Descriptive analytics

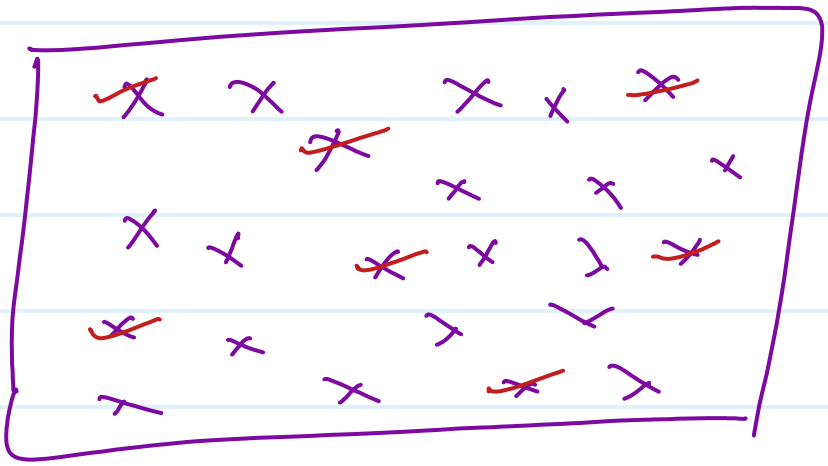
Population - N, μ

Sampling - n, $\bar{x}$

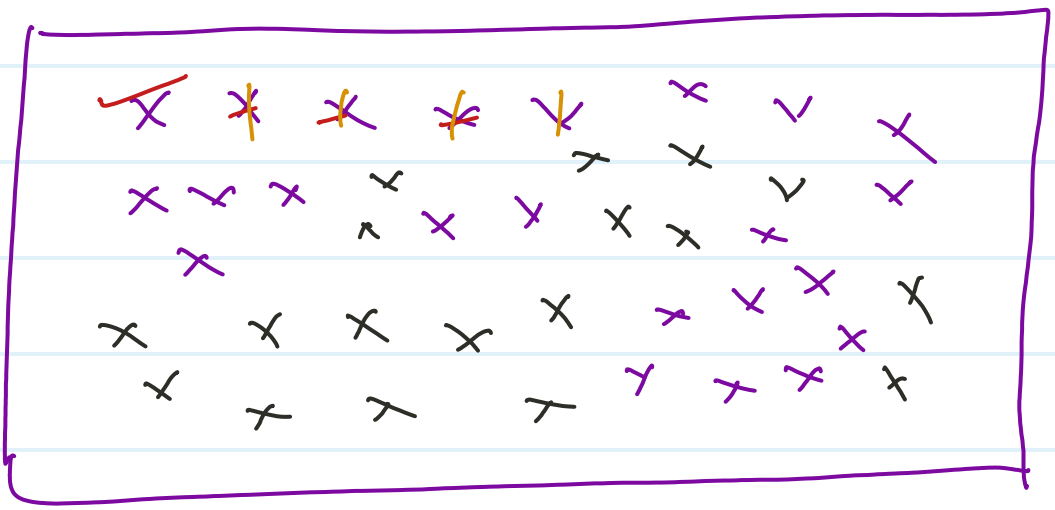
✡ Sampling Types

1. Simple random sampling
2. Stratified sampling
3. Systematic sampling
4. Convenience sampling

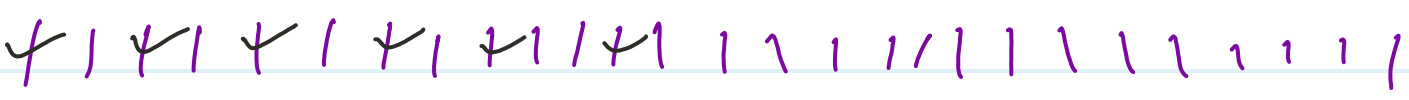
① simple random sampling



② stratified samp.



③ systematic samp.



④ Convi. Samp.

