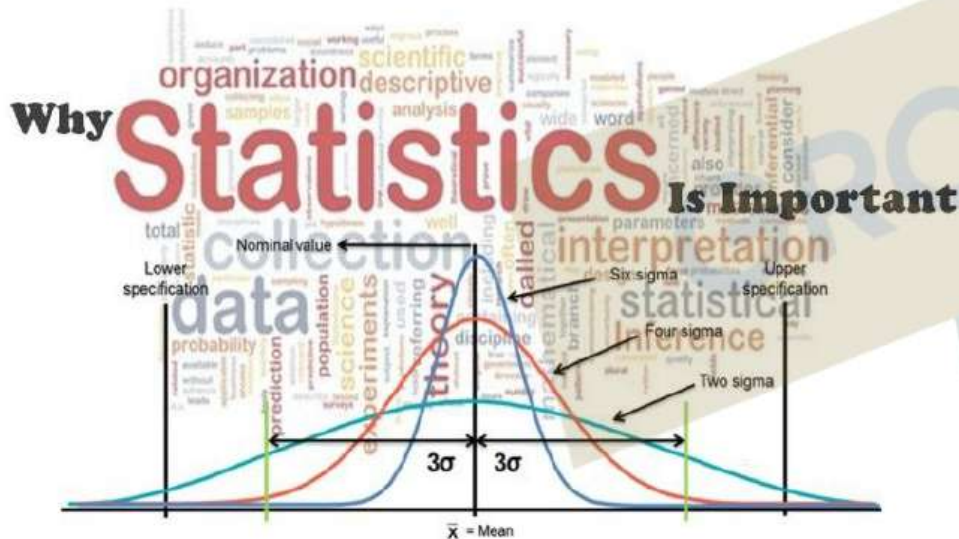




# INTRODUCTION TO STATISTICS

# APA ITU STATISTIK



**Statistik** sebagai keilmuan adalah bidang ilmu yang mempelajari tentang :

- Pengumpulan data
- Analisis data
- Penyajian data

Sehingga kita dapat melakukan pengambilan kesimpulan atas suatu pola atau trend tertentu.

Di sisi lain merupakan sebagai angka, statistik adalah :

- **angka-angka** yang dikumpulkan
- dikelompokkan
- dan ditabulasi

sehingga didapatkan informasi berkaitan dengan masalah tertentu

## Transcription Pengantar Statistic dan Statistics Descriptive

Time	Transcription
0:03 - 1:04	<p>Halo semuanya kali ini kita akan berkenalan dengan “Apa itu statistik?”. Statistik sebagai sebuah bidang keilmuan adalah bidang yang mempelajari mengenai pengumpulan data, analisis data penyajian data sehingga kita dapat melakukan pengambilan kesimpulan atas suatu pola atau tren tertentu, misalnya “Apakah benar tanaman herbal bisa menyembuhkan penyakit?”, statistik akan mempelajari bagaimana caranya mengumpulkan data eksperimen yang tidak bias dan membantu menghitung “Apakah obat yang digagas efektif ataupun tidak?”. Sedangkan statistik sebagai angka itu adalah angka-angka yang dikumpulkan, dikelompokkan ataupun ditabulasi sehingga dapat memberikan informasi. Jadi statistik itu bermakna ganda baik sebagai keilmuan ataupun sebagai angka perhitungan.</p>

