Percentiles and Quartiles

Percentages :

1, 2, 3, 4, 5

How many numbers are even ? $\boxed{2}$

Percentage of Even numbers = $\dfrac{\text{\# satisfying the condition}}{\text{Total \# of observation}}$

$= \dfrac{2}{5} = 0.4 \Rightarrow 40\%$

1, 3, 4, 5, 6, 7, 8, 9, 11, 13, 14, 15

Percentage of Prime Number ?? $\dfrac{5}{12} \times 100 = 41.66\%$

Percentile

GATE, CAT, JEE, NEET

A percentile is a value below which certain percentage of observation lie 5.25

Avg 8 $\boxed{45\%}$ <10 Agree? 80%

2, 2, 3, 4, $\boxed{5, 5,}$ 5, 6, 7, 8, 8, 8, 8, 8, 9, 9, 10, 11, 11, 12 CAT

What is the percentile ranking of 10 ? $\boxed{x = 10}$ JIM

$\boxed{99\%}$

Percentile Rank of $x = \dfrac{\text{\# of values below } x}{n} \times 100$

<8 $= \dfrac{16}{20} \times 100 = 80\%$

$= \frac{9}{\cancel{20}_{1}} \times \cancel{100}^{5} = 45\%$

What values exists at the percentile ranking of 25% ??

$$\text{Value} = \frac{\text{Percentile}}{100} \times (n+1)$$

$n=20$

$\underline{5.85}$

$$= \frac{25}{100} \times 21 = 5.25 \rightarrow \text{Index position}$$

$\Downarrow$

$\boxed{5}$

Median

$$75\% = \frac{75}{100} \times 21 = 15.75 \Rightarrow \underline{9}$$

↳ Index position

$$50\% = \frac{50}{100} \times 21 = \underline{10.5} \Rightarrow 8$$

Five Number Summary {Outliers}

1) Minimum
2) First Quartile (Q1) ⇒ 25%
3) Median
4) Third Quartile (Q3) ⇒ 75%
5) Maximum

Inter Quartile Range

IQR

{ outliers Removal }

IQR

Dataset

1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 9, $\boxed{\cancel{27}}$ ✓

$Q3 = \frac{\cancel{75}^{15}}{\cancel{100}_{\cancel{4}_{1}}} \times (\cancel{20}^{1})$

$= 15$ index

$= 7$

[Lower Fence ⟷ Upper Fence]

↑ ↑↑↑   ↑ ↑↑

$$\left.\begin{aligned}\text{Lower Fence} &= Q1 - 1.5\,(IQR)\\ \text{Upper Fence} &= Q3 + 1.5\,(IQR)\end{aligned}\right\}$$

$\frac{25}{100} \times (20) = 5$ index
⇓
3

$IQR = Q3 - Q1 = 75\% - 25\%$
$= 7 - 3 = \underline{\underline{4}} \Rightarrow IQR$
↳ value

$Q1 = 3$

Median =

$Q3 = 7$

Maximum =

Lower fence = $3 - 1.5(4) = 3 - 6 = -3$

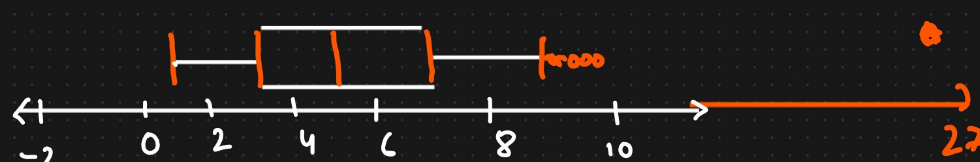
Upper fence = $7 + 1.5(4) = 7 + 6 = 13$

$[-3 \longleftrightarrow 13]$

1, 2, 2, 2, 3, 3, 4, 5 $\boxed{5, 5}$ 6, 6, 6, 6, 7, 8, 8, 9

Minimum : 1
Q1 = 3
Median = 5
Q3 = 7
Maximum : 9

<u>Box plots</u>

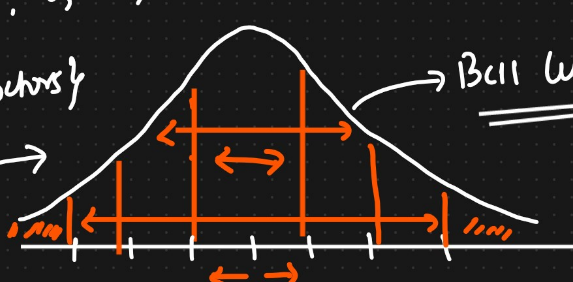


* <u>Gaussian Distribution</u> / <u>Normal Distribution</u> {Smoothening of histogram}

2015 : Kernel density Estimator

$X = \{1, 4, 5, 6, 7, 8, \ldots\}$

Domain knowledge {Doctors}



Bell Curve

3 people

→ 75 people

Weight, Height

{ 68 – 95 – 99.7 }

⇓

Rule

⇓ 68%

Properties { 3 Sigma Limit }

⇓

Empirical Formula.

