Descriptive stats. ⇒

1. central tendency
2. Dispersion measure.
3. PDF | PMF

Dispersion measure

1. Range.
2. Variance / standard deviation
3. mean absolute dev.
4. IQR (quartile)

min, Q_1, Q_2, Q_3 max.

five number summary

** Measure of dispersion describe the spread of the data

1. Range $\Rightarrow$ $\boxed{\text{maximum} - \text{minimum}}$

Consider two Points in the estimation.

Simple measure of variation (dispersion)

Ex:-1. $\boxed{1}$, 1, 1, 1, 2, 2, 2, 3, 3, 4, 4, 4, 4, $\boxed{5}$

Range of data $\Rightarrow$?

$\Rightarrow 5 - 1$

$\Rightarrow 4$

Ex:-2 $\boxed{1}$ 1, 1, 2, 2, 2, 2, 3, 4, 4, 5, $\boxed{120}$

Range of the data = ?

$\Rightarrow 120 - 1$

$\Rightarrow 119$

Range is sensitive to outlier

② **Variance** ⇒ Variance indicate how close or far the data points from the mean.

Population variance ⇒

$$\boxed{\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}$$

Sample variance ⇒

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

$$\Downarrow$$

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

(sample variance with the bessel's correction)

Ex:-

10, 12, 14, 15, 17, 18, 18, 24

$N = 8$

mean $= \frac{128}{8} = 16$

$$\boxed{\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$$

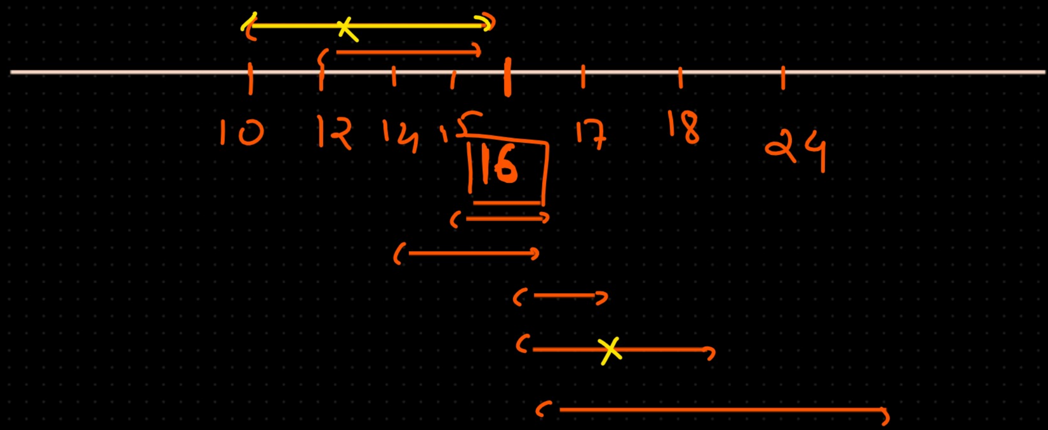
$n-1 \Rightarrow$ bessal's correction

$8-1 = 7$

$$\Rightarrow \frac{(10-16)^2 + (12-16)^2 + (14-16)^2 + (15-16)^2 + (17-16)^2 + (18-16)^2 + (18-16)^2 + (24-16)^2}{8}$$

$$\Rightarrow \frac{(-6)^2 + (-4)^2 + (-2)^2 + (-1)^2 + (1)^2 + (2)^2 + (2)^2 + (8)^2}{8}$$

$$\Rightarrow \frac{36 + 16 + 4 + 1 + 1 + 4 + 4 + 64}{8} = \frac{130}{8} = \boxed{16.25}$$



