FULLSTACK INTENSIVE BOOTCAMP

Data Analyst

MySkill

Let's Learn Basic Statistic

Oleh: Wildan Zakaria

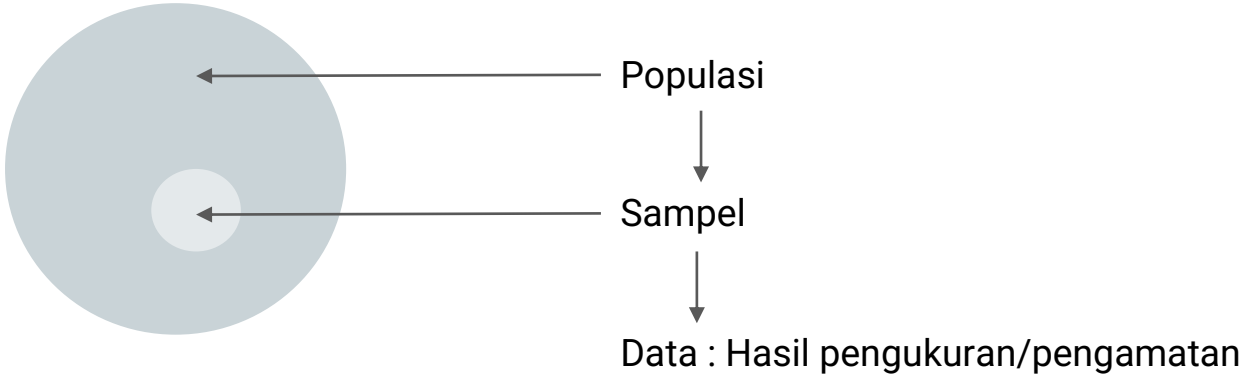
MINI PORTOFOLIO

Course Summary

Statistik : nilai-nilai ukuran data fakta yang mudah dimengerti
Statistika : ilmu yang berkaitan dengan cara pengumpulan, pengolahan, analisis, dan penarikan kesimpulan atas data.

Type of statistics

- Statistika deskriptif : Mendeskripsikan dan merangkum data
- Statistika inferensial : Membuat dugaan dari sampel data



Course Summary

Tipe data/observasi

1. Kualitatif/kategorical
- Nominal : tidak berurutan, contoh jenis kelamin laki-laki dan perempuan
- Ordinal/rank : berurutan, contoh rating buruk, cukup baik, baik, sangat baik

2. Kuantitatif/numerical
- Diskrit : terhitung/integer, contoh jumlah orang dirumah
- Kontinu : terukur mencakup nilai bilangan rill, contoh panjang meja 1,5 cm

Metode sampling
- Random sampling : pengambilan sampel yang dilakukan secara acak
- Sistematik sampling :pengambilan sampel dilakukan dengan mengurutkan terlebih dahulu data dari populasi kemudian dipilih secara berkala/berjarak.
- Stratified sampling : pengambilan sampel dilakukan secara acak namun sebelumnya telah dibuat menjadi beberapa karakteristik/kategori terlebih dahulu.

Course Summary

Parameter distribusi

- **Ukuran pemusatan**

Mean : nilai rata-rata
Median : nilai tengah.
Mode/modus : nilai yang sering muncul
Kuartil atas Q3, P75 : nilai yang membagi 2 bagian dengan 25% atas/kanan dan 75% bawah/kiri
Kuartil tengah Q2, P50 : nilai yang membagi 2 bagian dengan jumlah yang sama.
Kuartil bawah Q1, P25 : nilai yang membagi 2 bagian dengan 75% atas/kanan dan 25% bawah/kiri

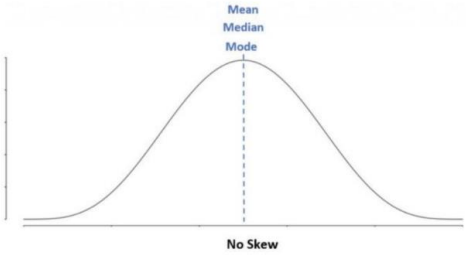
- **Ukuran penyebaran**

Range : jarak antara nilai terbesar dengan nilai terkecil.
Standar deviasi : mengukur jarak rata-rata dari sekelompok data dari nilai-nilai kelompok tersebut.
Variansi : nilai rata-rata dari perbedaan dari setiap data dari mean.