median

$[1, 2, 3, 4, 5, 6, 7, 8]$

$$\frac{4+5}{2} = 4.5$$

mode

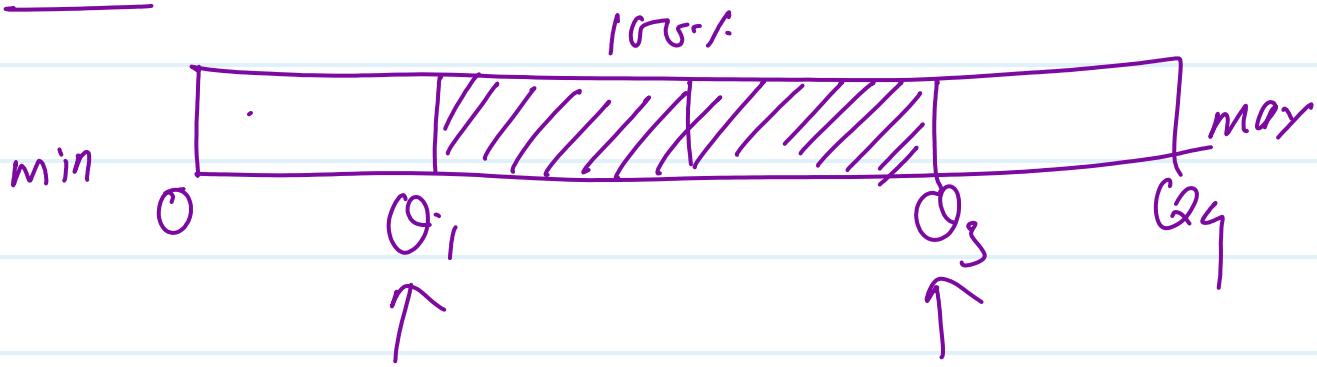
uni-modal

$[1, 2, 3, 4, 4, 5, 6, 4, 6, 6]$

Bi-modal

$$[2, 20, 15, 35, 40, 55, 57, 77]$$

IQR



Variance

$$N \rightarrow \quad \sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

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

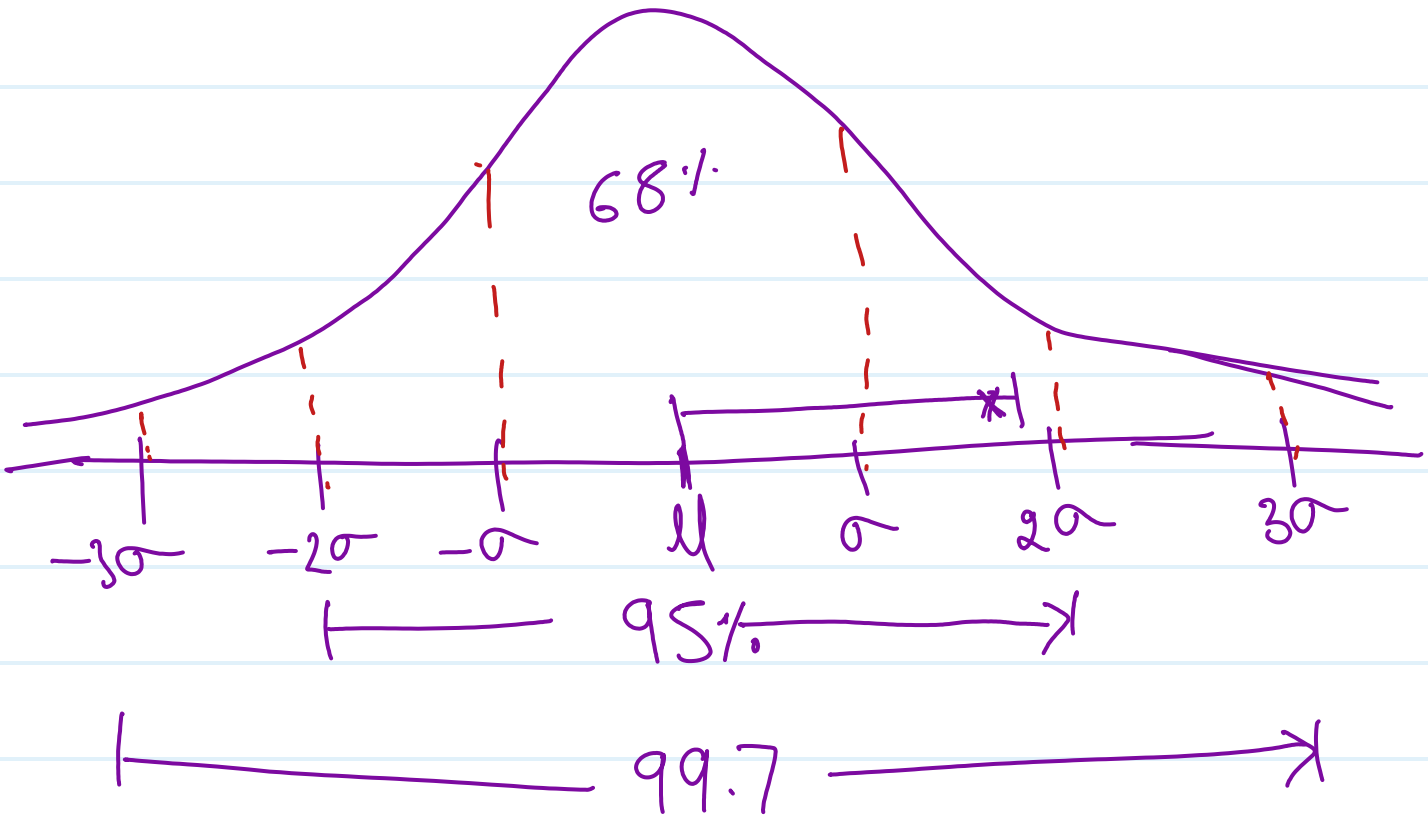
$(n-1)$ Degree of freedom

OR Bessels correction

SD

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

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

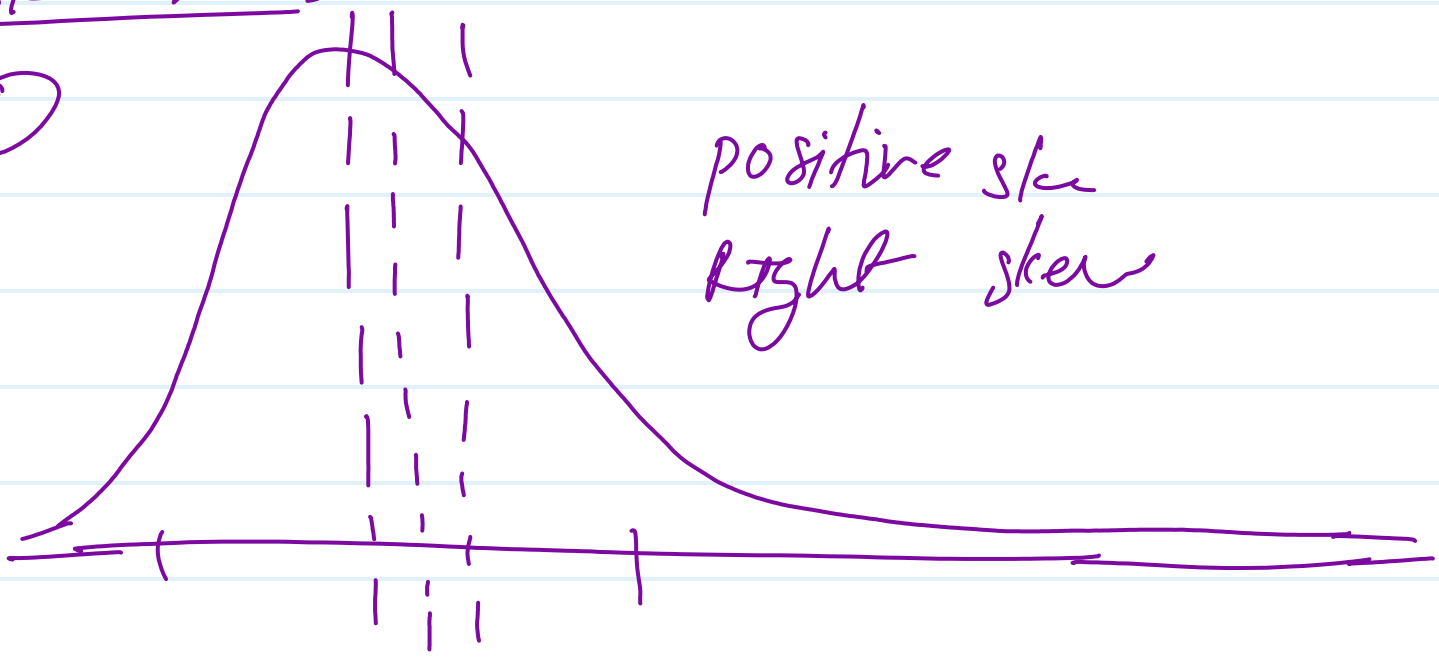


68 - 95 - 99.7

It called Empirical rule of 1 D.