= Squre

Spread of the Data around mean

Data Point

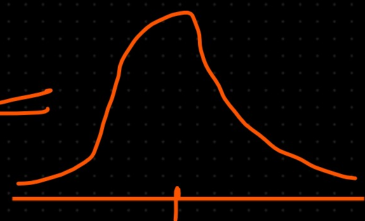
Variance = 16.25

data Point

20.15

Data Point

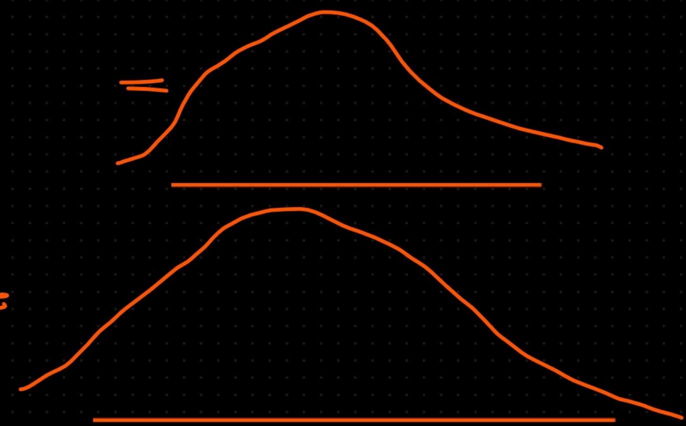
= 40.30



$N = 8$

$\Downarrow$

$\boxed{n=6}$ $\Rightarrow$ Sample variance



$\Downarrow$

Banal's. $\Rightarrow$ $n-1$

$= 6-1 = 5$

Standard deviation $\Rightarrow$ $\sqrt{\text{variance}}$

Population standard dev $\Rightarrow$ $\sqrt{16.25} \Rightarrow \boxed{4.03}$

Variance $\Rightarrow$ $$\frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n}$$

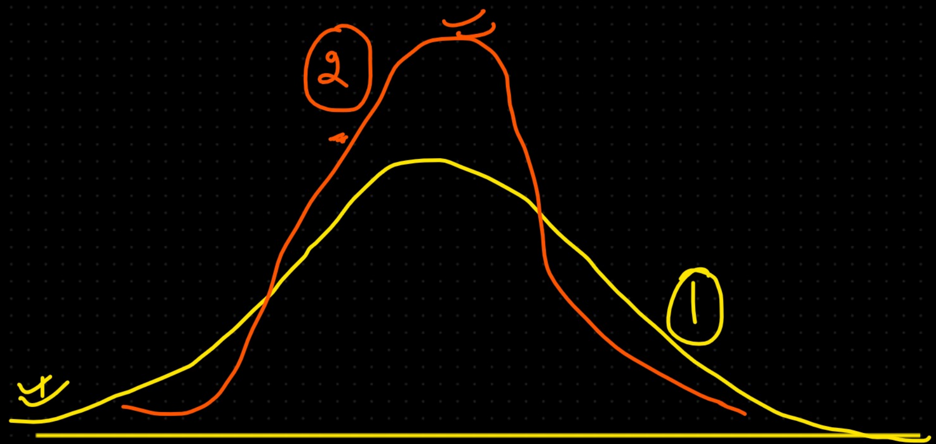
$\Rightarrow$ it's very high.

$\Rightarrow$ Compress that value.

$\Rightarrow$ square root of that value.

1) the more the data are spread out, the greater the Range, variance and std dev.

2) the more the data are concentrated, the smaller the Range, variance and std dev.



Mean absolute deviation. (MAD)

$$\boxed{\frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{n}}$$

$|-2|$

$\Rightarrow +2$

$|-5|$

$\Rightarrow +5$

$|10|$

$\Rightarrow 10$

$$\frac{\text{mean absolute deviation}}{\sum_{i=1}^{n} \frac{|x_i - \bar{x}|}{n}}$$

$-ve \Rightarrow$ abosulte $\Rightarrow +ve$

Variance $(x_i - \bar{x})^2 \Rightarrow$ (very high value) $\Rightarrow$ cancling the neg. effect around the mean.