# TIPE DATA

## Numerik

*contoh numerik diskrit :*

- banyaknya saudara
- banyaknya mobil yang dimiliki

*contoh numerik kontinu :*

- suhu udara
- tinggi badan

## Categorical

*contoh :*

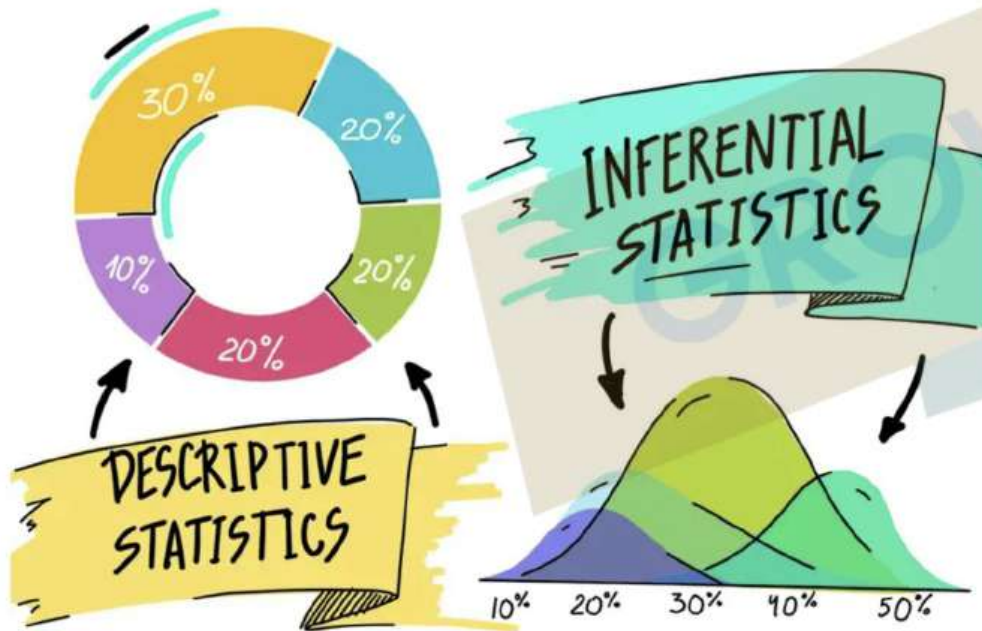
- warna baju
- nama kota
- makanan favorit
- nomor telepon



## Transcription Pengantar Statistic dan Statistics Descriptive

Time	Transcription
1:05–2:37	<p>Untuk mempelajari statistik kita perlu tahu dulu apa saja tipe-tipe data yang ada. Secara umum data dapat dibagi menjadi dua tipe yaitu data numerik dan data categorical. Data numerik atau biasanya disebut sebagai data kuantitatif adalah data dengan sifat angka yang mana kita dapat melakukan operasi aritmatika. Tipe data numerik dapat dipecah lagi menjadi diskrit dan kontinu, yang mana diskrit berarti angkanya adalah bilangan bulat seperti banyaknya saudara, banyaknya mobil yang dimiliki. Sedangkan kontinu adalah angka yang bisa mengandung koma, misalkan suhu udara hari ini adalah 37,85 derajat atau tinggi badan 170,2 senti. Sedangkan data kategorikal atau yang biasa disebut sebagai data kualitatif adalah data yang bukan numerik dan kita tidak bisa melakukan operasi aritmatika, contohnya adalah warna baju di sini kita tidak bisa melakukan perkalian antara warna merah dengan warna hijau contoh lainnya adalah nama kota, makanan favorit dan nomor telepon, di sini nomor telepon meskipun angka tetap masuk ke data kategorikal karena kita tidak bisa melakukan operasi aritmatika dengan nomor HP, penjumlahan dengan nomor HP tidak menghasilkan angka yang bermakna.</p>

# INTRODUCTION TO STATISTICS



Teknik untuk menyampaikan rangkuman mengenai data menggunakan angka, grafik maupun tabel disebut sebagai **Statistik deskriptif**.

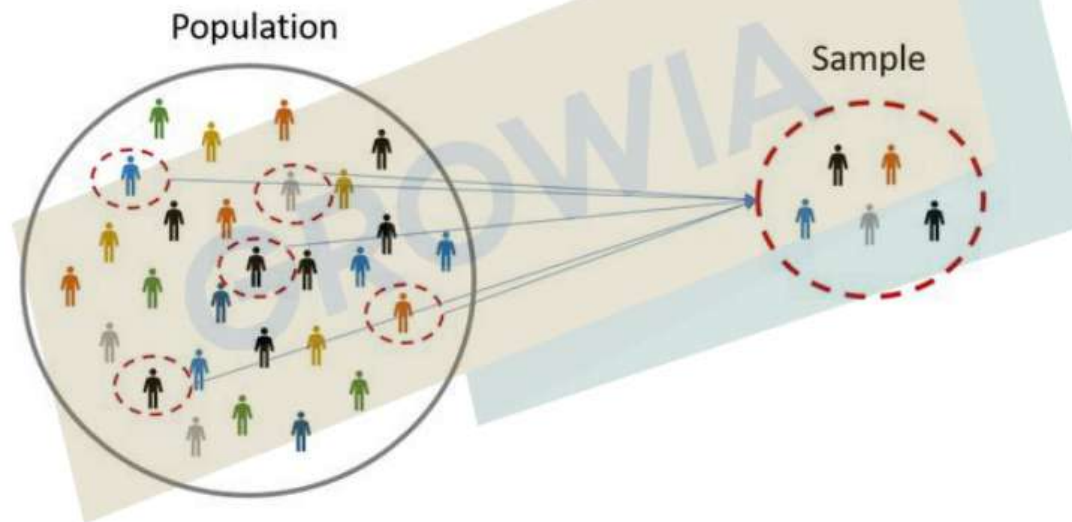
Teknik untuk menarik kesimpulan tentang populasi dari sampel data disebut sebagai **Statistik inferensial**