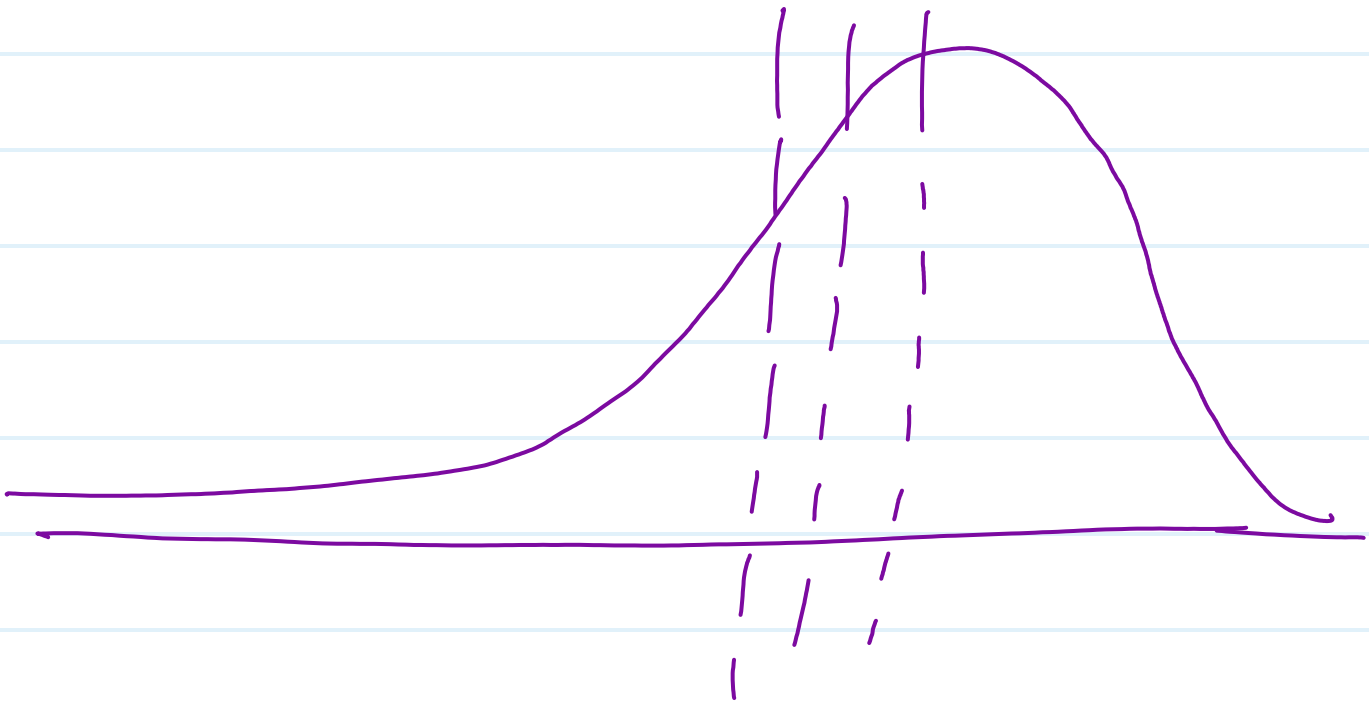
Skewness

① positive skew / Right skew



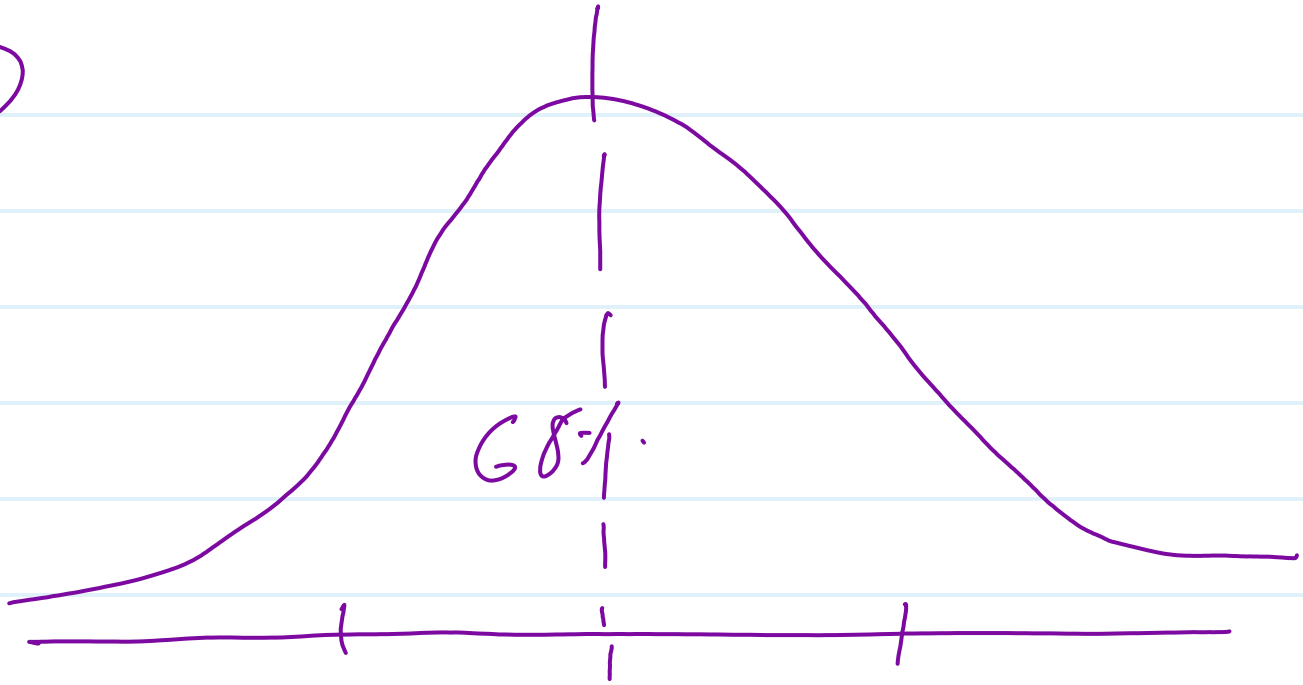
mean > median > mode

②



mean < median < mode

(3)