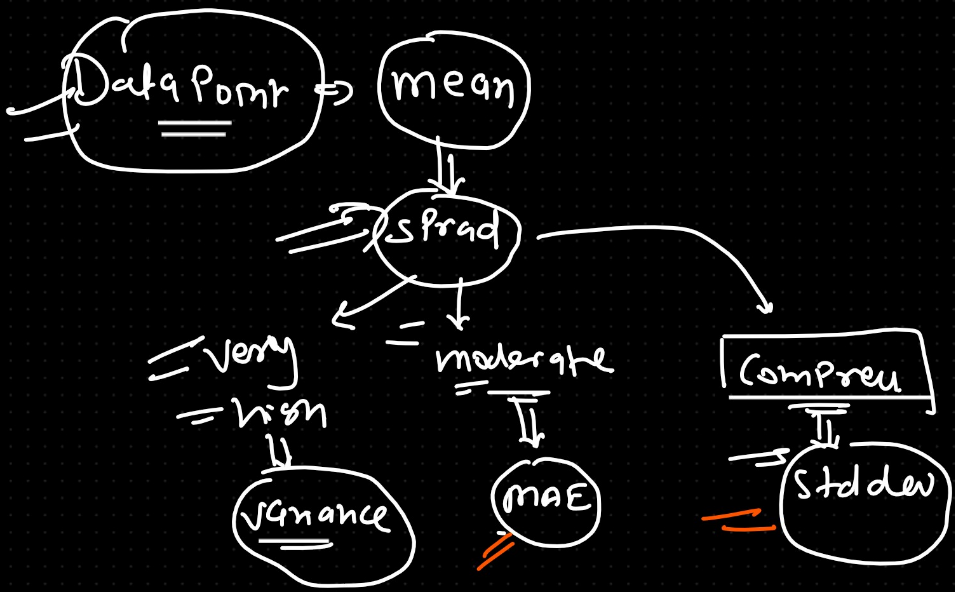
MAD $= |X_i - \bar{X}| \Rightarrow$ absolute value $\Rightarrow$ Cancelling neg. effect around the mean.

Std dev $\sqrt{\text{variance}} \Rightarrow$ compress the value of the variance.

Variance ⇒ MSE
std dev ⇒ RMSE
MAD ⇒ MAE

} loss

Assignment ⇒ (3 to 5 diff b/w variance | std dev | MAD)



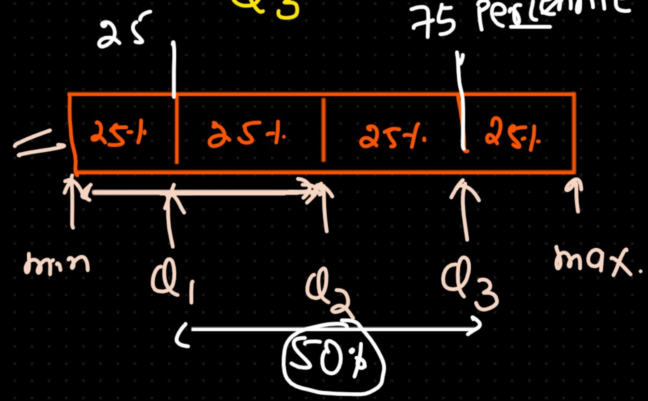
$(x_i - \bar{x})^2$ $|x_i - x|$

Quartile measure :

Q_1

Q_2

Q_3



$$IQR = Q_3 - Q_1$$

lowerfence $\Rightarrow$ $\boxed{Q_1 - 1.5 \times IQR}$

upperfence $\Rightarrow$ $\boxed{Q_3 + 1.5 \times IQR}$

for idenfication of outliers

Ques.

11, $\boxed{12, 13}$, 16, $\boxed{16}$, 17, $\boxed{18, 21}$, 22

$Q_1 = ?$

$Q_2 = ?$

$Q_3 = ?$

$\boxed{n = 9} \Rightarrow$ Odd

$$Q_1 = 25\% \times n = \frac{25}{100} \times n = \frac{1}{4} \times (n+1) = \frac{n+1}{4} = \frac{10}{4}$$

$= 2.5$

2.5

$\Downarrow$

12.5

$$Q_1 = \frac{12+13}{2} = 12.5$$

$$Q_2 = 50\% = (n+1) \times \frac{1}{2} = \frac{10}{2} = 5 \Rightarrow \text{Position}$$

$Q_2 = 16 \Rightarrow$ 5 Position

$\Downarrow$

median

$Q_3 = 75\% \Rightarrow \frac{(n+1)\times 3}{4} = \frac{10\times 3}{4} = 7.5$ Position

$\Rightarrow \frac{18+21}{2} = 19.5 \Rightarrow 7.5$

Quartile ⇒ divide the entire data into 4 segment

Q_2

11	12	13	16	16	17	18	21	22.