## Transcription Pengantar Statistic dan Statistics Descriptive

Time	Transcription
2:39–3:15	Setelah mengetahui tipe data selanjutnya saya akan menjelaskan dua jenis analisis statistik yang umumnya digunakan. Yang pertama adalah deskriptif statistik dan yang kedua adalah statistik inferensial. Yang dimaksud dengan statistik deskriptif adalah teknik untuk menyampaikan rangkuman mengenai data menggunakan angka, grafik maupun tabel. Sedangkan inferensi fokus kepada penarikan kesimpulan tentang populasi dari sampel data yang ada.



# POPULATION AND SAMPLES



**Populasi** : Keseluruhan dari objek yang ingin diamati

**Sampel** : Bagian dari populasi



## Transcription Pengantar Statistic dan Statistics Descriptive

Time	Transcription
3:16–3:47	<p>Apa sih yang dimaksud dengan sampel dan populasi? Populasi adalah keseluruhan dari objek yang ingin diteliti sementara sampel adalah bagian dari populasi. Misalkan ketika teman-teman ingin menganalisis data mungkin saja teman-teman tidak bisa mengambil data dari seluruh subjek melainkan hanya beberapa saja, contohnya ketika melakukan pengujian obat, kita hanya bisa memilih berapa puluh pasien yang bisa menjadi subjek percobaan obat yang baru dari mungkin ribuan atau jutaan penderita penyakit tertentu.</p>

# STATISTIK UTAMA DALAM STATISTICS DESCRIPTIVE

## Ukuran pemusatan data:

- mean
- median
- mode

## Ukuran penyebaran data:

- range
- variansi dan standard deviasi
- inter quartile range

## Transcription Pengantar Statistic dan Statistics Descriptive

Time	Transcription
03.49 - 04.17	Sekarang kita akan membahas mengenai statistika deskriptif. Dimulai dari aspek perhitungan, dalam statistika deskriptif umumnya ada dua angka statistik yang sering digunakan untuk mendeskriptifkan data, ada yang disebut sebagai ukuran pemusatan data seperti mean, median, mode atau modus dan ada yang disebut sebagai ukuran penyebaran data seperti range, variasi dan interquartile range.



# Ukuran pemusatan data

*Merupakan ukuran nilai yang menyatakan pusat dari data*

## Mean

sering dinyatakan sebagai rata-rata

## Median

nilai yang tengah dari data

## Modus

nilai yang paling sering muncul

## Transcription Pengantar Statistic dan Statistics Descriptive

Time	Transcription
04.18–05:10	<p>Apa itu ukuran pemusatan data? Pada dasarnya ukuran pemusatan data adalah nilai yang menyatakan pusat dari data. Pusat berarti akan ada banyak nilai data yang tercatat di sekitar pusat. Angka pemusatan pertama yang akan kita bahas adalah mean atau bisa disebut sebagai rata-rata. Perhitungan mean cukup sederhana yakni kita menjumlahkan seluruh nilai dan membaginya dengan banyaknya sampel pengamat. Kedua median, median adalah nilai tengah dari data. Nilai median lebih sering digunakan apabila dalam data kita terdapat nilai pencilan ekstrim. Ketiga modus, modus adalah nilai yang paling sering muncul. Nilai modus lebih cocok digunakan apabila data kita bersifat numerical discrete</p>

# Ukuran pemusatan data

Murid	Tono	Joko	Andi	Desi	Budi
Nilai Ujian	80	80	100	20	30

Perhitungan mean :  
 $(80 + 80 + 100 + 20 + 30) / 5 = 62$

Perhitungan median  
 nilai ujian diurutkan dari terendah  
 20, 30, **80**, 80, 100

