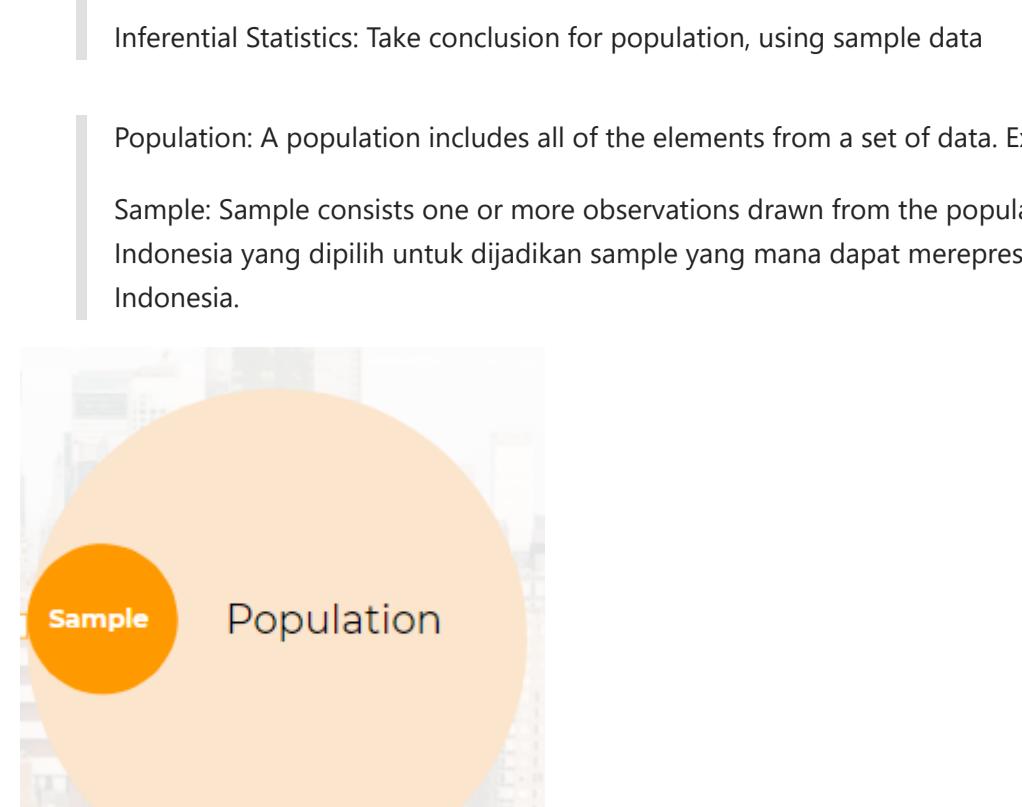


## What is Statistics?

Statistics (Statistik) is the discipline that concerns the collection, organization, analysis, interpretation and presentation of data.

Statistics (Statistik) is the knowledge of how to analyze data



Jadi ketika terdapat data, dan dari data tersebut di treatment dengan menggunakan ilmu statistika maka akan menghasilkan informasi yang bermanfaat.

In Business Case (example):

1. How to distribute promo for potential loyal customer?
2. How is the customer satisfaction if we use model A or Model B?
3. Why user become churn?
4. Is the product going better or worse?

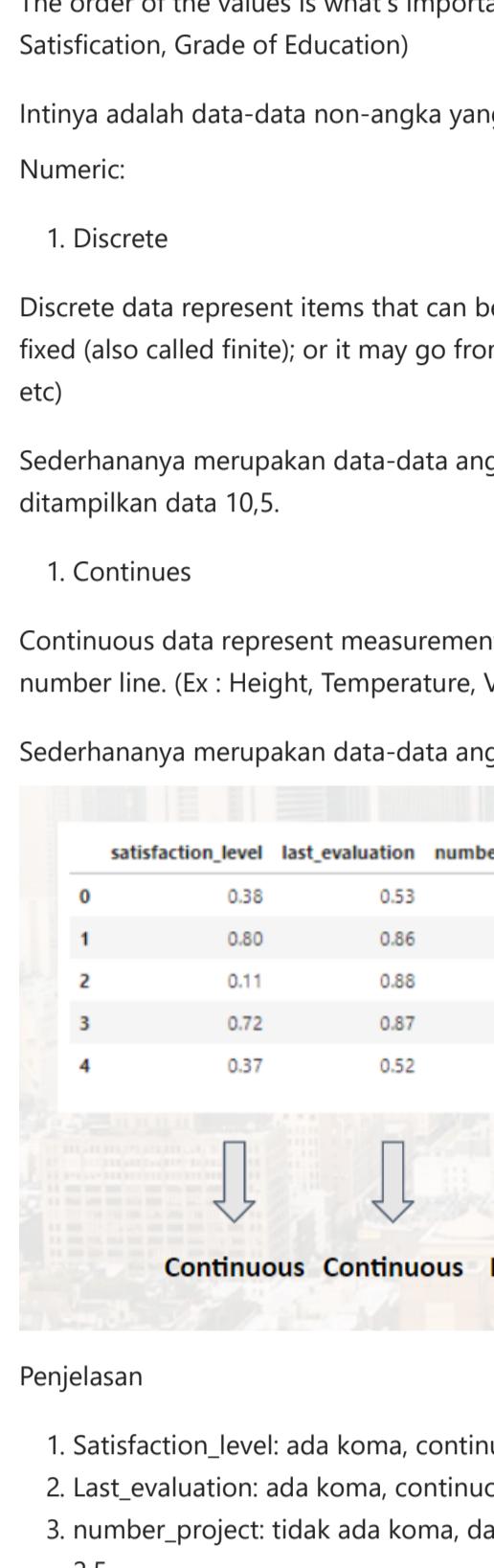
## How to make conclusion?

