① Quartile & Percentile

Q_1 = 25%

median = 50%.

Q_3 = 75%.

min =

max =

Percentile

$$\frac{80 \times 30}{100} \Rightarrow 24$$

[2, 3, 5, 4, 6, 7, 8, 10, 12, 3, 4, 5, 2, 1, 4]

$n = 15$

$$\text{Percentile} = \frac{11}{n+1} \times 100$$

$$= \frac{11}{16} \times 100$$

$$\Rightarrow 68\%.$$

$6^{th} = \frac{5}{16} \times 100$

$\Rightarrow 31.25\%$

Percentile Rank.

$\Rightarrow 40\%$

$\Rightarrow \frac{40}{100} \times n+1$

$= \frac{40}{100} \times 16 \Rightarrow 6.4$ index

7 number

Inter quartile Range

$IQR = Q_3 - Q_1$

Box and whiskes plot

5 number summery

Q_1

Q_3

median

lower limit

upper limit

[2, 3, 7, 4, 5, 3, 6, 7, 8, 9, 11, 10, 4, 3, 1, 4, 30, 89]

$Q_1 = 25\%$

$= \frac{25}{100} \times 19$

$\Rightarrow 4.75$

$Q_3 = \frac{75}{100} \times 19$

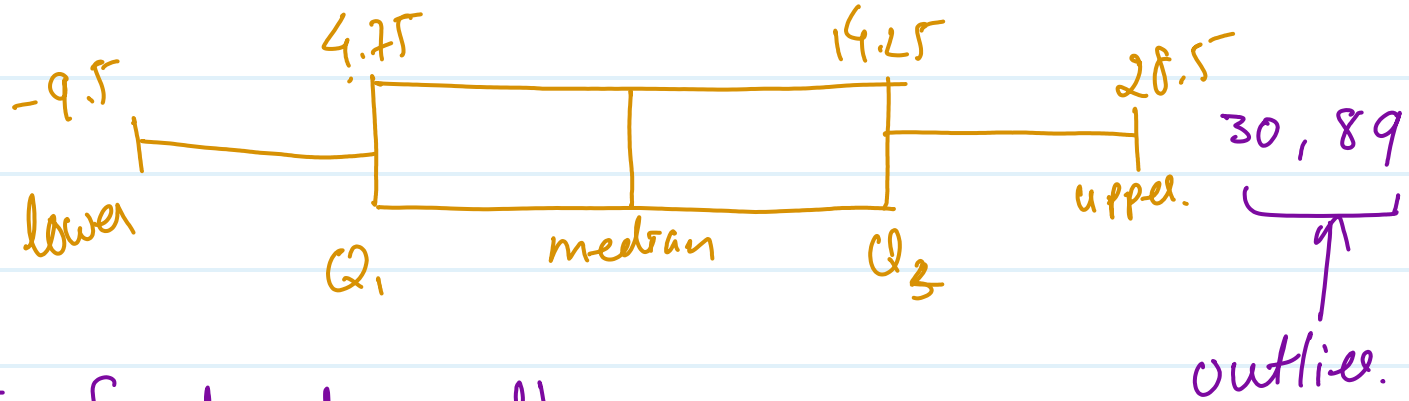
$= 14.25$

$IQR = 14.25 - 4.75$

$= 9.5$

$$\text{lower limit} = Q_1 - 1.5\,IQR$$
$$= 4.75 - 1.5 \times 9.5$$
$$= -9.5$$

$$\text{upper limit} = Q_3 + 1.5\,IQR$$
$$= 14.25 + 1.5 \times 9.5$$
$$= 28.5$$



To find out outlier.

$$[2, 3, 4, 6, 7, 8, \underset{6}{40}] \quad \frac{36}{7}$$

mean = 10

mean = 4.9

mean

median = 6

mode

$$[2, 3, 4, \frac{1}{6}, 6, 7, 8\frac{1}{6}, 40\frac{1}{6}]$$

(1 ↓ 6, 10 ↓ 6)