Perhitungan modus  
 Hitung berapa banyak kemunculan

Nilai	Frekuensi
80	2
100	1
30	1
20	1

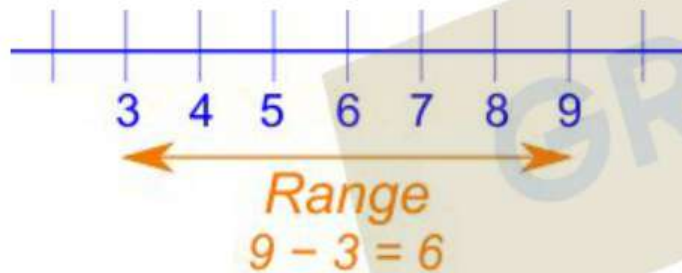


## Transcription Pengantar Statistic dan Statistics Descriptive

Time	Transcription
05:11–06:20	<p>Misalkan kita memiliki data ujian dari 5 orang murid, untuk menghitung mean atau rata-rata dari data kita jumlahkan semua nilai anak di sini lalu dibagi dengan banyaknya murid yakni 5 dan hasilnya adalah 62. Jadi <math>80 + 80</math> hingga terakhir <math>+ 30</math> dibagi dengan 1 2 3 4 5 murid hasilnya 62. Sementara untuk menghitung median, pertama datanya kita urutkan terlebih dahulu dari nilai terendah di sebelah kiri hingga nilai terbesar di sebelah kanan karena kita memiliki 5 murid maka nilai tengahnya adalah posisi yang di tengah yaitu tiga, tepat ditengah disini. Berapa nilai ujian pada urutan ke-3? Nilai ujiannya adalah 80 maka median dari data ini adalah 80. Sedangkan untuk perhitungan modus kita hitung dulu seluruh nilai yang ada berapa kali nilainya muncul. Di sini nilai 80 muncul sebanyak 2 kali sedangkan nilai lainnya hanya muncul 1 kali oleh karena itu modus dari data ini adalah 80.</p>

# UKURAN PENYEBARAN DATA

## - Range



*Range* adalah ukuran variabilitas yang paling sederhana dan sangat mudah dihitung.

**Rumus :** *maksimum - minimum*

## Transcription Pengantar Statistic dan Statistics Descriptive

Time	Transcription
06:21-06:54	Berikutnya mari kita bahas mengenai ukuran penyebaran data dimulai dari range. Range merupakan ukuran penyebaran yang paling sederhana dan sangat mudah dihitung. Kita hanya perlu mengambil nilai maksimum dan kita kurangi dengan nilai minimum akan tetapi metode ini memiliki suatu kelemahan karena kita hanya melihat maksimum dan minimum nilai anomali yang terlalu besar atau terlalu kecil akan mengganggu pemahaman kita mengenai penyebaran data yang kita miliki.



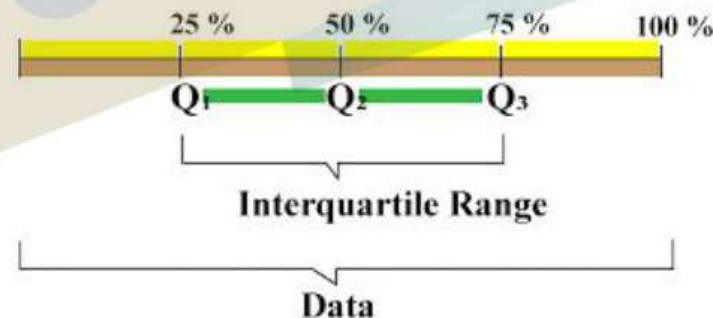
# UKURAN PENYEBARAN DATA

## - Interquartile Range (IQR)

Pada perhitungan median, kita menyusun data untuk mencari nilai yang tepat di tengah.