250+ people

{ 100 }

68

② Standard Normal Distribution

1, 2, 3, 4, 5

$\mu = 3$

$\sigma = 1$

⇓

$\left\{ \boxed{\mu = 0 \quad \sigma = 1} \right.$

$$Z \text{ score} = \frac{x - \mu}{\sigma}$$

$$= \frac{1-3}{1} = -2$$

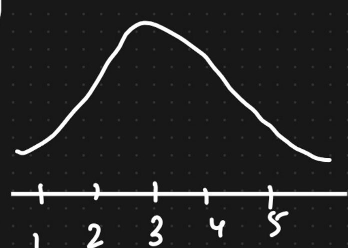
$$\frac{2-3}{1} = -1$$

$$\frac{3-3}{1} = 0$$

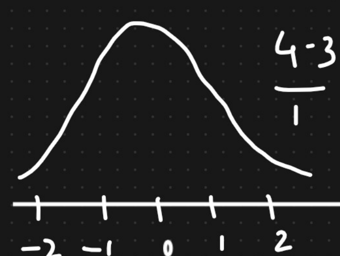
$$\frac{4-3}{1} = 1$$

{1, 2, 3, 4, 5}

{-2, -1, 0, 1, 2}



Z score ⟹



⇓

Normal Distributed Data

(μ, σ)

$\boxed{Z \text{ score}}$ ⟹

$(\mu = 0, \sigma = 1)$

Standard Normal Distributed

Why we need to convert ???

Purpose ?? Reason ??

Machine Learning

Age (yrs)	Weight (kgs)	Distance (Kms)
1	70	120
2	60	200
3	30	300
4	20	400
5	50	500
7	80	600

{ Same Scale }

feature scaling

Normalization

Standardization

ML problem

⇓

Z-Score

$$Z \text{ score} = \frac{x-\mu}{\sigma} \quad \{\text{Standardization}\}$$

$\boxed{\mu=0,\ \sigma=1}$

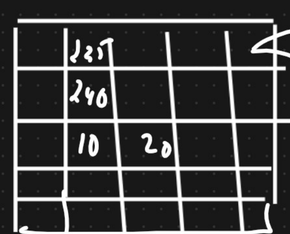
{ Min Max Scaler }

Normalization : Dataset features values ⟹ $\boxed{0 \text{ to } 1}$

{ feature Scaling }

Deep learning

CNN



→ 0-255

→ $\boxed{0-1}$

Z score vs Normalization

→ 0 to 1

0 to 5

0 to 10

$\boxed{\mu=0,\ \sigma=1}$

$$Z\text{score} = \frac{x-\mu}{\sigma} \longrightarrow \text{Standardization}$$

Probability ÷ It is a measure of the likelihood of an event.

What is a probability of coin landing on heads?

$$\text{Probability} = \frac{\text{\# of ways an event can occur}}{\text{\# of possible outcomes}}$$

$$= \frac{1}{2} = 0.5$$

① **Mutual exclusive event** ÷ Two events are mutually exclusive if they cannot occur at the same time

Red King ◇

$\boxed{K\heartsuit}$ → Black

♡

Eg ÷ Toss a coin → {H} {T}

Rolling a dice → {1} {2} {3}

Pulling a card from the deck → Non mutual exclusive.

Additive Rule {Mutual Exclusive}

If I flip a coin, what is the probability of coin landing heads or tail

$Pr(H) = \frac{1}{2}$ $\quad$ $Pr(T) = \frac{1}{2}$

$$P(H \text{ or } T) = Pr(H) + Pr(T)$$

$$= \frac{1}{2} + \frac{1}{2} = 1$$

(x) When picking randomly from a deck of cards, what is the probability of choosing a card that is a heart or a king

$P(\text{Heart}) = \frac{13}{52}$ } $P(\text{King}) = 4/52$ } $P(\text{King and Heart}) = 1/52$

Addition Rule for non mutual exclusive events

$$\overset{K \text{ or } \heartsuit}{P(A \text{ or } B)} = \overset{P(K) + P(\heartsuit) - P(K \text{ and } \heartsuit)}{P(A) + P(B) - P(A \text{ and } B)}$$

$$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = 4/13$$

{ Mutual inclusive }

↓↓↓↓

$P(\text{K or Queen}) = P(K) + Pr(Q)$

$$= \frac{4}{52} + \frac{4}{52} = 8/52 = 2/13$$

Independent Event & Non Independent Events

↓

{ Two if they do not affect one another }

Toss

1	2	3	4
H	T	H	H

→ Dependent Events

↙ 1

3 marble

Bag of marbles { 4 marbles, } {2 R, 1 wh, 1 black }

1 2

Red ← 1

Multiplication Rule for Independent Events:

Rolling a dice

What is the probability of rolling a "5" and then a "3"

$$\boxed{Pr(A \text{ and } B) = Pr(A) * Pr(B)}$$

$$Pr(5 \text{ and } 3) = Pr(5) * Pr(3)$$

$$= \frac{1}{6} * \frac{1}{6} = \boxed{\frac{1}{36}}$$

(*) What is the probability of drawing a King and then drawing a queen from a deck of cards, without putting King back?

⇓

Dependent Event

52

Baye's theorem

$Pr(K) = \frac{4}{52}$

Conditional probability

$Pr(A \text{ and } B)$

⇓

$$Pr(K \text{ and } Q) = \boxed{P(K) * Pr(Q/K)}$$

$Pr(Q) = \frac{4}{51}$

$$= \frac{4}{52} * \frac{4}{51}$$

Naive Bayes