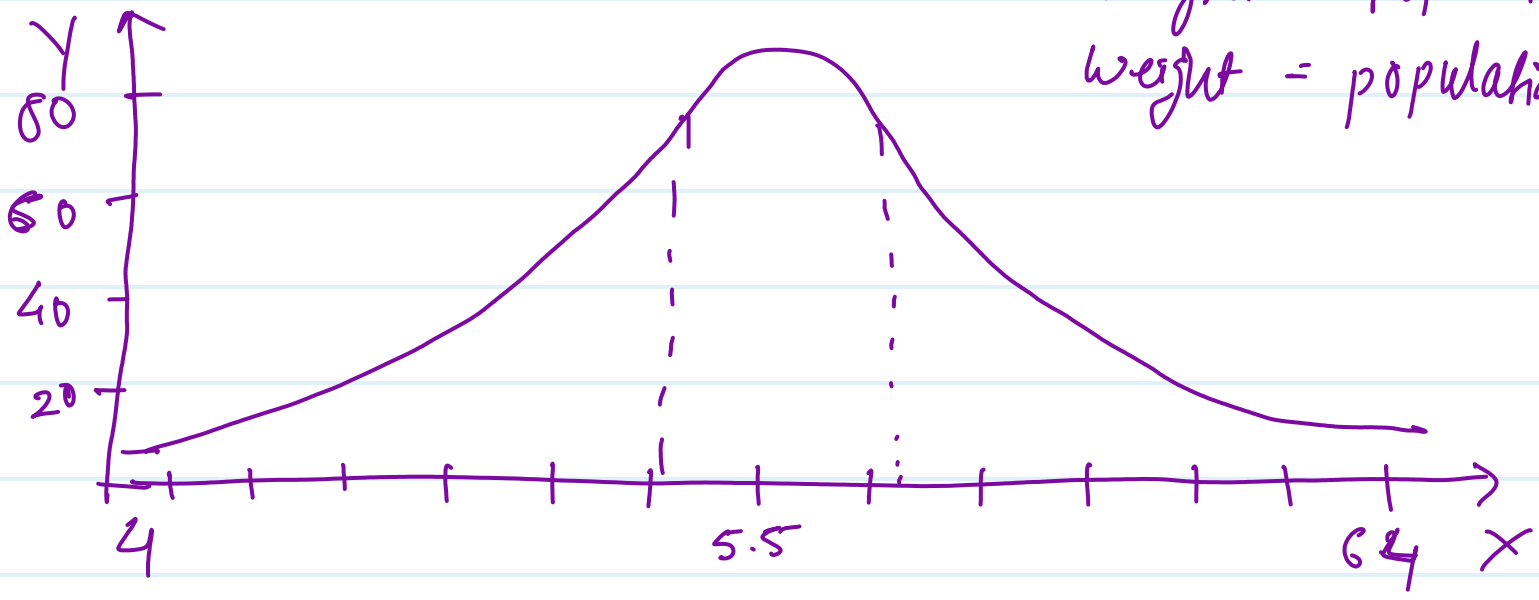
X
P
F
P
–
–
F
–
P
F

mode

missing value handeling

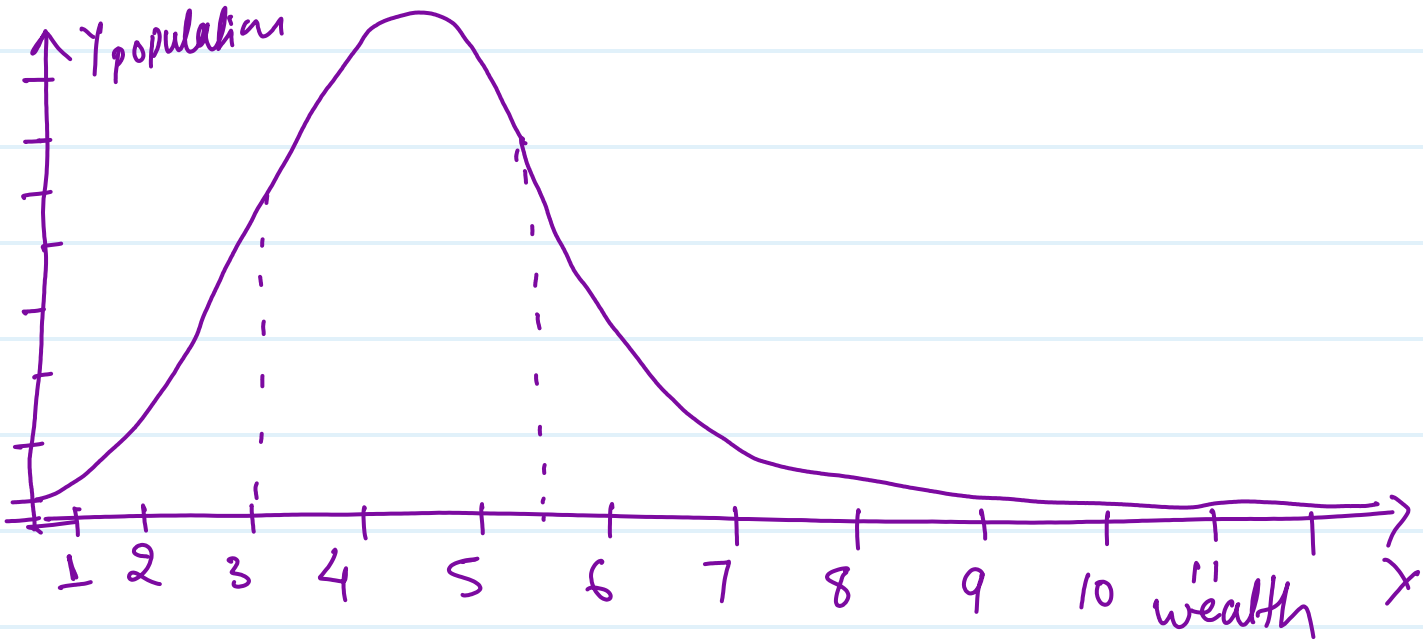
① mean
② median
③ mode
④ Random Imputation