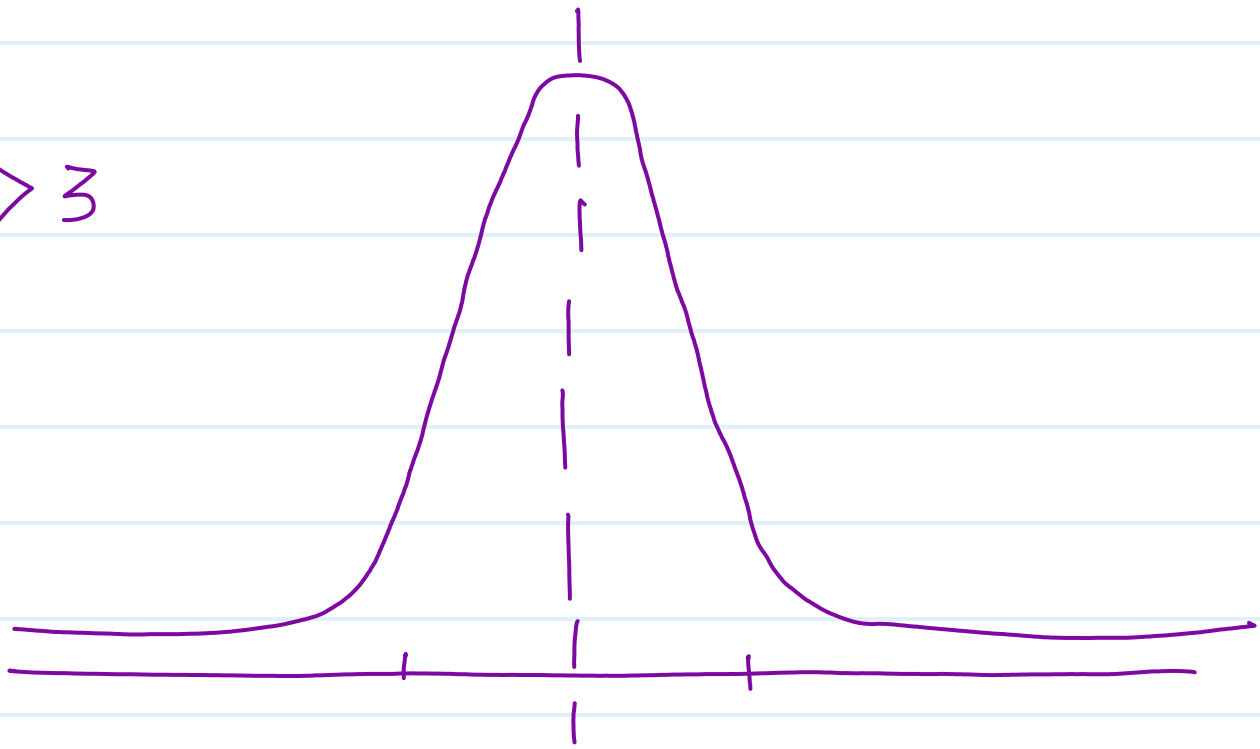


mean = median = mode

Kurtosis — $K = 3$

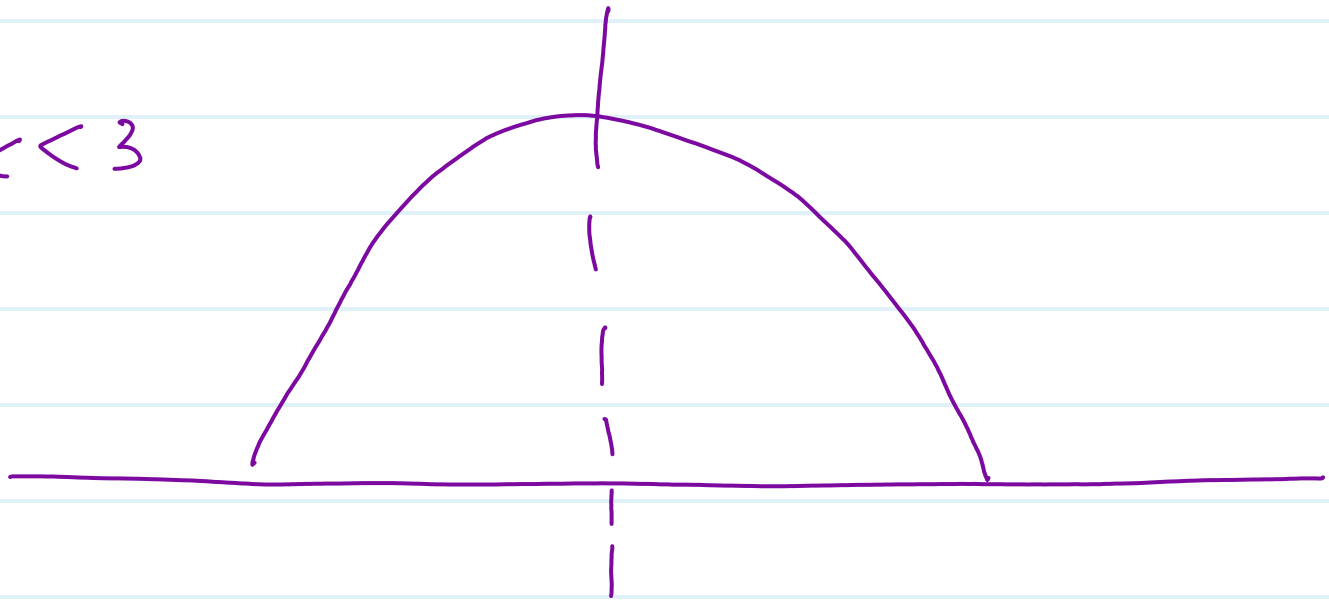
(1) Lepto kurtosis —

$K > 3$