Q_1

$Q_1 = 25$ Percentile $\rightarrow \underset{\text{Pos}}{2.5} \Rightarrow \underset{\text{value}}{12.5}$

$Q_2 = 50$ Percentile $\Rightarrow \underset{\text{Pos}}{5} \Rightarrow \underset{\text{value}}{16} \Rightarrow$ median

$Q_3 = 75$ Percentile $\Rightarrow \underset{\text{pos}}{7.5} \Rightarrow \underset{\text{value}}{19.5}$

$$\boxed{IQR = Q_3 - Q_1 = 19.5 - 12.5 \Rightarrow 7}$$

five number summary. $\Rightarrow$

min	11
Q_1	12.5
Q_2	16
Q_3	19.5
max	22.

$IQR = Q_3 - Q_1$

$= 19.5 - 12.5$

$= 7$

lower fence. $\Rightarrow$ $Q_1 - IQR \times 1.5$

$\Rightarrow 12.5 - (7 \times 1.5) \Rightarrow 12.5 - (10.5)$

$\Rightarrow 2$

upper fence $\Rightarrow$ $Q_3 + (IQR \times 1.5)$

$\Rightarrow 19.5 + (7 \times 1.5) \Rightarrow 19.5 + (10.5) \Rightarrow 30$

five number summary. ⇒ min 11

Q_1 12.5

Q_2 16

Q_3 19.5

max 22.

IQR. 7

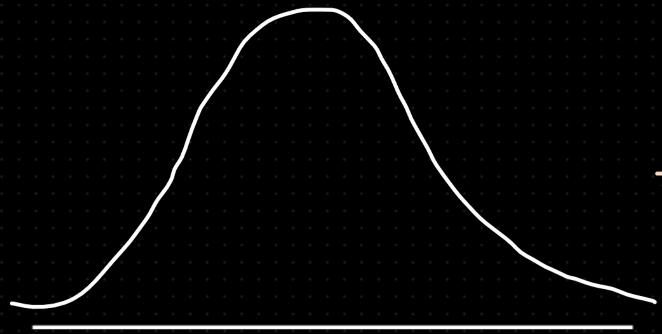
lower fence 2

upper fence 30 ⇒

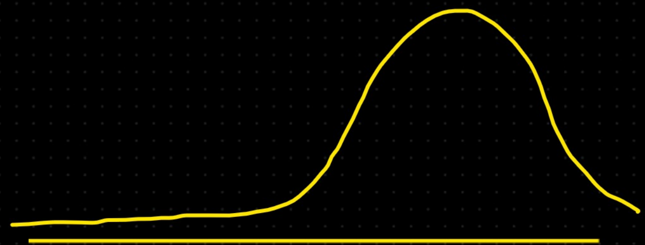
$\boxed{-5\ 0\ 1}$ 5 10 $\boxed{11\ 12\ 13\ 16\ 16\ 17\ 18\ 21\ 22.}$ 30, $\boxed{50, 65}$

right skewed.

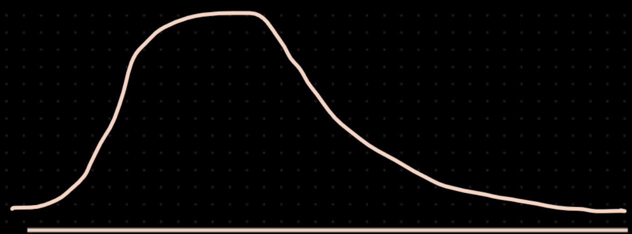
left skewed.