Normal Distribution

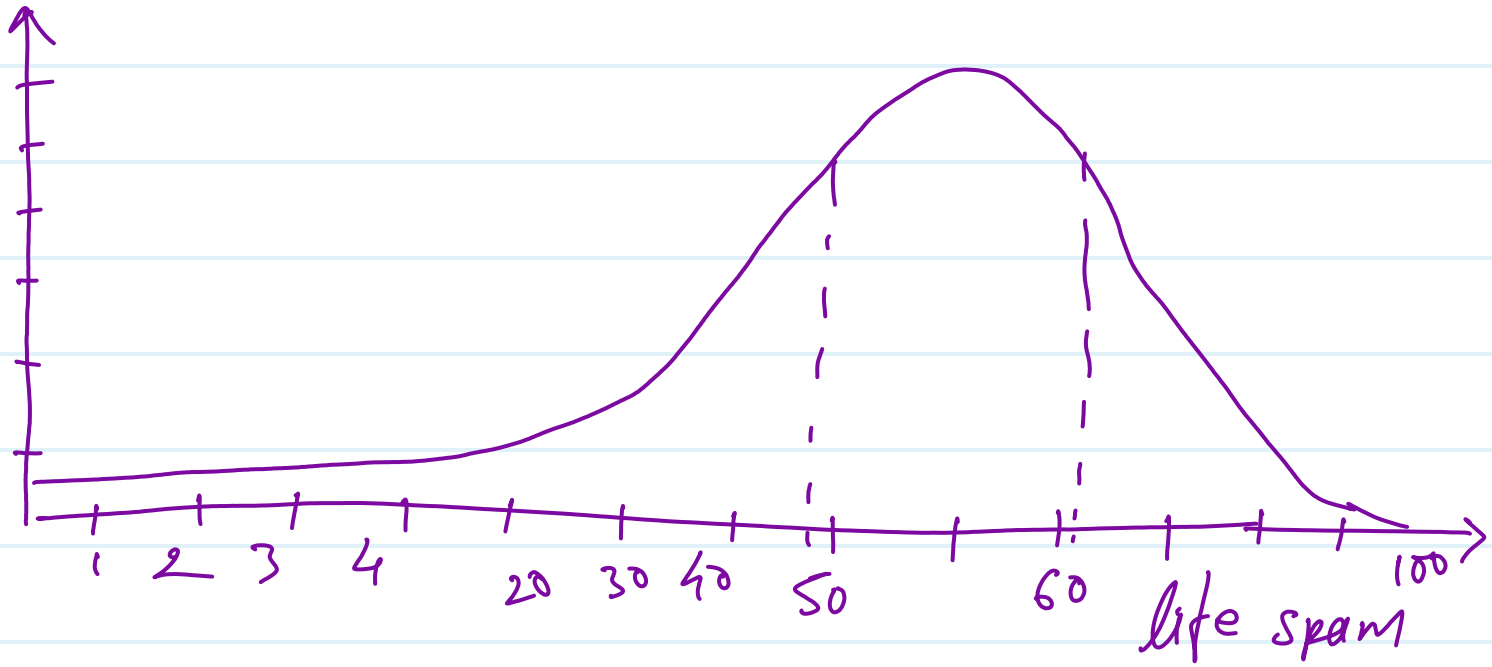


Hight = population
weight = population.