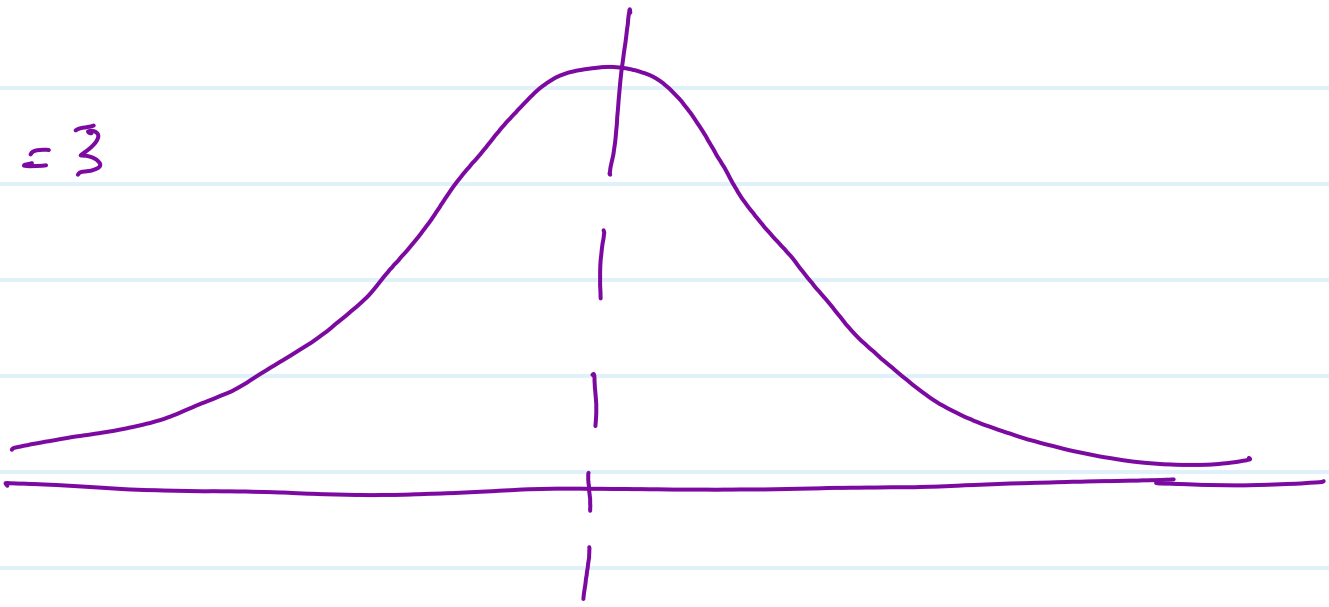
② platykurtosis

$k < 3$

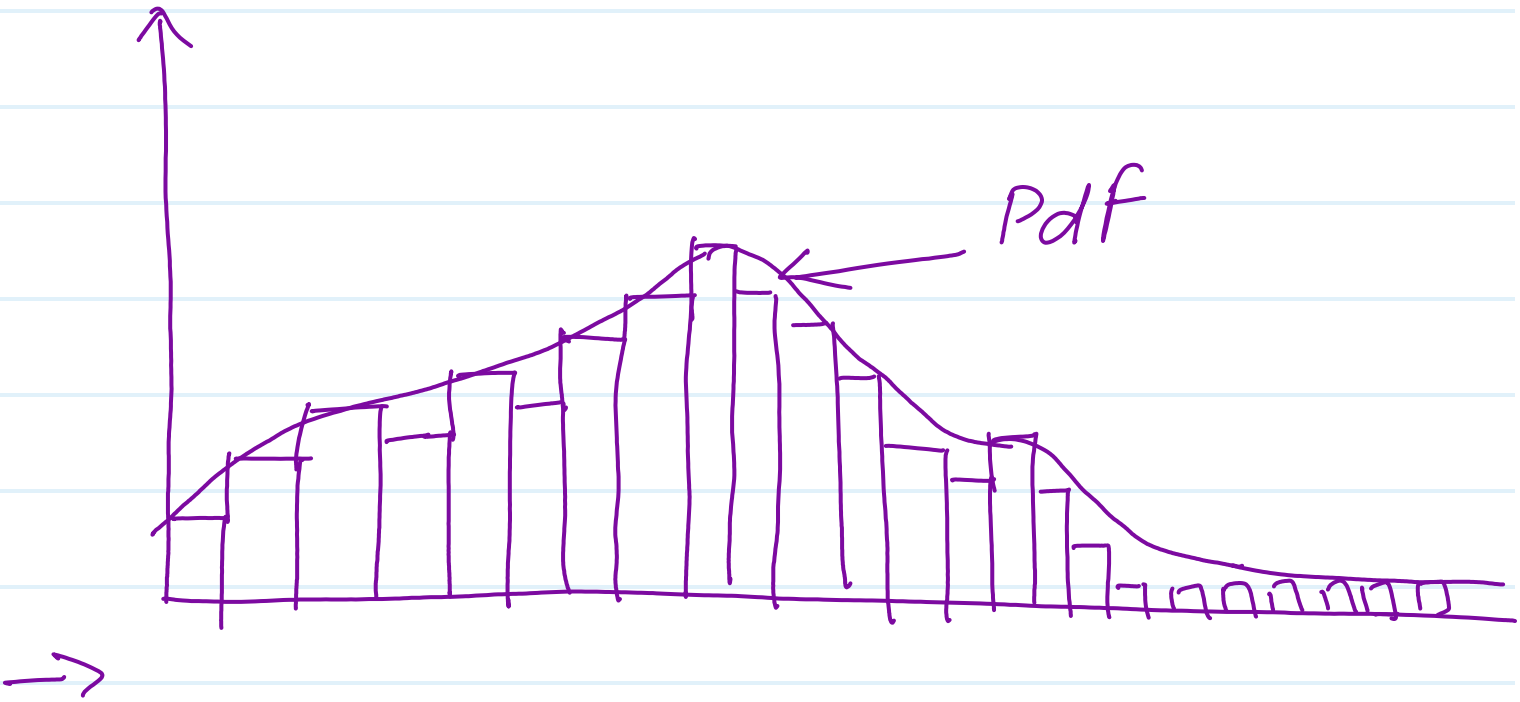


③ mesokurtosis

$k = 3$

