara definisi, statistika adalah bidang studi yang mempelajari bagaimana mengumpulkan dan menganalisa data. Jika mengambil definisi yang lebih lengkap, maka statistika dapat didefinisikan sebagai ilmu yang mempelajari cara pengumpulan data, menganalisis data untuk mendapatkan kesimpulan informasi sampai dapat dijadikan dasar pembuatan keputusan atau kebijakan.

Di dalam mempelajari statistika, akan mempelajari bagaimana caranya **mengorganisir** dan **membuat kesimpulan** dari data. Kegiatan mengorganisir dan membuay kesimpulan dari data disebut sebagai **statistika deskriptif**. Ada banyak cara bagaimana kita menyimpulkan suatu data, misalnya dengan membuat **grafik** atau dengan **angka**, misalnya mengambil suatu **rata-rata** dari data.

Selanjutnya ketika harus mengambil keputusan dari data yang sudah diolah, kemudian akan menggunakan **statistika inferensial** sehingga dapat mengambil keputusan yang **benar** dari data yang juga sudah diolah dengan **benar**.

Beberapa Konsep Dasar Statistika

Populasi, sampel dan observasi/pengamatan

Observasi: Adalah suatu unit yang diukur dengan data. Beberapa contoh diantaranya adalah:

- Siswa
- Warga negara
- Hewan
- Kendaraan

Populasi: Adalah koleksi dari keseluruhan observasi. Beberapa contoh diantaranya adalah:

- Semua siswa yang ada di sekolah
- Semua warga negara Indonesia,
- Semua hewan yang ada di hutan lindung,
- Semua kendaraan di Jakarta

Sampel: Adalah sub koleksi dari populasi. Beberapa contoh diantaranya adalah:

- 5 siswa dari masing-masing kelas di suatu sekolah
- 100 warga negara Indonesia yang diambil dari beberapa wilayah
- Beberapa species hewan tertentu di suatu hutan lindung
- 3 jenis kendaraan di Jakarta

Statistik dan Parameter

Setiap kita mengkaji suatu permasalahan, biasanya kita menggunakan beberapa individu dari grup-grup tertentu. Misalnya, ketika kita ingin mengetahui prestasi siswa di suatu sekolah untuk mata pelajaran matematika, kita bisa saja menghitung semua nilai siswa lalu ambil nilai rata-ratanya. Atau dalam konteks pemasaran, kita ingin tahu untuk setiap segmen pasar, berapa besar pendapata yang bisa kita peroleh dari tiap segmen. Namun ada kalanya kita hanya butuh sebagian kecil dari grup dikarenakan beberapa keterbatasan seperti biaya pengambilan data yang terlalu mahal atau bisa karena membutuhkan waktu analisis yang singkat karena harus membuat keputusan saat itu juga. Jika kasus seperti ini biasanya kita hanya mengambil sebagian kecil dari grup. Misalnya kita hanya mengambil 10 siswa dari tiap kelas untuk menghitung prestasi siswa atau kita hanya mengambil 20% dari tiap segmen pasar kita. Hal ini yang memunculkan dua istilah untuk kedua kasus ini: **parameter** dan **statistik**

Parameter adalah **penjelasan atas populasi** sedangkan statistik hanya **menjelaskan sampel dari populasi**. Untuk kasus terkait segmentasi pasar, mengukur revenue dari setiap orang di dalam segmen adalah parameter sedangkan jika kita hanya mengambil 20% dari tiap segmen hal ini dikatakan sebagai statistik.

Data Kualitatif dan Data Kuantitatif

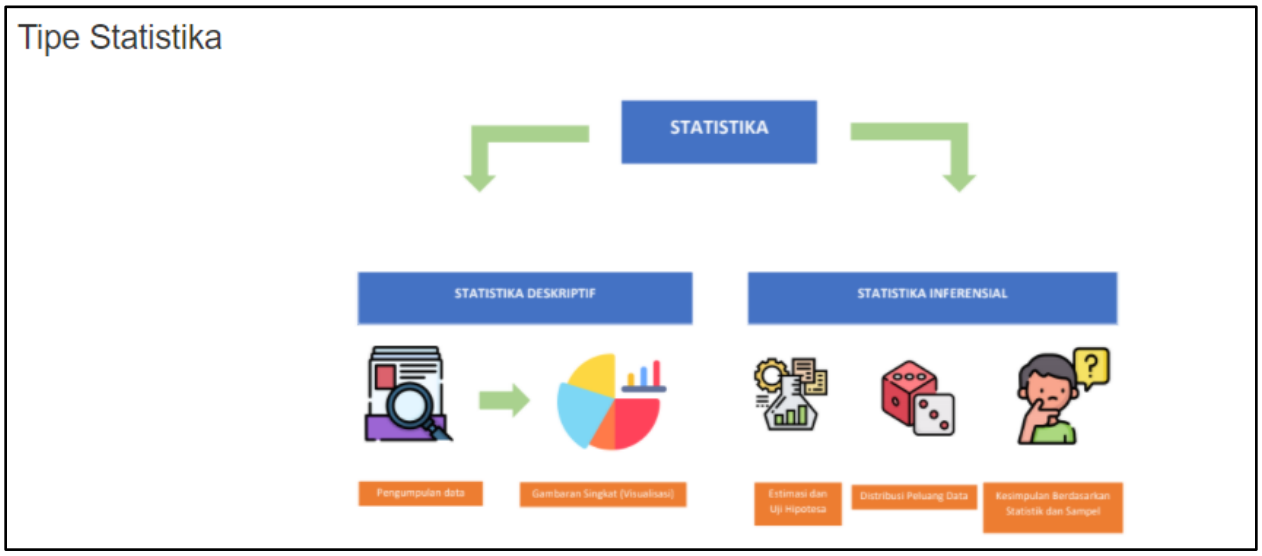
Terdapat dua kategori data yang terdapat pada populasi atau sampel, yaitu data kualitatif dan data kuantitatif.