Descriptive Statistics: Take a description or overview of the data using some value and some visualization

Inferrential Statistics: Take conclusion for population, using sample data

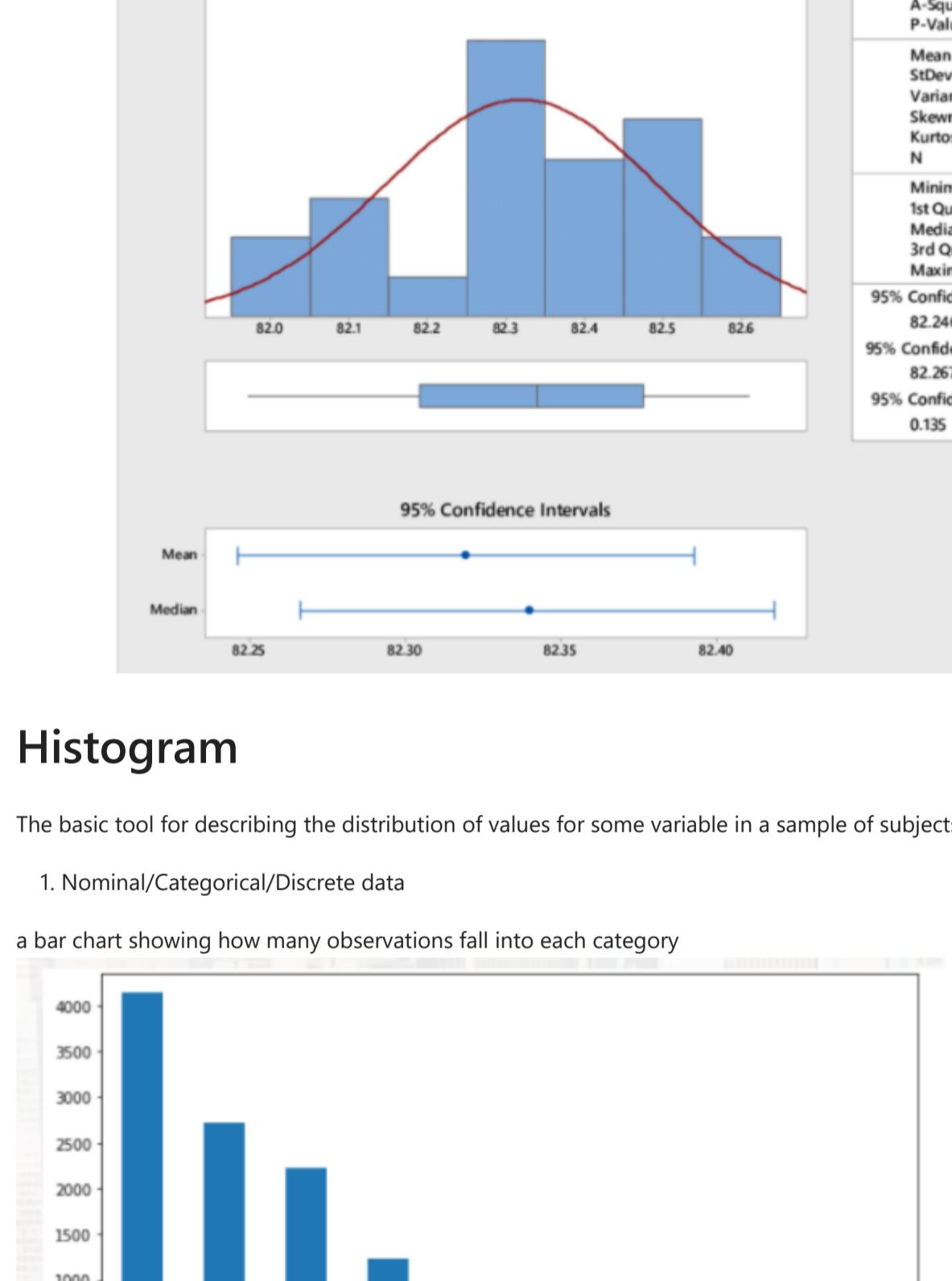
Population: A population includes all of the elements from a set of data. Ex: Seluruh Mahasiswa Universitas Indonesia.