⇒ Nooutlier

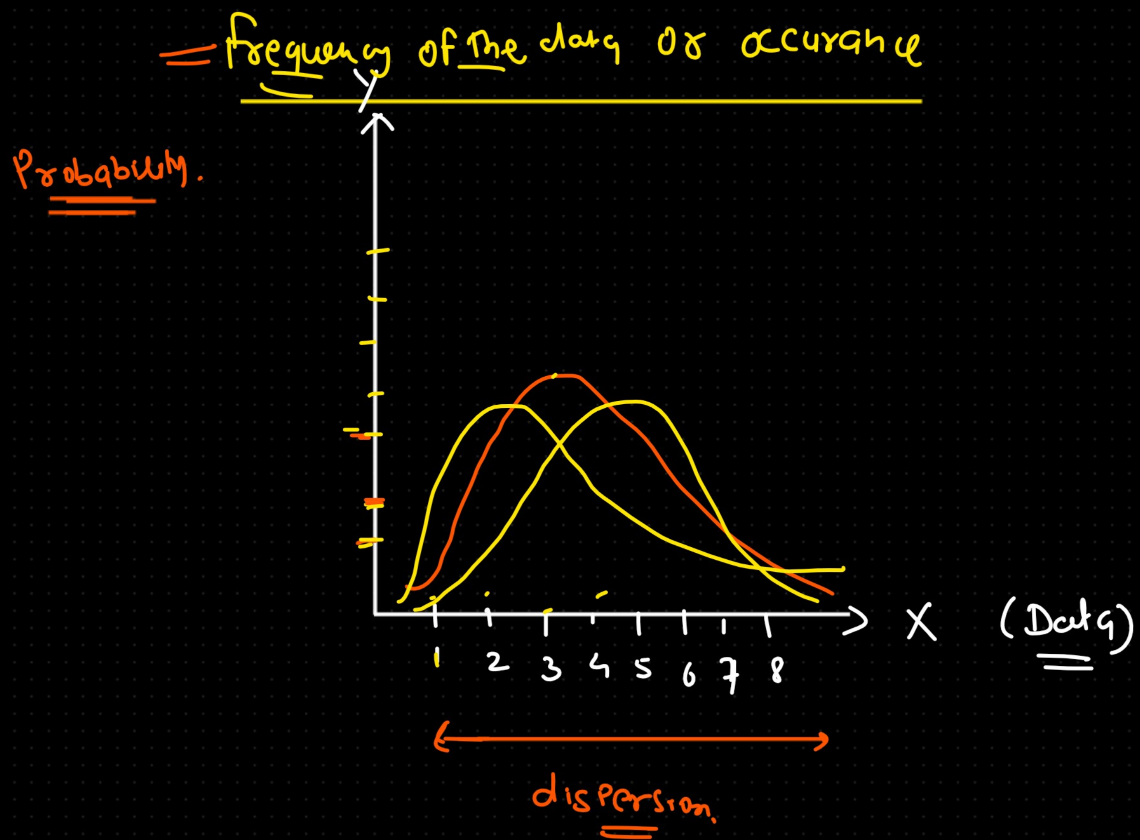


⇒ left skewd mean.
outlier will be on left hand side



⇒ right skewed mean outlier will be on right hand side

Data ⇒



= Frequency of the data or occurance
Y
Probability.
1
2
3
4
5
6
7
8
X
(Data)
dispersion.

Data points => { 10, 12, 14, 18, 24, 26, 30, 35, 36, 37, 40, 41, 42, 43, 50, 51 }