Data kualitatif adalah data yang diperoleh dari mengkategorikan atau menjelaskan suatu atribut dari populasi atau sampel. Biasanya disebut juga sebagai data kategorik. Beberapa contoh data kualitatif misalnya warna rambut, golongan darah, nama jalan, nama produk yang digunakan dan lain sebagainya.Biasanya data kualitatif selalu disebutkan dalam bentuk kata ataupun simbol.

Data kuantitatif adalah data yang diperoleh dari ukuran atau hitungan di suatu populasi atau sampel. Data kuantitatif selalu berbentuk angka. Data seperti gaji, berat badan, populasi di suatu negara, dan jumlah pelanggan yang dimiliki suatu *e-commerce* termasuk data kuantitatif. Data kuantitatif sendiri dapat dibagi menjadi dua yaitu data diskrit dan data kontinu.

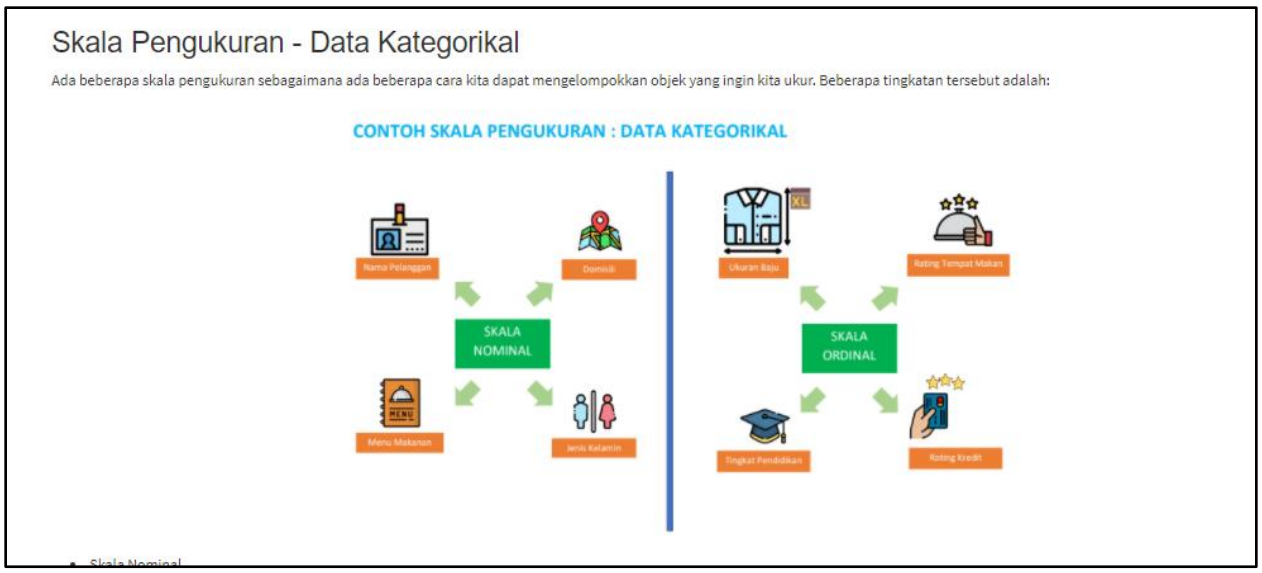
Data yang diperoleh dari hasil perhitungan adalah data diskrit: jumlah pelanggan, jumlah produk, dan jumlah telefon yang diterima oleh *customer service* per harinya adalah beberapa contoh data diskrit.

Data yang diperoleh dari hasil perhitungan namun dapat memuat rasio, desimal, atau bilangan irasional adalah data kontinu: berat badan, tinggi badan, waktu, dan gaji adalah beberapa contoh data kontinu.



Statistika deskriptif digunakan untuk **melakukan eksplorasi pada data**, biasanya menggunakan teknik **visualisasi data** sebagai alat bantu untuk memahami bagaimana bentuk distribusi dan hubungan antara satu titik data dengan titik data lainnya.

Statistika inferensial digunakan untuk **melakukan pemgambilan keputusan** atas suatu simpulan terkait dengan data yang sedang dianalisa. Statistika inferensi memungkinkan kita **mengambil kesimpulan dari suatu populasi dengan menggunakan sampel yang diambil dari populasi** tersebut.