Sample: Sample consists one or more observations drawn from the population. Ex: Beberapa Mahasiswa Universitas Indonesia yang dipilih untuk dijadikan sample yang mana dapat merepresentasikan seluruh mahasiswa Universitas Indonesia.



Only  $x\%$  of population is taken, and we will take an inference from the parameter for population.

## Parameter vs Statistics



Parameter merupakan rata-rata populasi

Statistics merupakan rata-rata sample.

Dengan menggunakan statistics, diharapkan mampu merepresentasikan populasi atau parameter.

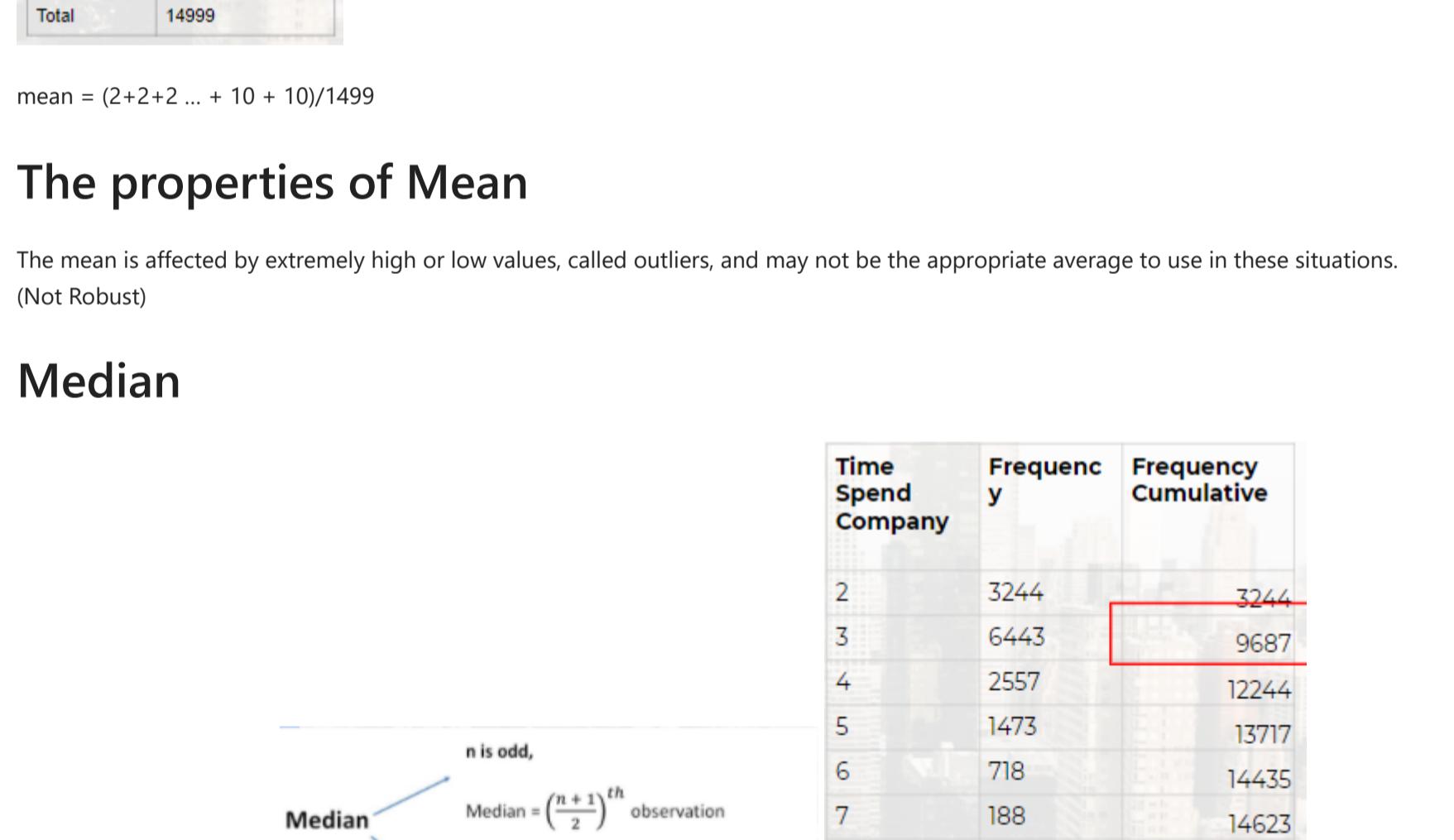
## What is Data?