**Median** adalah salah satu angka khusus dari quartile (Median =  $Q_2$ ).

Jika median membagi data menjadi 2 bagian, maka Quartile 1 adalah nilai tengah dari setengah data di sisi kiri quartile pada data yang sudah diurutkan.



$$\text{IQR} = \text{Quartile3} - \text{Quartil 1}$$

## Transcription Pengantar Statistic dan Statistics Descriptive


Time	Transcription
06:21-07:55	<p>Berikutnya ada metode yang lebih tidak sensitif dengan pencilan ekstrim dari pada range yaitu interquartile range atau yang biasa disebut sebagai IQR. Jika range adalah selisih dari maksimum dan minimum, maka IQR adalah selisih dari quartile 3 dengan quartile 1. Apa itu quartile? Sebelumnya kita sudah menghitung median yang mana median itu sebenarnya adalah kuartil 2. Bayangkan jika teman-teman memiliki angka 1 sampai 100 secara berurut, maka median itu adalah angka yang ke-50, sedangkan quartile 1 adalah angka yang ke-25 dan quartile 3 adalah angka yang ke-75. Jadi IQR nya adalah angka yang ke-75 dikurang yang ke-25. Untuk menghitung kuartil kita harus mengurutkan datanya terlebih dahulu dari yang kecil ke yang besar.</p>

# UKURAN PENYEBARAN DATA

## - Variance

Variansi secara sederhana menyatakan rata-rata jarak kuadrat observasi di data dengan rataannya.

Standard deviasi adalah akar dari variansi dan lebih mudah untuk di interpretasi.

 **The Sample Variance Formula**

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

$x_i$  are the individual data values  
 $\bar{x}$  is the sample mean  
 $n$  is the number of data values

© Maths at Home [www.mathsat-home.com](http://www.mathsat-home.com)



## Transcription Pengantar Statistic dan Statistics Descriptive

Time	Transcription
07:56–8:25	Berikutnya Variance atau variansi, secara sederhana variansi yang menyatakan rata-rata jarak kuadrat atas setiap data observasi dengan rataannya. Ini adalah formula dari variansi, $X_i$ adalah pengamatan ke $i$ $\bar{X}$ adalah rata-rata selisihnya kita kuadratkan, simbol Sigma di sini menunjukkan penjumlahan dan $n$ adalah banyaknya sampel pengamatan.

# Ukuran persebaran data

Murid	Tono	Joko	Andi	Desi	Budi
Nilai Ujian	80	80	100	20	30

## Range

$$100 - 20 = 80$$

## IQR

20, 30, 80, 80, 100

$$Q1 = (20 + 30)/2 = 25$$

$$Q3 = (80 + 100)/2 = 90$$

$$IQR = 90 - 25 = 65$$

$$\text{Variansi} = \frac{(80 - 62)^2 + (80 - 62)^2 + (100 - 62)^2 + (20 - 62)^2 + (30 - 62)^2}{(5-1)} = 1220$$

$$\text{Standard deviasi} = \text{akar}(1220) = 34.9...$$

## Transcription Pengantar Statistic dan Statistics Descriptive

Time	Transcription
8:26:9:25	<p>Biar lebih jelas ayo kita lihat contoh berikut ini, disini kita kembali menggunakan data ujian murid. Pertanyaan pertama berapa range dari data ini? Maksimumnya adalah 100 dan minimumnya adalah 20 maka kita katakan range dari data ini adalah 80. Kemudian berapa IQR dari data ini? Jadi datanya kita urutkan kemudian kita hitung Q1 dan Q3 nya berapa? Di sini, di sebelah kiri median ini mediannya 80 di sebelah kirinya hanya terdapat 2 angka, kalau dua angka seperti ini nilai tengahnya apa? Nilai tengahnya adalah <math>20 + 30</math> dibagi dengan 2 yaitu 25. Sedangkan Q3 juga serupa, kita ambil nilai tengah diantara kedua nilai ini, maka quartile 3 dari data ini adalah <math>80 + 100</math> dibagi dua yaitu 90, maka interquartile range dari data ini adalah 65.</p>