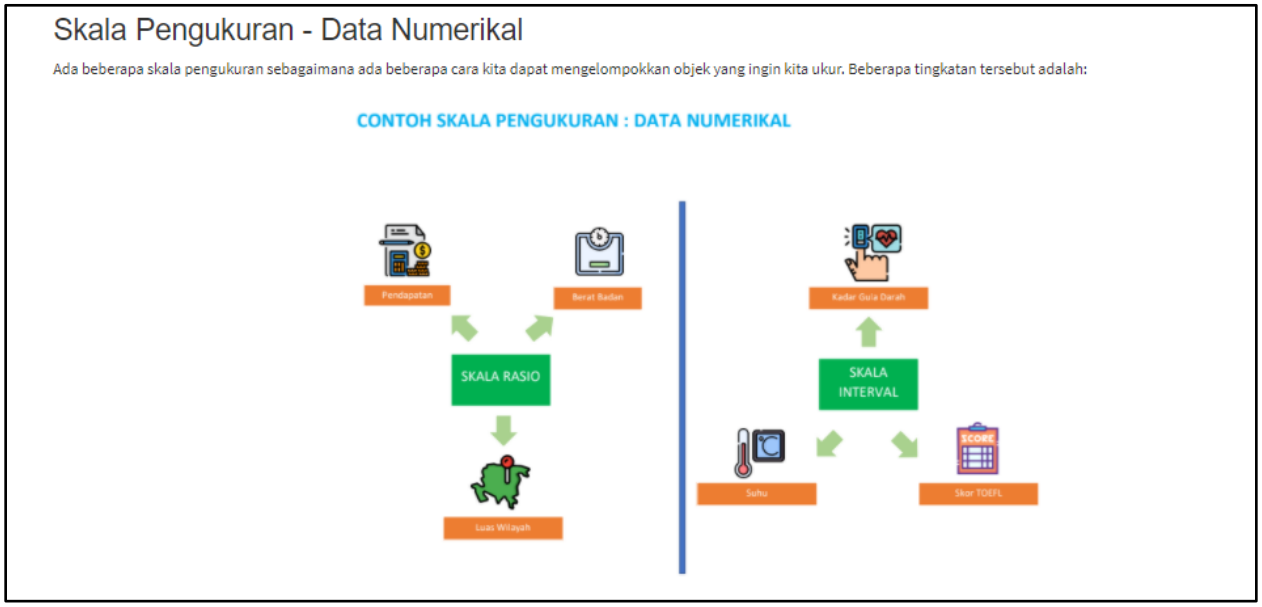


- Skala Nominal

Skala nominal adalah skala yang digunakan untuk **mengkategorikan suatu objek pengamatan dengan objek pengamatan lainnya**. Sebagai contoh yang termasuk skala nominal seperti gender, kategori barang, ras, status pernikahan dan lain sebagainya

- Skala Ordinal

Skala ordinal adalah skala yang digunakan untuk **mengurutkan suatu objek pengamatan** dimana suatu titik pengamatan memiliki nilai yang lebih rendah atau tinggi dibanding nilai lainnya. Sebagai contoh yang termasuk skala nominal adalah kelas, jabatan, tingkat pendidikan, dan sebagainya



- Skala Interval

Skala interval adalah skala yang digunakan untuk tidak hanya untuk mengklasifikasikan maupun memberikan tingkatan pada suatu titik pengamatan, namun kita dapat **mengukur seberapa besar nilai antara suatu titik pengamatan dengan titik pengamatan lainnya**. Beberapa contoh yang termasuk skala interval diantaranya adalah suhu tubuh dan jarak