Individual units of information

Organized in data matrix:

	satisfaction_level	last_evaluation	number_project	average_monthly_hours	time_spend_company	Work_accident	left	promotion_last_5years	sales	salary
Observation	0	0.38	0.53	2	157	3	0	1	0 sales	low
	1	0.80	0.86	5	262	6	0	1	0 sales	medium
	2	0.11	0.88	7	272	4	0	1	0 sales	medium
	3	0.72	0.87	5	223	5	0	1	0 sales	low
	4	0.37	0.52	2	159	3	0	1	0 sales	low

## Type of Variables



Tipe data secara umum dibagi 2:

1. Category yang merupakan data-data non-angka
2. Numeric yang merupakan data-data angka.

Category:

1. Nominal

Nominal scales are used for labeling variables, without any quantitative value. "Nominal" scales could simply be called "labels." (Ex : Male-Female, Color, Occupation)

Intinya data-data non angka yang mana tidak ada tingkatan (lebih besar atau lebih kecil). Misalnya data gender, laki-laki dan perempuan setara, tidak ada yang lebih besar dan tidak ada yang lebih kecil.

1. Ordinal

The order of the values is what's important and significant, but the differences between each one is not really known. (Ex : Feeling, Satisfaction, Grade of Education)

Intinya adalah data-data non-angka yang mana ada tingkatannya, misalnya jenjang pendidikan, SD < SMP < SMA < S1

Numeric:

1. Discrete

Discrete data represent items that can be counted; they take on possible values that can be listed out. The list of possible values may be fixed (also called finite); or it may go from 0, 1, 2, on to infinity (making it countably infinite). (Ex : Number of Student, Number of Vehicle, etc)

Sederhananya merupakan data-data angka yang mana tidak decimal atau countable. Misalnya tidak mungkin jumlah kematian bayi ditampilkan data 10,5.

1. Continuous

Continuous data represent measurements; their possible values cannot be counted and can only be described using intervals on the real number line. (Ex : Height, Temperature, Velocity, etc)

Sederhananya merupakan data-data angka yang mana terdapat decimal. Misalnya tinggi badan 165,5 cm

satisfaction_level	last_evaluation	number_project	average_monthly_hours	time_spend_company	Work_accident	left	promotion_last_5years	sales	salary
0	0.38	0.53	2	157	3	0	1	0 sales	low
1	0.80	0.86	5	262	6	0	1	0 sales	medium
2	0.11	0.88	7	272	4	0	1	0 sales	medium
3	0.72	0.87	5	223	5	0	1	0 sales	low
4	0.37	0.52	2	159	3	0	1	0 sales	low

Penjelasan

1. Satisfaction\_level: ada koma, continuous
2. Last\_evaluation: ada koma, continuous
3. Number\_project: tidak ada koma, dan tidak mungkin ada koma, karena merepresentasikan jumlah project yang diambil, tidak mungkin 2,5
4. Average\_monthly\_hours: karena menghitung rata-rata dan data yang ditampilkan sudah dibulatkan, walaupun tidak ada koma tetapi continuous
5. Time\_spend\_company: casenya sama seperti nomor 4
6. Work\_accident: merepresentasikan sudah pernah atau belum terjadi kecelakaan kerja. Jadi nominal
7. Left: sama seperti nomor 6
8. Promotion\_last\_5years: sama seperti nomor 6
9. Sales: jelas nominal karena tidak ada tingkatan
10. Salary: ordinal, karena terdapat tingkatan dan non-angka

## Statistics Descriptive

The first step in analyzing a set of data is to have a good idea of what the data looks like. This is the job of descriptive statistics

### Summary Report for Diameter

#### Anderson-Darling Normality Test

A-Squared: 0.26  
P-Value: 0.646

Mean: 82.326  
StDev: 0.174  
Variance: 0.030

Kwensian: -0.31911  
Kurtosis: -0.58949

N: 24

Minimum: 81.970  
1st Quartile: 82.240  
Median: 82.340  
3rd Quartile: 82.475  
Maximum: 82.600

95% Confidence Interval for Mean: 82.246  
82.393

95% Confidence Interval for Median: 82.267  
82.419

95% Confidence Interval for StDev: 0.135  
0.244

### 95% Confidence Intervals

Mean: [82.25, 82.40]  
Median: [82.30, 82.35]

Continuous

Continuous

Discrete

Continuous

Nominal

Nominal

Ordinal

Continuous

Continuous