Right skewed Distribution



Left skewed Distribution



Probability Distribution

① Descrete probability

② continous probability

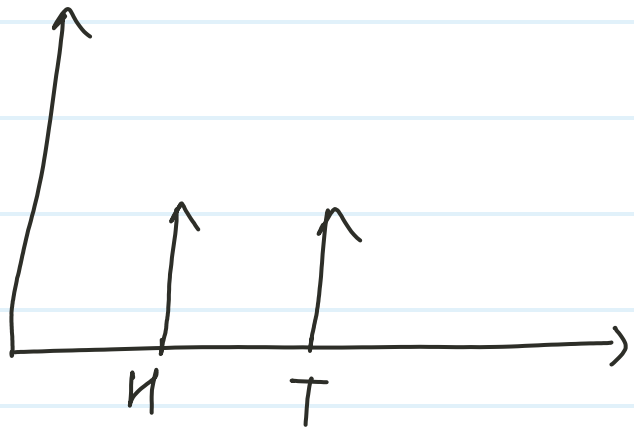
① Descrite prob.

① Bernaulli's Dist.

coin = T/H = $\frac{1}{2}$

experiment = single time

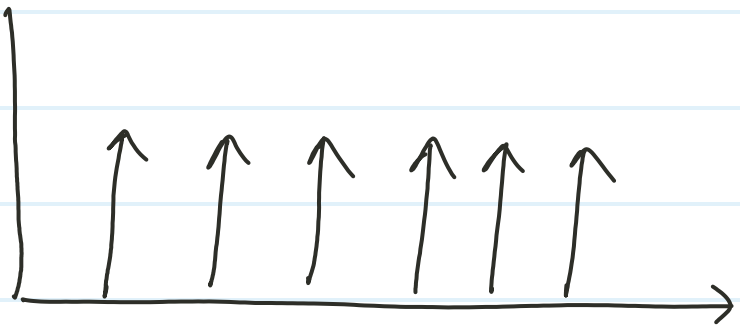
outcome = fix



(1) Binomial Distribution

experiment = multiple time

outcome = fixed

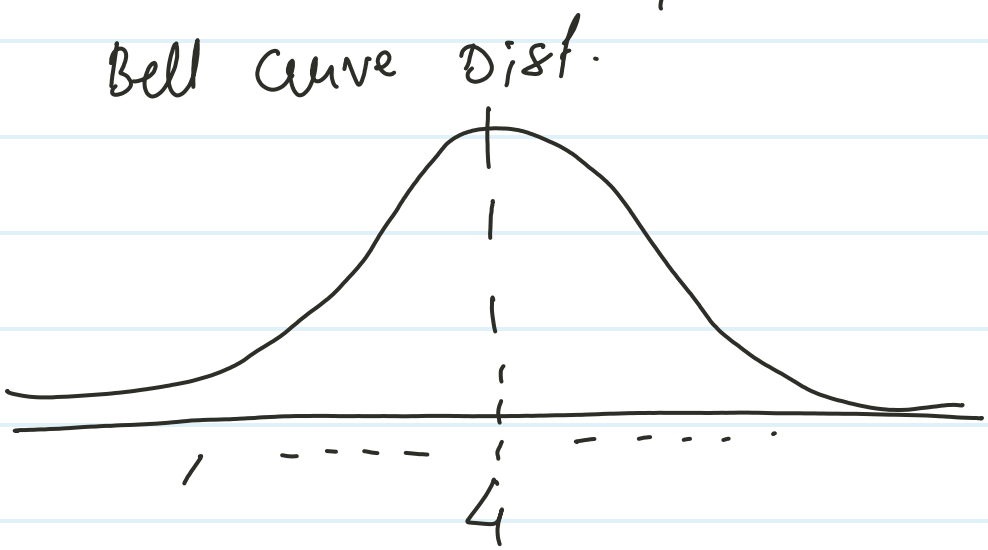


(3) Poission Distribution.