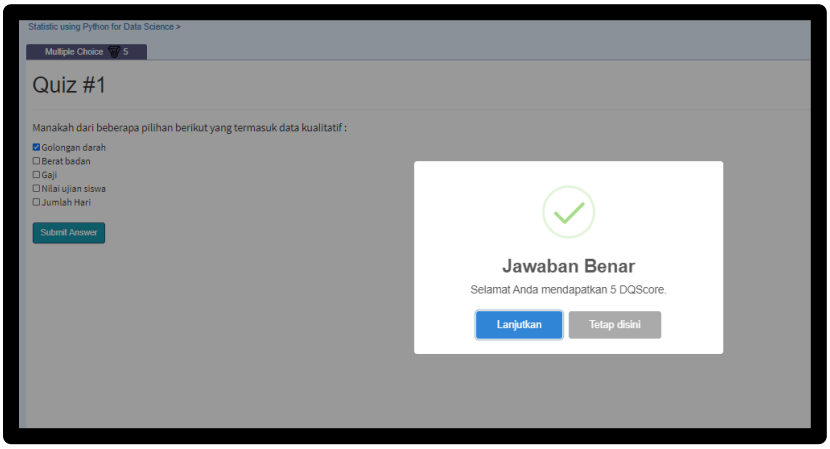
- Skala Rasio

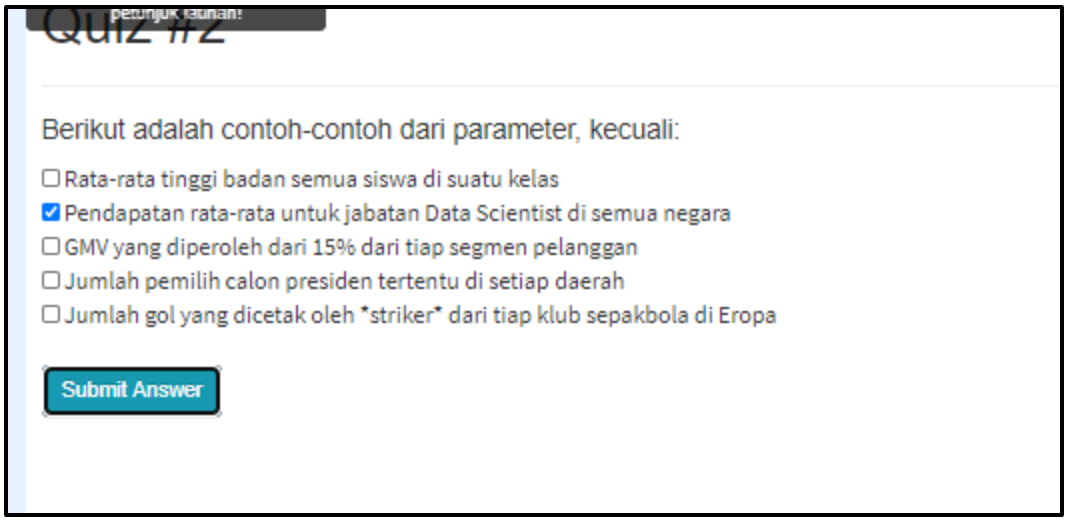
Skala rasio memiliki kemiripan dengan skala interval, perbedaannya terletak pada **nilai 0** pada skala rasio. Berbeda dengan skala interval yang tidak memiliki nilai 0 yang tidak pasti.

Package Statistika di Python

Untuk melakukan perhitungan statistika, kita dapat menggunakan beberapa *package* atau *library* berikut yang tersedia di Python. Diantaranya adalah:

- numpy: digunakan untuk melakukan analisa data numerik dan perhitungan berbasis vektor atau matriks
- pandas: digunakan untuk melakukan pengolahan data tabular
- matplotlib: digunakan untuk melakukan ploting atau penggambaran grafik, dapat digunakan sebagai alat bantu dalam analisa data
- statsmodels: digunakan untuk melakukan uji hipotesa, eksplorasi data maupun pemodelan statistika
- scipy: digunakan untuk melakukan uji statistika, juga dapat digunakan untuk melakukan pemodelan statistika



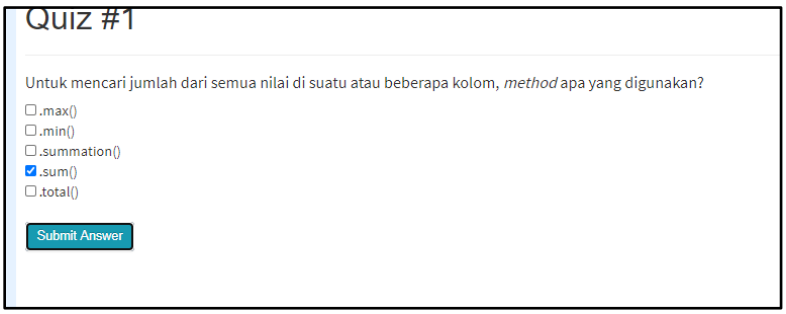


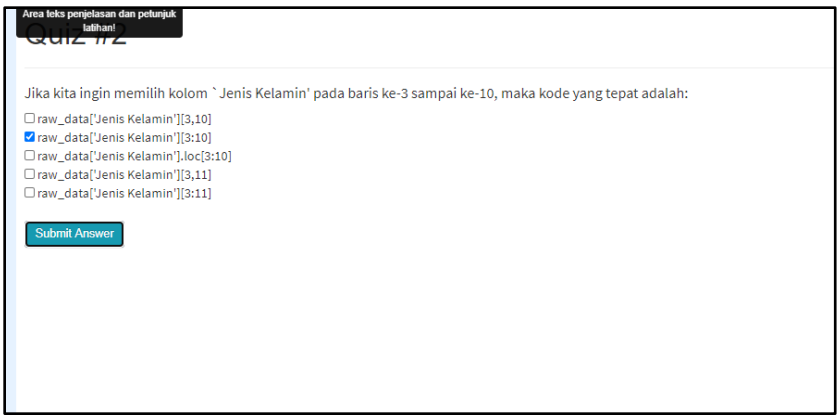
Pengenalan Numpy dan Pandas

Kedua *library* ini, numpy dan pandas, adalah *library* yang umum digunakan untuk melakukan pengolahan data, mulai dari membaca data dalam format tertentu (csv, xlsx, xls, tsv) atau dari sumber tertentu (RDBMS, No-SQL), melakukan inspeksi (mengecek data yang hilang, inkonsistensi pada data) dan pengolahan data (transformasi nilai, *encoding*, normalisasi) sampai visualisasi data untuk membuat laporan atau mempersiapkan data untuk membuat model.

Numpy adalah *library* yang biasanya digunakan untuk manipulasi *array* atau vektor. Perhitungan yang melibatkan operasi pada objek berbentuk matriks, vektor atau bahkan multidimensi vektor (misalnya data gambar dengan skema warna RGB) dapat kita lakukan dengan menggunakan numpy.