500

value
size

=50 / 5
=> 10

NO

⇓

bin

⇓

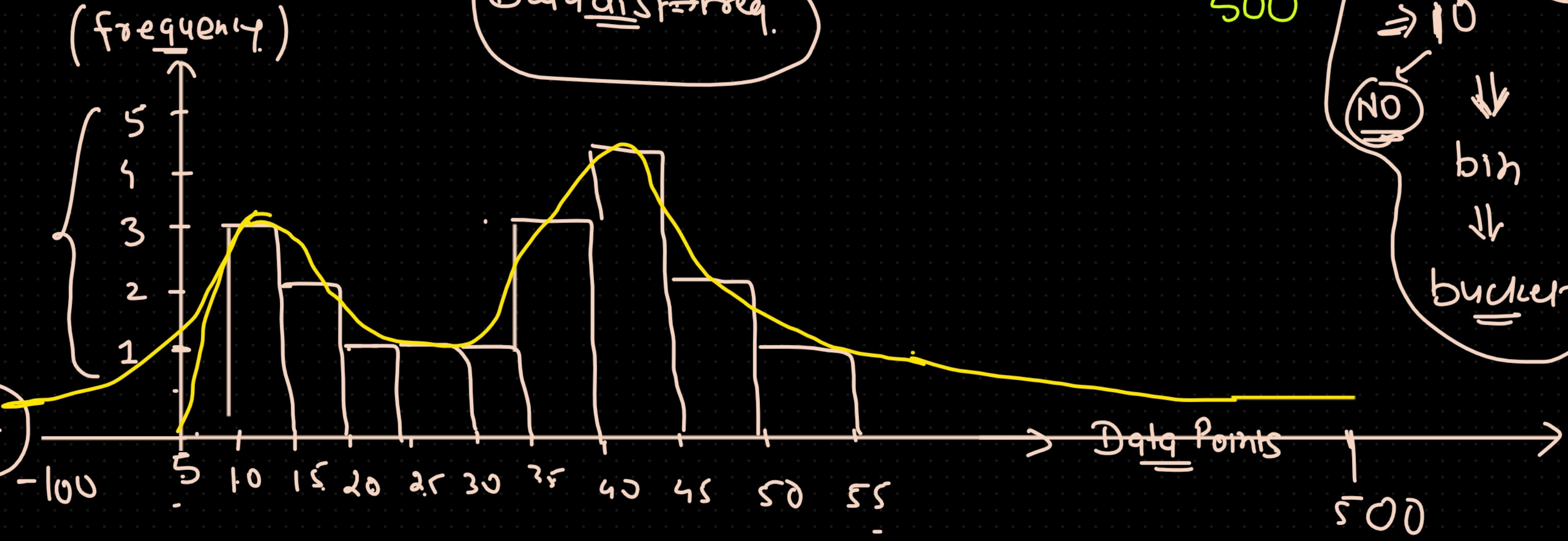
bucket

frequency distribution of data

Data dist → freq.

Histogram

Prob. → pdf



bins (diff is 5)

⇓

11

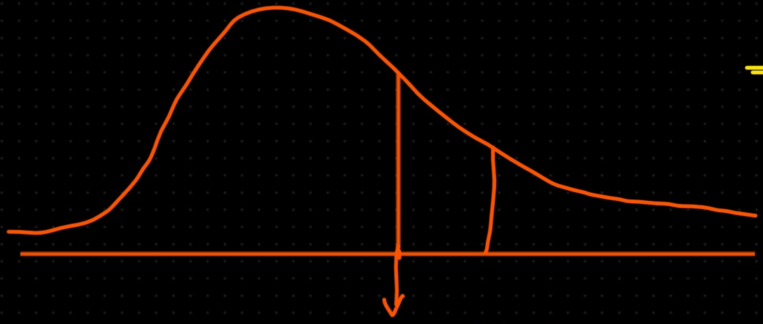
Positive outlier

my data dist will be right skewed.

-ve outlier

my data dist will be left skewed

right skewed ⇒ mean is greater ✓
median is greater ✗



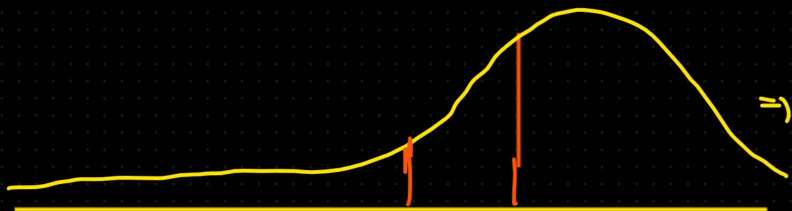
⇒ right skewed

[2, 3, 5, $\boxed{8}$, 10, 100, 550]

mean = 96.8
median = 8
mode =

$\boxed{\text{median} < \text{mean}}$

$\boxed{\text{median} < \text{mean}}$



⇒ left skewed

$\boxed{\text{mean} < \text{median}}$

[-300, -150, $\boxed{2, 3}$, 6, 9]

mean = -71.66
median = 2.5
mode =

$\boxed{\text{mean} < \text{median}}$

- Probability | PDF / PDF / PMS

Random variable ⇒ continues | Discreate

Diff-2 based on concept