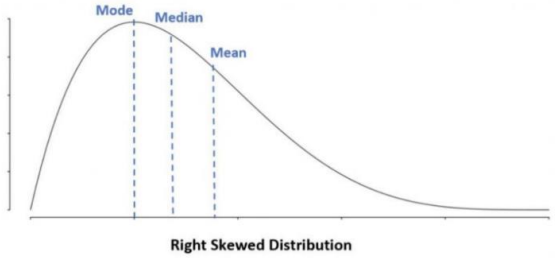
Course Summary

- **Karakteristik distribusi**

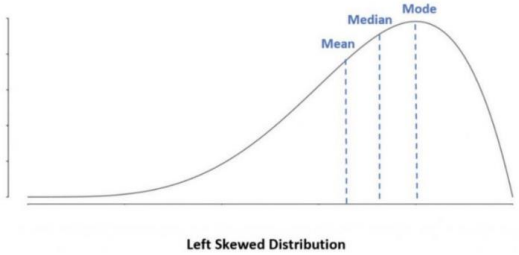
Simetris : mean = median = mode
Left skewed/skew positif : mode > median > mean
Right skewed/skew negative : mode < median < mean
Puncak : berpuncak tunggal (modus = 1), berpuncak jamak (modus > 1)



Distribusi simetris: mean = median = mode



Distribusi menceng kanan: mode < median < mean



Distribusi menceng kiri: mean < median < mode

Course Summary

- **Outlier/pencilan :** Data yang memiliki karakteristik yang berbeda jauh dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim baik untuk variabel tunggal atau variabel kombinasi.
- **IQR (Inter Quartil Range)** : jarak antar quartil yang berguna untuk nge-define apakah nilai dalam data tersebut termasuk outlier atau masih dapat di consider dalam perhitungan.(mengidentifikasi fraud)
- **Uji hipotesis** : metode statistik yang digunakan dalam pengambilan keputusan statistik dengan menggunakan data eksperimen.

  H0 : Hipotesis yang ingin diuji. Dapat berupa: hasil penelitian sebelumnya, info dari buku, atau hasil percobaan orang lain.

  H1 : Hipotesis yang ingin dibuktikan.
- **Regresi :** menentukan/menaksir parameter-parameter yang terlibat dalam suatu model matematik yang linear terhadap parameter-parameter tersebut. (memprediksi/forecasting nilai suatu variable)

15 MINS: LET'S DO THIS!

Date	Voucher Redeem
2023-01-01	98
2023-01-02	112
2023-01-03	1
2023-01-04	41
2023-01-05	106
2023-01-06	111
2023-01-07	86
2023-01-08	142
2023-01-09	143
2023-01-10	88
2023-01-11	32
2023-01-12	66
2023-01-13	92
2023-01-14	94
2023-01-15	118
2023-01-16	53
2023-01-17	119
2023-01-18	12
2023-01-19	37
2023-01-20	90
2023-01-21	128
2023-01-22	79
2023-01-23	86
2023-01-24	45
2023-01-25	33
2023-01-26	54
2023-01-27	115
2023-01-28	129
2023-01-29	56
2023-01-30	83

Case

You are a part of CRM Team and asked to evaluate the redeemed voucher of the day

1. What are the
   - mean
   - median
   - mode
   - std.dev
   - and upper & lower threshold (Cari tau apakah ada outlier of the last month performan)
2. Is there any outlier from the last month performance?
3. If yes? how many?

Answer

Mean	81.63
Median	87
Mode	86
Var	1460.309195
Std Dev S	38.21399214
Percentile 25	53.25
Percentile 75	111.75
IQR	58.5
LOWER THRESHOLD	-34.5
UPPER THRESHOLD	199.5
COUNTIF	0

Note : no outlier from the last month performance

15 MINS: LET'S DO THIS!

Transaction ID	service duration
1	1
2	4
3	8
4	5
5	5
6	1
7	10
8	5
9	6
10	1
11	1
12	10
13	2
14	1
15	4
16	6
17	8
18	4
19	9
20	5
21	10
22	7
23	10
24	2
25	10
26	2
27	7
28	10
29	7
30	4
31	3
32	10
33	8
34	10
35	4
36	10
37	10
38	5
39	10
40	9
41	3
42	7
43	9
44	9
45	6
46	3
47	5
48	7
49	6
50	8

Task

In one money transfer company, the expected of P90 SLA is under 5 mins to ensure the customer satisfaction of the service provided. There are 50 transactions occured with each of SLA is attached. Is the company was achieved P90 Satisfaction Level Condition? What is your recommendation to product team to solve this condition?