Pandas adalah *library* yang biasanya digunakan untuk analisa data atau biasa disebut sebagai *data wrangling*. Biasanya data yang diolah oleh pandas berbentuk tabular atau tabel layaknya *spreadsheet* di Excel. Pandas menggunakan numpy sebagai *back-end* sehingga beberapa fungsi atau *method* dari numpy dapat digunakan di objek pandas.





Ukuran Pusat (Measures of Central Tendency)

Ukuran pusat (*measures of central tendency*) adalah statistika deskriptif yang dapat digunakan untuk membantu kita mengidentifikasi kasus-kasus tipikal di dalam sebuah sampel atau populasi Terdapat beberapa jenis ukuran pusat yang dapat digunakan untuk menganalisa data, diantaranya:

- Mean atau rerata
- Median
- Modus

Rata-rata (Mean)

Rata-rata atau *mean* adalah salah satu ukuran pusat yang nilainya diperoleh dengan cara menjumlahkan semua nilai titik data yang ada lalu dibagi oleh jumlah data. Secara matematis hal ini dapat dirumuskan sebagai berikut:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Dimana:

- $\bar{x}$ adalah rerata
- x1,x2,...,xn adalah titik data
- n adalah jumlah data

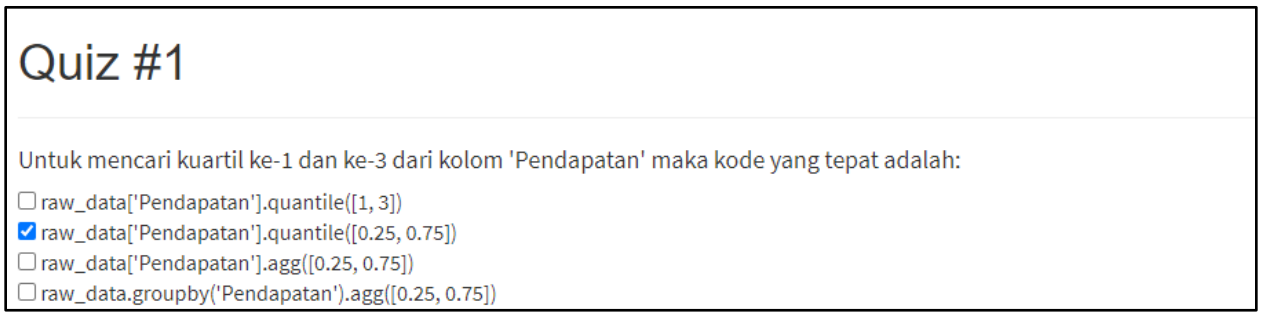
Median

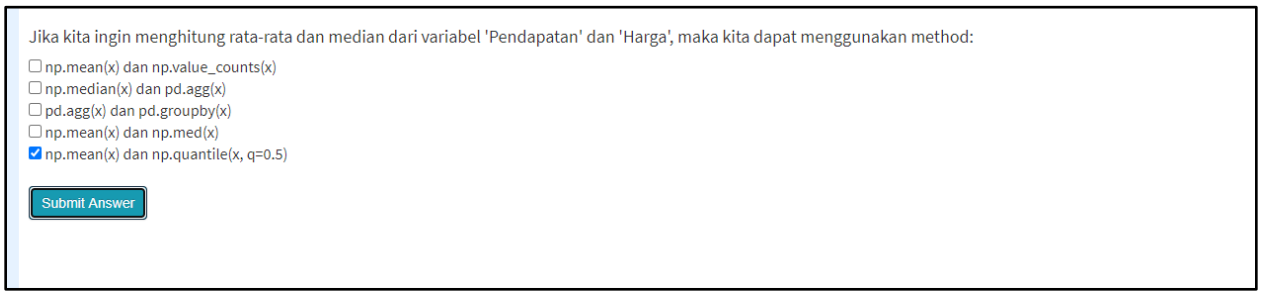
Median adalah salah satu ukuran pusat yang nilainya terletak di tengah titik data. Sebagai gambaran, jika kita memiliki titik data bernilai 1, 2, 3, 4, 4, 5, 6 maka median dari sekumpulan

titik data tersebut adalah 4. Namun jika kita memiliki titik data bernilai 1, 2, 3, 3 maka media dari sekumpulan titik data tersebut adalah:

$$\text{Median} = \frac{2+3}{2} = 2.5$$

Kita dapat menemukan nilai median dengan menggunakan method .median() pada numpy maupun pandas





Ukuran Sebaran (Measures of Dispersion)

Ukuran sebaran (*measure of dispersion*) adalah statistika deskriptif yang digunakan untuk membantu kita memahami sebaran titik data di dalam sebuah sampel ataupun populasi.