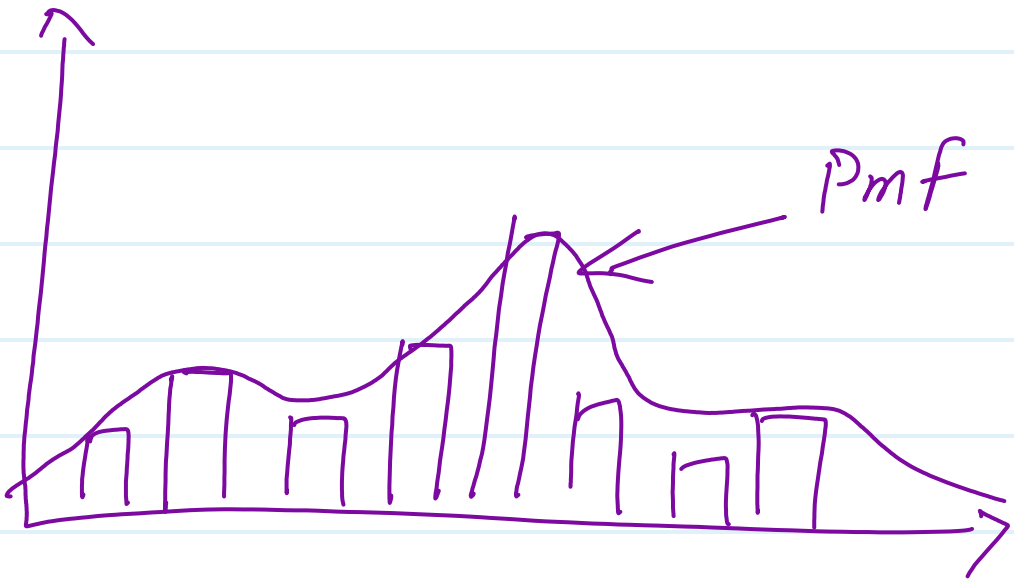


① Histogram



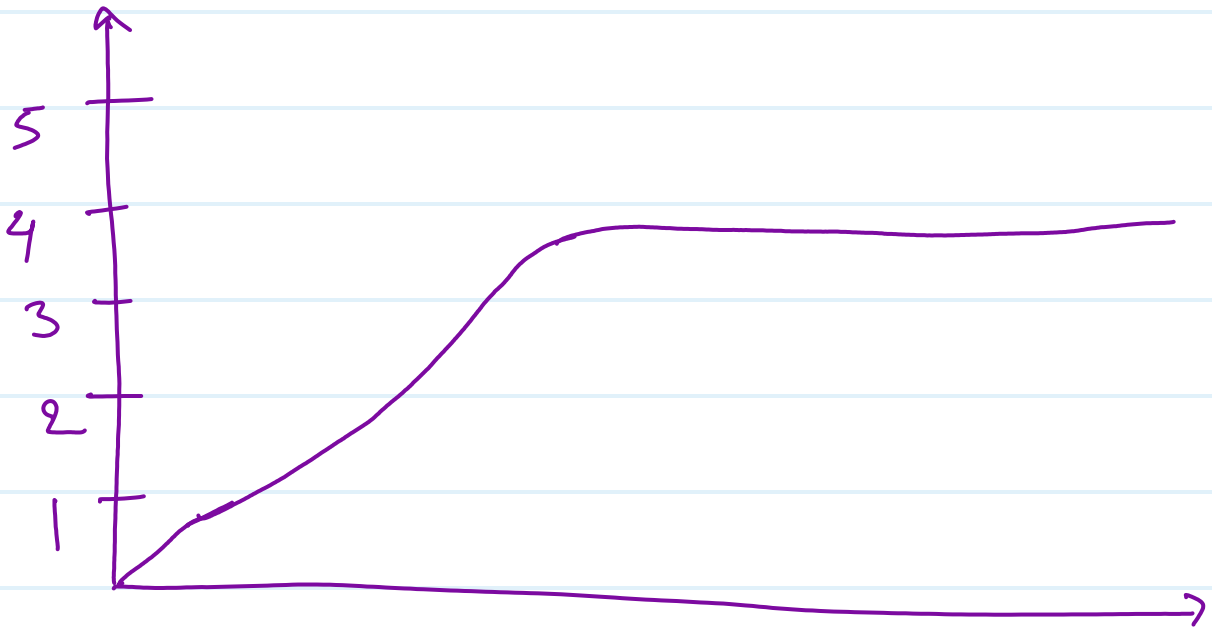
Probability density function

② Density histo



Prob. mass. function.