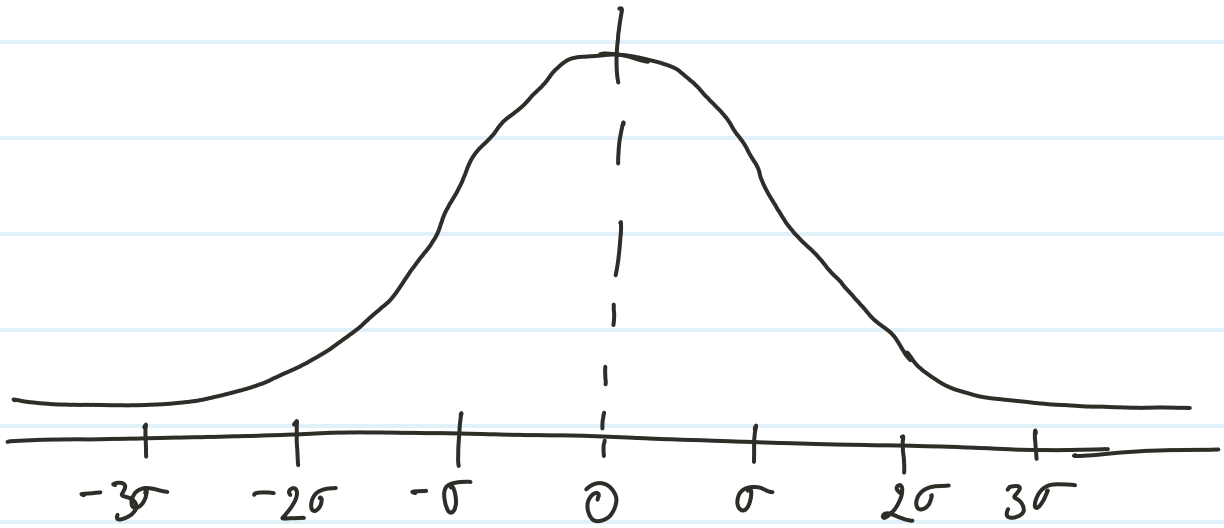
(2) Continous Distribution.

(1) normal Distribution / Gaussian Dist.

Bell curve Dist.



② standard normal Dist.



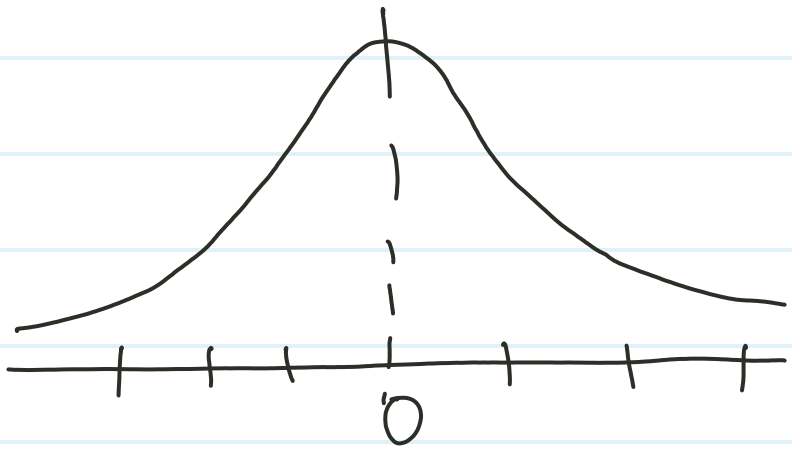
$$Z_{score} = \frac{(X_i - \mu)}{\sigma}$$

[1, 2, 3, 4, 5, 6, 7, 8, 9]