Terdapat beberapa ukuran sebaran yang biasanya digunakan tergantung pada jenis atau tipe datanya, yaitu:

- Tipe Data Nominal dan Ordinal
    - Proporsi Kategori (*Categorical Proportion*)
    - Persen Proporsi (*Percent Proportion*)
    - Rasio Variasi (*Variation Ratio*)
- Tipe Data Interval dan Rasio
    - Rentang (*Range*)
    - Variansi (*Variance*)

- Deviasi Baku (*Standard Deviation*)

Proporsi Kategori

Proporsi kategori adalah ukuran sebaran yang paling sederhana dari ukuran sebaran pada data nomisal dan ordinal. Secara matematis dapat dirumuskan sebagai:

$$\text{Proporsi} = \frac{N_{\text{kategori ke-i}}}{N_{\text{total}}}$$

Ukuran Sebaran pada Data Interval dan Rasio

Rentang (*range*)

Rentang adalah jarak antara nilai maksimum dengan nilai minimum. Semakin besar jarang antara nilai maksimum dan minimum semakin besar pula sebaran datanya. Secara matematis dapat dituliskan sebagai berikut:

Range=max(X)−min(X)

Variansi

Variansi adalah ukuran sebaran pusat yang diperoleh dengan cara menghitung jarak antara tiap titik data pada sampel atau populasi dengan nilai mean. Secara matematis variansi dirumuskan sebagai berikut:

Variansi Populasi

$$s^2 = \frac{\Sigma_{i=1}^{N}(X_i - \bar{X})^2}{N} = \frac{(X_1 - \bar{X})^2 + \ldots + (X_N - \bar{X})^2}{N}$$

Dimana N merupakan jumlah data dan $X_1, \ldots, X_N$ adalah titik data pada sampel atau populasi.

Variansi Sampel

$$s^2 = \frac{\Sigma_{i=1}^{N}(X_i - \bar{X})^2}{N-1} = \frac{(X_1 - \bar{X})^2 + \ldots + (X_N - \bar{X})^2}{N-1}$$

Untuk menghitung variansi kita dapat menggunakan method .var() dari numpy maupun pandas.

Note: Perhatikan bahwa nilai variansi keduanya berbeda. Hal ini karena secara *default* pandas menggunakan variansi sampel sedangkan numpy menggunakan variansi populasi. Untuk memperoleh hasil yang sama kita dapat menggunakan parameter ddof=0 pada method .var() untuk memperoleh variansi populasi.

Deviasi Baku (Standard Deviation)

Deviasi baku adalah ukuran sebaran pusat yang diperoleh dengan cara menarik akar kuadrat dari hasil perhitungan variansi. Hal ini dilakukan karena nilai variansi umumnya memiliki nilai yang lebih besar daripada nilai aslinya sebagai efek dari pengkuadratan dan ini menjadikan variansi sulit untuk diinterpretasikan. Secara matematis deviasi baku dapat dirumuskan sebagi berikut:

Deviasi Baku Populasi

$$s = \sqrt{\frac{\Sigma_{i=1}^{N}(X_i - \bar{X})^2}{N}} = \sqrt{\frac{(X_1 - \bar{X})^2 + \ldots + (X_N - \bar{X})^2}{N}}$$

Deviasi Baku Sampel

$$s = \sqrt{\frac{\Sigma_{i=1}^{N}(X_i - \bar{X})^2}{N-1}} = \sqrt{\frac{(X_1 - \bar{X})^2 + \ldots + (X_N - \bar{X})^2}{N-1}}$$

Kita dapat menghitung deviasi baku menggunakan method .std dari numpy maupun pandas.

Quiz

Untuk menghitung variansi dan standar deviasi sampel pada kolom 'Pendapatan' dan 'Harga', maka kode yang dapat digunakan adalah:

- [ ] `raw_data[['Pendapatan', 'Harga']].agg([np.var, np.std])`
- [ ] `np.var(raw_data[['Pendapatan', 'Harga']], ddof=1)`
- [ ] `raw_data[['Pendapatan', 'Harga']].agg([np.var, np.std])`
- [ ] `raw_data[['Pendapatan', 'Harga']]([np.var, np.std], ddof=1)`
- [x] `raw_data[['Pendapatan', 'Harga']].agg([np.var, np.std], ddof=1)`

Submit Answer

Korelasi

Korelasi adalah salah satu metode statistika yang dapat digunakan untuk mengukur seberapa besar hubungan antara satu variabel dengan variabel lainnya. Sebagai contoh, misalnya mencari hubungan antara tinggi badan dengan berat badan, mencari hubungan antara gender dengan penghasilan dan masih banyak aplikasi penggunaan korelasi.

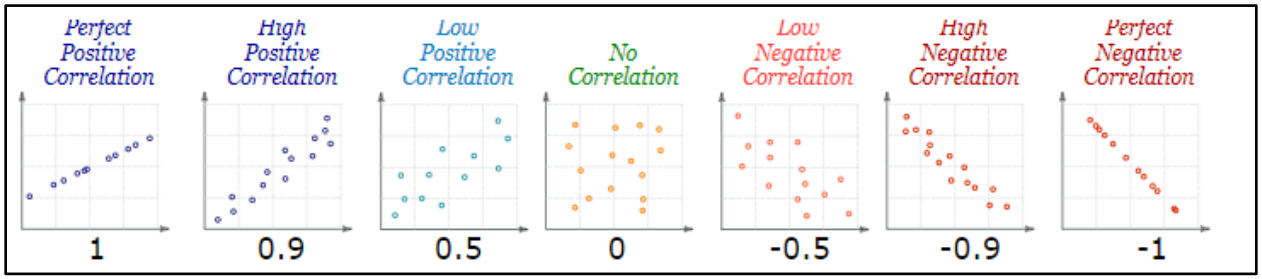
Terdapat beberapa metode yang dapat digunakan untuk menghitung korelasi antara sepasang variabel tergantung tipe dari sepasang variabel tersebut. Diantaranya adalah:

1. Korelasi Pearson
2. Korelasi Spearman
3. Korelasi Kendall

Korelasi Pearson

Korelasi Pearson atau sering juga disebut sebagai *Pearson's product moment correlation* adalah pengukuran korelasi parametrik yang menghasilkan koefisien korelasi. Koefisien korelasi ini dapat digunakan untuk mengukur kekuatan hubungan atau asosiasi linier antara dua variabel. Artinya jika hubungan kedua variabel tidak linier maka koefisien korelasi Pearson tidak dapat digunakan untuk mengukur kekuatan hubungan antara kedua variabel.

Selain itu nilai dari koefisien pearson dapat digunakan untuk mengukur arah dari hubungan tersebut: **positif atau negatif**. Hubungan antar variabel dikatakan **positif jika nilai salah satu variabel naik maka nilai variabel lainnya juga naik**. Sebaliknya, hubungan antar variabel dikatakan **negatif jika nilai salah satu variabel naik maka nilai variabelnya turun**. Gambar berikut dapat menjelaskan maksud dari kekuatan dan arah dari korelasi antar kedua variabel.



Sumber Gambar: https://www.mathsisfun.com

Beberapa asumsi yang harus dipenuhi untuk menggunakan korelasi Pearson diantaranya adalah:

1. Nilai berskala interval/rasio
2. Terdapat hubungan linier antara kedua variabel, yaitu kita dapat menggambarkan hubungan kedua variabel sebagai garis lurus.
3. Kedua variabel berdistribusi normal
4. Homoskedastis, atau data berdistribusi seragam dalam garis regresi

Secara matematis, korelasi pearson dapat dirumuskan sebagai berikut:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

Korelasi Spearman

Korelasi Spearman atau sering juga disebut sebagai *Spearman's rank correlation* adalah **pengukuran korelasi non-parametrik**. Artinya kita mencoba mengukur

hubungan antara kedua variabel tanpa menghiraukan asumsi seperti distribusi dari kedua variabel dan asumsi lainnya. Secara kriteria memiliki kemiripan dengan korelasi Pearson walau korelasi Spearman bisa digunakan untuk data bertipe ordinal. Perbedaannya hanya terletak pada pengubahan data dalam bentuk ranking sebelum menghitung nilai korelasinya.

Secara matematis, korelasi Spearman dapat dihitung menggunakan rumus berikut:

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

Dimana:

- ρ adalah nilai korelasi Spearman
- di adalah nilai beda antara kedua variabel
- n adalah jumlah data

Korelasi Kendall

Korelasi Kendall atau sering juga disebut juga sebagai *Kendall's tank correlation* atau korelasi Tau (τ) adalah pengukuran korelasi non-parametrik.

Secara matematis, korelasi Kendall dapat dihitung menggunakan rumus berikut:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

Dimana:

- n_c adalah jumlah pasangan konkordan
- n_d adalah jumlah pasangan diskordan
- n adalah jumlah data

Menghitung Korelasi

Untuk menghitung korelasi antara kedua variabel, kita dapat menggunakan method `.corr()` dari pandas sebagaimana contoh berikut:

```
[ ]  # menghitung korelasi dari setiap pasang variabel pada raw_data
     raw_data.corr()
```

	ID Pelanggan	Jenis Kelamin	Pendapatan	Harga	Jumlah	Total	Tingkat Kepuasan
ID Pelanggan	1.000000	0.151375	0.110958	-0.028707	0.011289	-0.039968	-0.245717
Jenis Kelamin	0.151375	1.000000	0.192849	0.457555	-0.104168	0.238051	-0.088339
Pendapatan	0.110958	0.192849	1.000000	0.322443	0.399825	0.592044	-0.312663
Harga	-0.028707	0.457555	0.322443	1.000000	-0.138883	0.645455	-0.380798
Jumlah	0.011289	-0.104168	0.399825	-0.138883	1.000000	0.636097	0.017568
Total	-0.039968	0.238051	0.592044	0.645455	0.636097	1.000000	-0.268345
Tingkat Kepuasan	-0.245717	-0.088339	-0.312663	-0.380798	0.017568	-0.268345	1.000000

Perhatikan bahwa hanya variabel bertipe numerik saja yang dihitung nilai korelasinya. Selain itu secara *default method* .corr() memiliki parameter method bernilai pearson sehingga nilai korelasi di atas adalah korelasi Pearson. Untuk menggantinya menjadi korelasi Kendall atau korelasi Spearman kita cukup menambahkan parameter method='kendall' atau method='spearman' pada *method* .corr() sebagaimana contoh berikut:

```
[ ]  # mencari korelasi 'kendall' untuk tiap pasang variabel
     raw_data.corr(method='kendall')
```