## Transcription Pengantar Statistic dan Statistics Descriptive

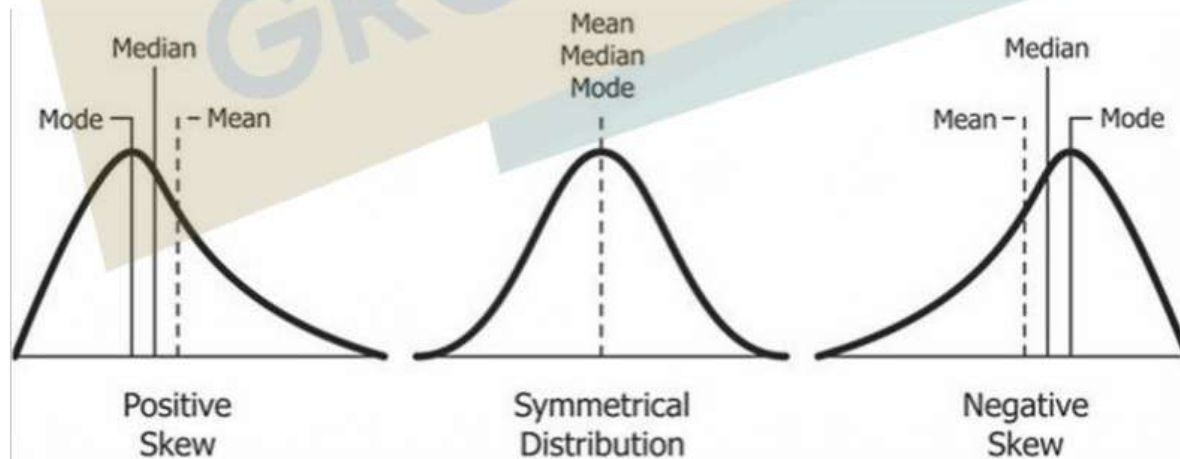
Time	Transcription
09:26–10:23	Selanjutnya variansi, variansi dihitung sebagai berikut, pada bagian sebelumnya kita tahu bahwa rata-rata dari nilai ujian murid adalah 62, maka perhitungan variansi dilakukan sebagai berikut nilai ujian yang pertama 80 dikurang dengan rata-ratanya kemudian kita kuadratkan ditambah sekali lagi 80 itu $80 - 62$ kita kuadratkan ditambah 100 dikurang 62 dibulatkan sampai nilai yang terakhir yaitu budi (30) kita bagi dengan 5 dikurang 1 atau dibagi dengan 4 dan didapat variansinya 1220 akan tetapi nilai variansi ini agak susah untuk diinterpretasikan maka nilai yang lebih umum digunakan atau lebih mudah untuk diinterpretasikan adalah standar deviasi. Standar deviasi adalah akar dari variansi 34,9 untuk kasus data ini jadi data nya sekitar 30-an.



## OTHER MEASUREMENT

### - Skewness

Skewness adalah ukuran yang menyatakan kemencengan/ketidak simetrisan dari suatu data.



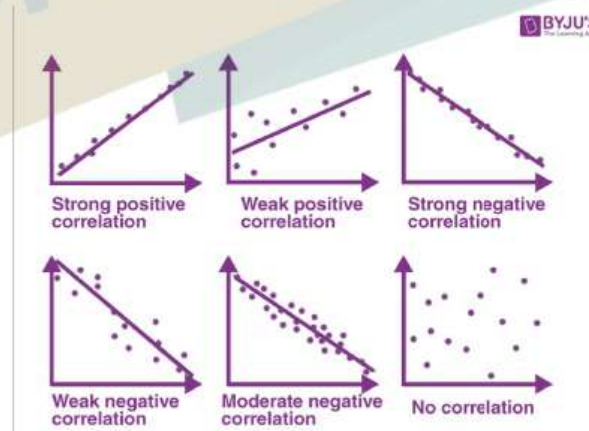
## Transcription Pengantar Statistic dan Statistics Descriptive

Time	Transcription
10:24–11:57	<p>Selain ukuran pusat data dan ukuran persebaran data, ada juga statistik-statistik lain yang digunakan untuk mendeskripsikan data, misalnya adalah skewnes. Skewnes adalah ukuran yang menyatakan kemencengan atau ketidak simetrisan dari suatu data. Perhatikan tiga grafik berikut, grafik ini ini adalah grafik distribusi, ketika kurvanya semakin tinggi artinya semakin banyak nilai yang terkumpul di area ini, sebelah kanan berarti nilai yang lebih tinggi sementara sebelah kiri berarti nilai yang rendah. Pada grafik di kiri ini, ini kita sebut sebagai positive skew. Data dengan positive skew atau yang kita sebut sebagai menceng ke kanan adalah kondisi ketika, data dominan berkumpul di nilai yang rendah. Negative skew adalah kebalikannya yakni data yang bernilai besar itu lebih dominan sementara data yang bernilai kecil itu lebih sedikit. Ditengah-tengahnya ada data dengan skewness nol yaitu data tersebar secara simetris kiri dan kanan. Ketika data memiliki distribusi yang simetris sempurna, maka mean, median dan modus dari data adalah sama. Sedangkan pada positive skew, mean akan lebih tinggi dari median karena mean akan lebih tertarik oleh nilai yang tinggi meskipun frekuensinya lebih kecil.</p>