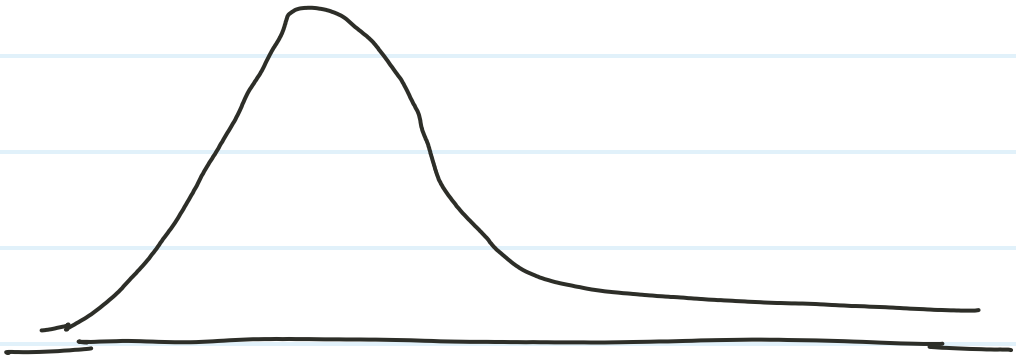
$\sigma = 1$

$\mu = 5$

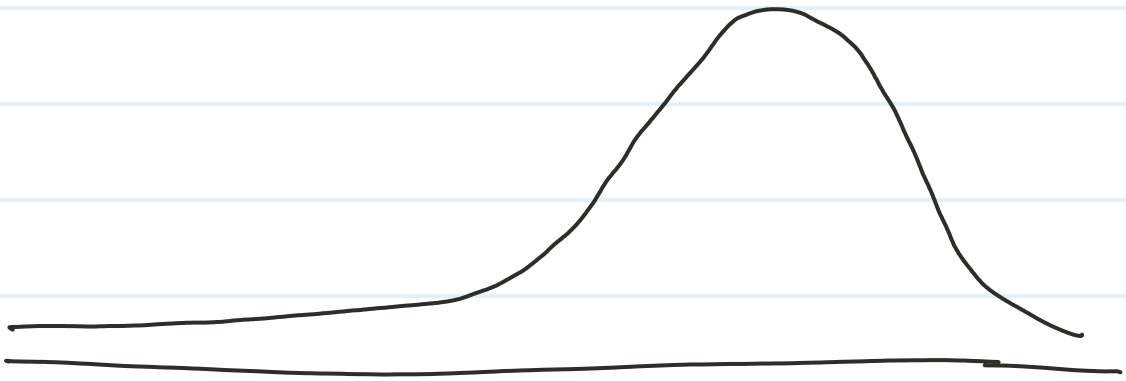
$1 - 5 = -4$
$2 - 5 = -3$
$3 - 5 = -2$
$4 - 5 = -1$
$5 - 5 = 0$
$6 - 5 = 1$
$7 - 5 = 2$
$8 - 5 = 3$
$9 - 5 = 4$



③ Right skewed Dist. / Log normal Dist.
positive Dist.



④ Left skewed Dist. / Negative Dist.



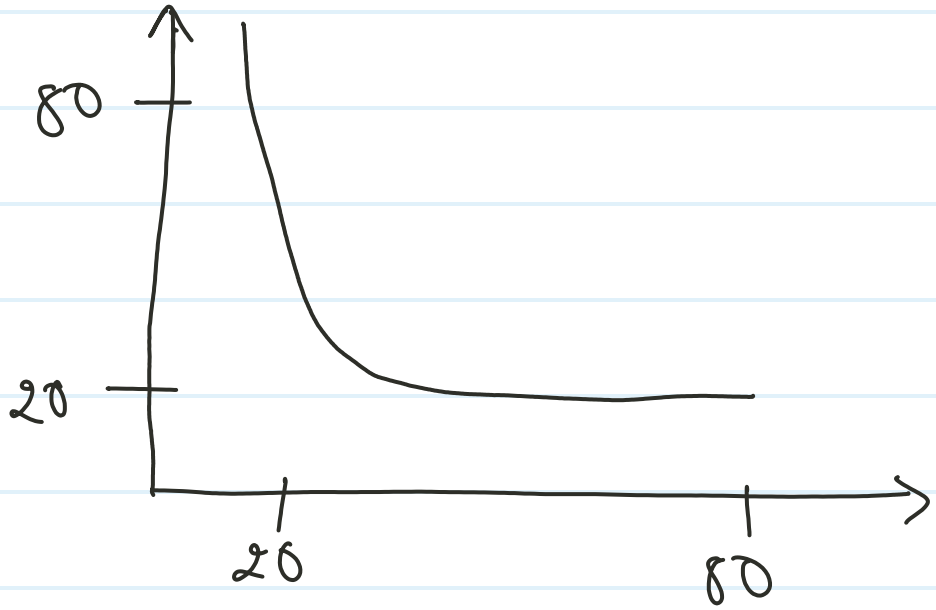
⑤ pareto distribution / Powa law Dist.

20 - 80

→ 80 work. — 20 emp.
20 work — 80 emg

→ 80 Run — 20 batch

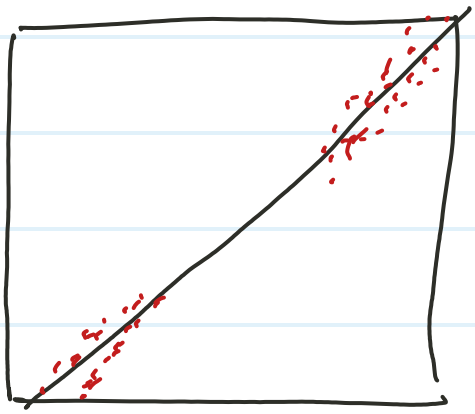
→ 20 Run — 80 batch



① Box - Cox transformation
② log normal trasformation.

Quantile - Quatile plot (Q-Q. plot)

non-Gaussian Distribution



— Gaussian Distribution