	ID Pelanggan	Jenis Kelamin	Pendapatan	Harga	Jumlah	Total	Tingkat Kepuasan
ID Pelanggan	1.000000	0.126650	-0.054998	-0.005753	0.024016	-0.065998	-0.183817
Jenis Kelamin	0.126650	1.000000	0.190245	0.415339	-0.090299	0.190245	-0.085796
Pendapatan	-0.054998	0.190245	1.000000	0.523053	0.501925	0.988506	-0.165588
Harga	-0.005753	0.415339	0.523053	1.000000	-0.098450	0.535078	-0.325659
Jumlah	0.024016	-0.090299	0.501925	-0.098450	1.000000	0.501925	-0.028923
Total	-0.065998	0.190245	0.988506	0.535078	0.501925	1.000000	-0.165588
Tingkat Kepuasan	-0.183817	-0.085796	-0.165588	-0.325659	-0.028923	-0.165588	1.000000

```
[ ]  # mencari korelasi 'spearman' untuk tiap pasang variabel
     raw_data.corr(method='spearman')
```

	ID Pelanggan	Jenis Kelamin	Pendapatan	Harga	Jumlah	Total	Tingkat Kepuasan
ID Pelanggan	1.000000	0.151375	-0.063711	-0.039149	0.046356	-0.069779	-0.238890
Jenis Kelamin	0.151375	1.000000	0.219508	0.463635	-0.098864	0.219508	-0.090784
Pendapatan	-0.063711	0.219508	1.000000	0.640000	0.607110	0.998470	-0.192463
Harga	-0.039149	0.463635	0.640000	1.000000	-0.130749	0.646194	-0.378933
Jumlah	0.046356	-0.098864	0.607110	-0.130749	1.000000	0.607110	-0.023874
Total	-0.069779	0.219508	0.998470	0.646194	0.607110	1.000000	-0.192463
Tingkat Kepuasan	-0.238890	-0.090784	-0.192463	-0.378933	-0.023874	-0.192463	1.000000

Interpretasi Nilai Korelasi dan Kaitannya dengan Kausalitas

Perhatikan nilai korelasi Pearson dari raw_data berikut:

```
[ ] # menghitung korelasi dari setiap pasang variabel pada raw_data
    raw_data.corr()
```

	ID Pelanggan	Jenis Kelamin	Pendapatan	Harga	Jumlah	Total	Tingkat Kepuasan
ID Pelanggan	1.000000	0.151375	0.110958	-0.028707	0.011289	-0.039968	-0.245717
Jenis Kelamin	0.151375	1.000000	0.192849	0.457555	-0.104168	0.238051	-0.088339
Pendapatan	0.110958	0.192849	1.000000	0.322443	0.399825	0.592044	-0.312663
Harga	-0.028707	0.457555	0.322443	1.000000	-0.138883	0.645455	-0.380798
Jumlah	0.011289	-0.104168	0.399825	-0.138883	1.000000	0.636097	0.017568
Total	-0.039968	0.238051	0.592044	0.645455	0.636097	1.000000	-0.268345
Tingkat Kepuasan	-0.245717	-0.088339	-0.312663	-0.380798	0.017568	-0.268345	1.000000

Perhatikan bahwa beberapa pasang variabel memiliki nilai korelasi yang positif maupun negatif. Selain itu terdapat nilai yang sangat kecil hingga mendekati nol, namun ada juga yang cukup besar diatas 0.5, misalnya korelasi antara 'Jumlah' dan 'Pendapatan'. Besar kecilnya suatu nilai korelasi dari sepasang variabel menandakan seberapa kuat **hubungan linier** antara kedua variabel tersebut. Sebagai acuan untuk mengukur seberapa kuat korelasi sepasang variabel, kita dapat menggunakan ukuran berikut:

- $0 < |r| < 0.49$: Hubungan lemah
- $0.50 < |r| < 0.79$: Hubungan sedang
- $0.80 < |r| < 1$: Hubungan kuat

Berdasarkan kriteria di atas, kita dapat menilai bahwa hubungan antara variabel 'Total' dan 'Jumlah' hubungannya sedang dan positif (r = 0.636097) sedangkan hubungan antara 'Pendapatan' dan 'Tingkat Kepuasan' hubungannya lemah dan negatif (r = -0.088339).

Namun, perlu diperhatikan bahwa walaupun kita dapat menilai kuat hubungan antara kedua variabel, namun kita tidak bisa menentukan arah dari hubugan tersebut. Sebagai contoh, walaupun antara variabel 'Total' dan 'Jumlah' memiliki hubungan yang sedang, tidak berarti kita bisa mengetahui apakah kenaikan nilai 'Total' berefek positif terhadap 'Jumlah' atau sebaliknya.

Penutup

Selamat! Kamu sudah menyelesaikan materi Statistic using Python for Data Science dengan baik. Pada materi ini, kamu sudah dapat memahami :

- Pengenalan mengenai stastistik.
- Cara membaca data dalam format CSV.

- Estimasi karakteristik (Mean, Median, Modus).
- Skala pengukuran data.
- Ukuran sebaran data.
- Perbedaan statistik deskriptif dan statistik inferensia.
- Mengenal jenis-jenis korelasi.