# OTHER MEASUREMENT

## - Correlation

Korelasi juga merupakan ukuran hubungan dan mengukur kekuatan dan arah hubungan linier antara dua variabel. Jika korelasi terdeteksi maka itu berarti ada hubungan atau pola antara nilai dua variabel target.





## Transcription Pengantar Statistic dan Statistics Descriptive

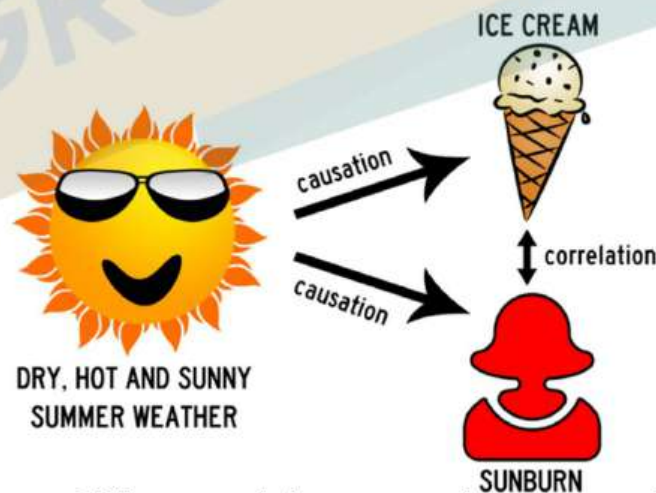
Time	Transcription
11:58–12:42	Ukuran lain yang dapat digunakan adalah korelasi. Korelasi adalah sebuah statistik yang mengukur kekuatan hubungan linier antara dua data. Korelasi memiliki nilai negatif 1 hingga 1 yang mana ketika korelasi bernilai negatif 1 artinya kedua variabel memiliki hubungan yang bertolak belakang. Sedangkan korelasi 1 berarti kedua variabel memiliki hubungan yang selaras ketika satu nilai naik maka nilai yang lainnya pun ikut naik begitu pula sebaliknya. Jika correlation bernilai sangat kecil mendekati 0 berarti tidak terlihat ada hubungan antar dua variabel tersebut.



# REMEMBER !

## - Correlation and Causality

Korelasi mengukur hubungan numerik antara dua variabel. Koefisien korelasi yang tinggi (mendekati 1), tidak berarti bahwa kita dapat dengan pasti menyimpulkan hubungan yang sebenarnya antara dua variabel.



## Transcription Pengantar Statistic dan Statistics Descriptive

Time	Transcription
12:42-end	<p>Meskipun begitu ada satu hal yang perlu diwaspadai mengenai korelasi bahwa korelasi tidak cukup untuk membuktikan kausalitas. Apa itu hubungan kausalitas? Kausalitas berarti sebab akibat sedangkan korelasi adalah hubungan berdasarkan kemiripan pola antar dua fenomena. Perhatikan contoh berikut ketika cuaca sedang sangat panas maka ada kecenderungan penjualan es krim pun akan meningkatkan karena orang ingin meredakan panas. Di saat sama karena cuaca panas maka potensi orang yang terkena sunburn atau terkena luka bakar karena terpapar sinar matahari terlalu lama itu pun juga akan meningkat. Banyaknya penjualan es krim dan kasus sunburn naik secara signifikan ketika cuaca sedang panas karena kedua fenomena meningkat bersama-sama maka ukuran korelasi akan menunjukkan nilai positif yang tinggi. Akan tetapi apakah tepat jika kita menyimpulkan bahwa penjual es krim akan mengakibatkan sunburn atau sebaliknya? Tentu tidak, jadi sebagai calon data analis kita harus berhati-hati sekali dalam menyimpulkan hubungan sebab akibat? Ingat! Bahwa karena dua hal terjadi secara bersamaan tidak berarti satu mengakibatkan yang lainnya.</p>