kde



$1 = 1$

$2 = 1 + 2 = 3$

3

$4 = 3 + 3 = 6$

$5 = 6 + 5 = 11 + 0$

✡ correlation / covariance

x_1 x_2 x_3 x_4 y

x_1 y

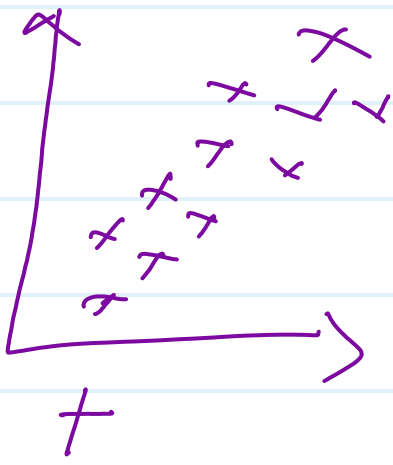
x_2 y

x_3 y

x_4 y

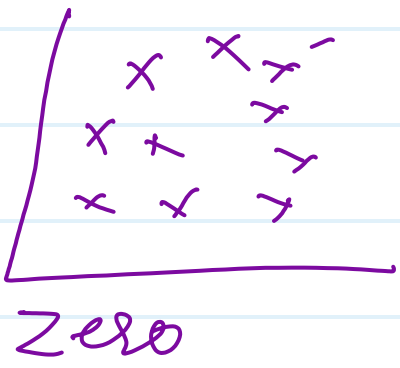
covariance :-

$-\infty$ $+\infty$

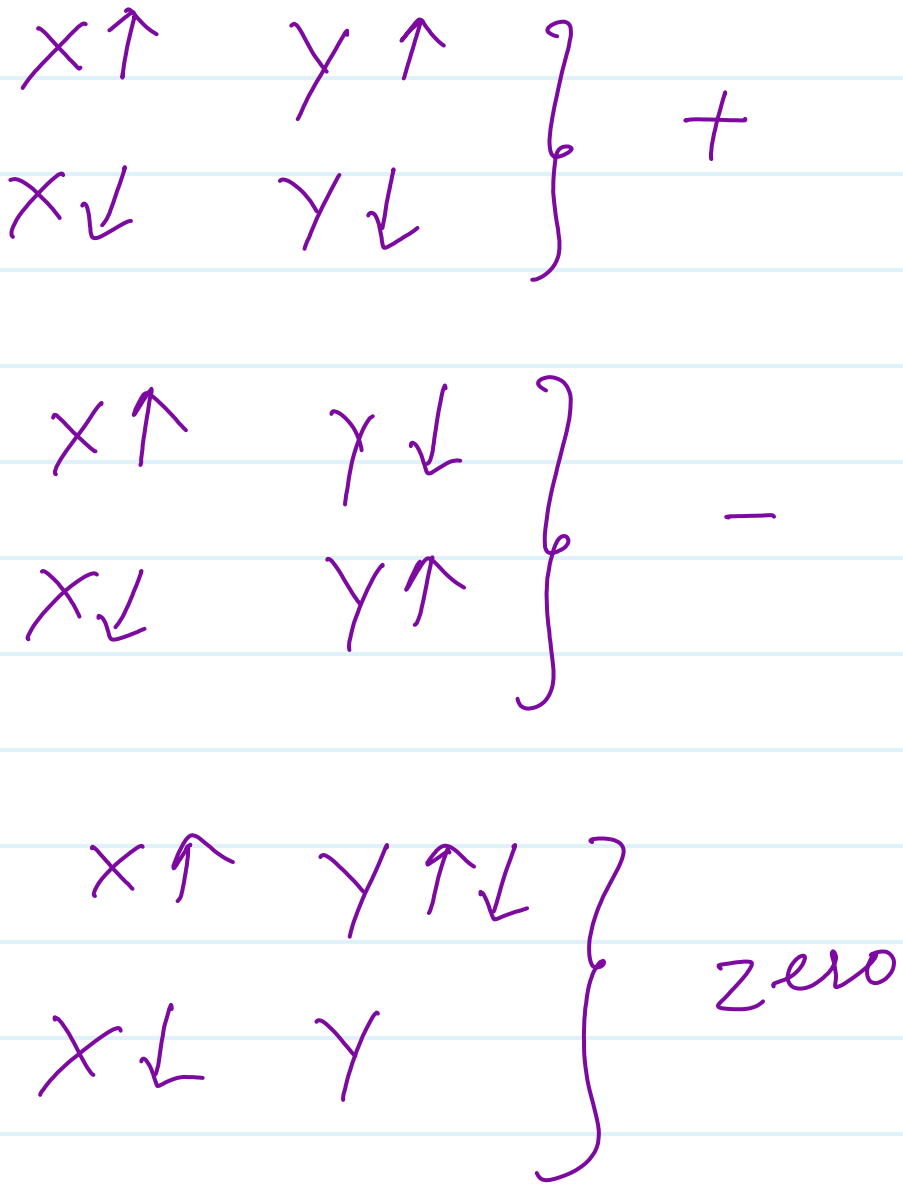


+

−



Zero



correlation

1. Pearson corr. coeff.
2. Spearman rank corr. coeff.

$X \quad Y \quad = \quad 0.85$

$X \quad Y \quad = \quad -0.9$

~~Assumption =~~

$X_1 \; X_2 \; X_3 \; X_4 \mid Y$

$X_1 \quad X_2 \quad = \quad 0.75$