Answer

P90 10

The company was not achived P90 Satisfication Level Condition, because P90 SLA from 50 transactions can under 10 mins. I recommendation to product team in companies to give P90 SLA under 10 mins to ensure the customer satisfaction of the service provided.

15 MINS: LET'S DO THIS!

Date	Visitor of the Public Service
2022-12-01	4
2022-12-02	2
2022-12-03	5
2022-12-04	10
2022-12-05	7
2022-12-06	8
2022-12-07	10
2022-12-08	5
2022-12-09	8
2022-12-10	4
2022-12-11	9
2022-12-12	6
2022-12-13	10
2022-12-14	8
2022-12-15	4
2022-12-16	6
2022-12-17	8
2022-12-18	5
2022-12-19	7
2022-12-20	3
2022-12-21	9
2022-12-22	7
2022-12-23	5
2022-12-24	4
2022-12-25	7
2022-12-26	7
2022-12-27	7
2022-12-28	5
2022-12-29	8
2022-12-30	5

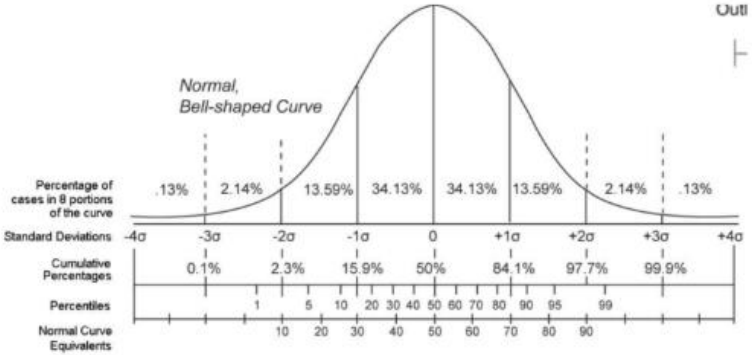
Case

If you are the leader of one institution in Surabaya, how many chair that you need to prepared to cover 68% of all visitior at least will get the seat?

Answer

Mean 6.4

Stdev 2.1



With assuming the data is normally distributed
We can prepare the chair to cover 68% of the seat at least for
mean + 1 * std. deviasi
6,4 + 1 * 2,1
8.6

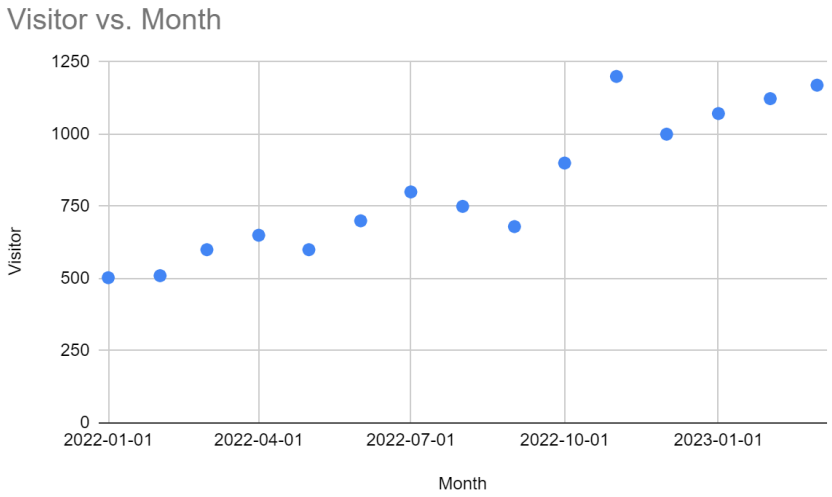
9 chair we need to prepare.

15 MINS: LET'S DO THIS!

Month	Visitor	
2022-01-01	503	
2022-02-01	510	
2022-03-01	600	
2022-04-01	650	
2022-05-01	600	
2022-06-01	700	
2022-07-01	800	
2022-08-01	750	
2022-09-01	680	
2022-10-01	900	modal belajar buat
2022-11-01	1200	bikin formula /
2022-12-01	1000	persamaan dari si linear regresi
2023-01-01	1071	x untuk prediksi
2023-02-01	1123	
2023-03-01	1170	

You are the manager of the museum, you asked to forecast and visualize the forecast of the 2023 Q1 potential visitor

1. Provide a forecast for Q1 2023
2. Provide the visualization (in scatter plot)



Follow me!

Instagram : @wildanzkkr
Twitter : @rekfoop
LinkedIn : wildan zakaria

Bootcamp Data Analysis
by @myskill.id