Pareto Dist. / Powerlaw Disd.

Cricket - 11

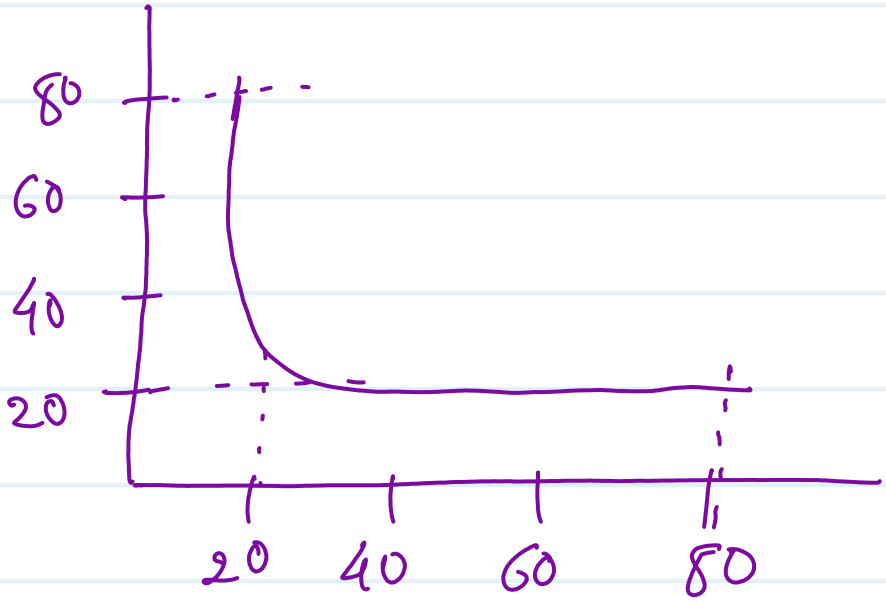
run - 80% — 20%

20% — 80%

Work

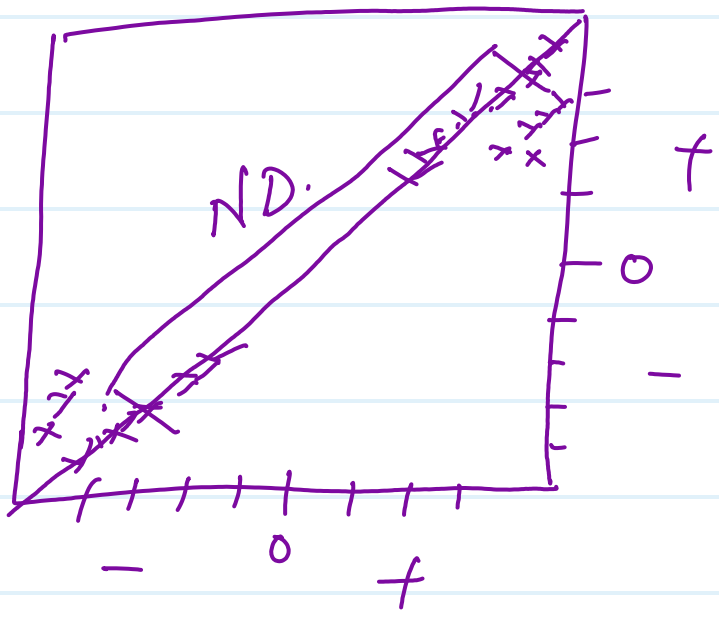
80% — 20%

20% — 80%



Box-Cox transformation

Q-Q



Percentile

[2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 9, 9, 10, 11, 11, 12]

percentile of 10 position!

$$10^{th} = \frac{16}{19} \times 100$$

$$= 84\%$$

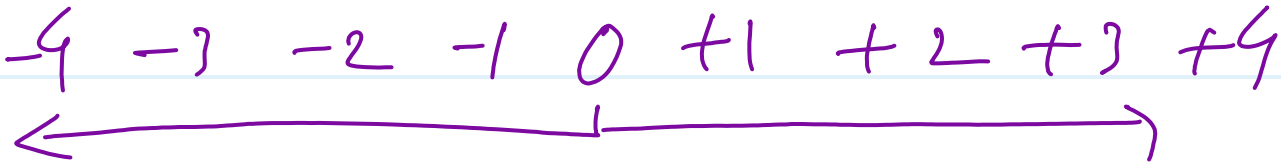
Q_1

Q_3

$$IQR = Q_3 - Q_1$$

Upper fence $= Q_3 + 1.5\ IQR$

lower fence $= Q_1 - 1.5\ IQR$



5 number summary

1. max
2. min
3. Q_1
4. Q_3
5. m/m