Probability continuation & Descriptive stats.

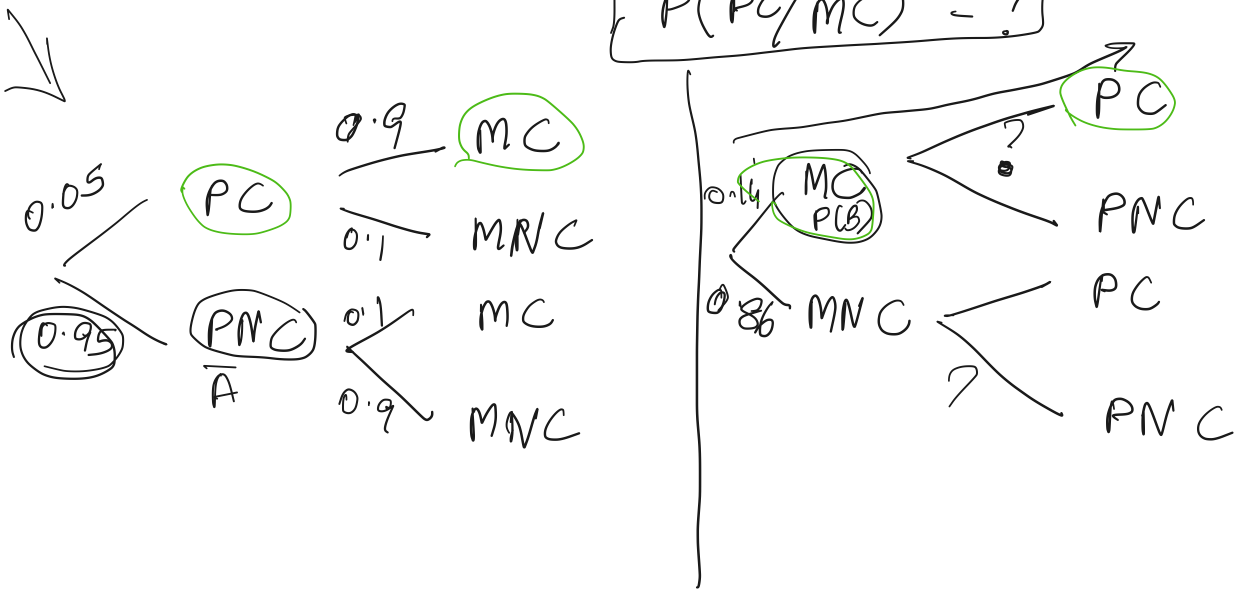
Conditional prob $P(A/B) = \frac{P(A \cap B)}{P(B)}$ ←

Bayes Theorem:

Event A: P.C.
Event $\bar{A}$: PNC
Event B = MC
Event $\bar{B}$ = MNC

$P(MC/PC) = P(B/A) = 0.9$
$P(MC/PNC) = P(B/\bar{A}) = 0.1$
$P(PC) = P(A) = 0.05$

$P(PC/MC) = ?$



A and $\bar{A}$ are M.E.

$P(MC) = P(B) = P(B \cap \bar{A}) + P(B \cap A)$

$\rightarrow P(B) = P(\bar{A}) * P(B/\bar{A}) + P(A) * P(B/A)$ ←

$P(B/A) = 0.9 \quad ; \quad P(A) = 0.05$

$P(B/A) = 0.1$; $P(A) = 0.95$.

$$\Rightarrow P(B) = (0.95) * (0.1) + (0.05) * (0.9)$$

$$P(B) = 0.14$$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A/B) * P(B)$$

$$P(B/A) = \frac{P(B \cap A)}{P(A)} \Rightarrow P(B \cap A) = P(B/A) * P(A)$$

$$\Rightarrow P(A \cap B) = P(B \cap A).$$

$$P(A/B) * P(B) = P(B/A) * P(A).$$

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$

$P(A/B)$: posterior. $P(B/A)$: likelihood. $P(A)$: prior. $P(B)$: evidence

$$= \frac{0.9 * 0.05}{0.14} = 0.32$$

$P(\text{P cheating} / \text{M P.C}) = 0.32.$

$P(\text{spam} / \text{Rolex}) =$

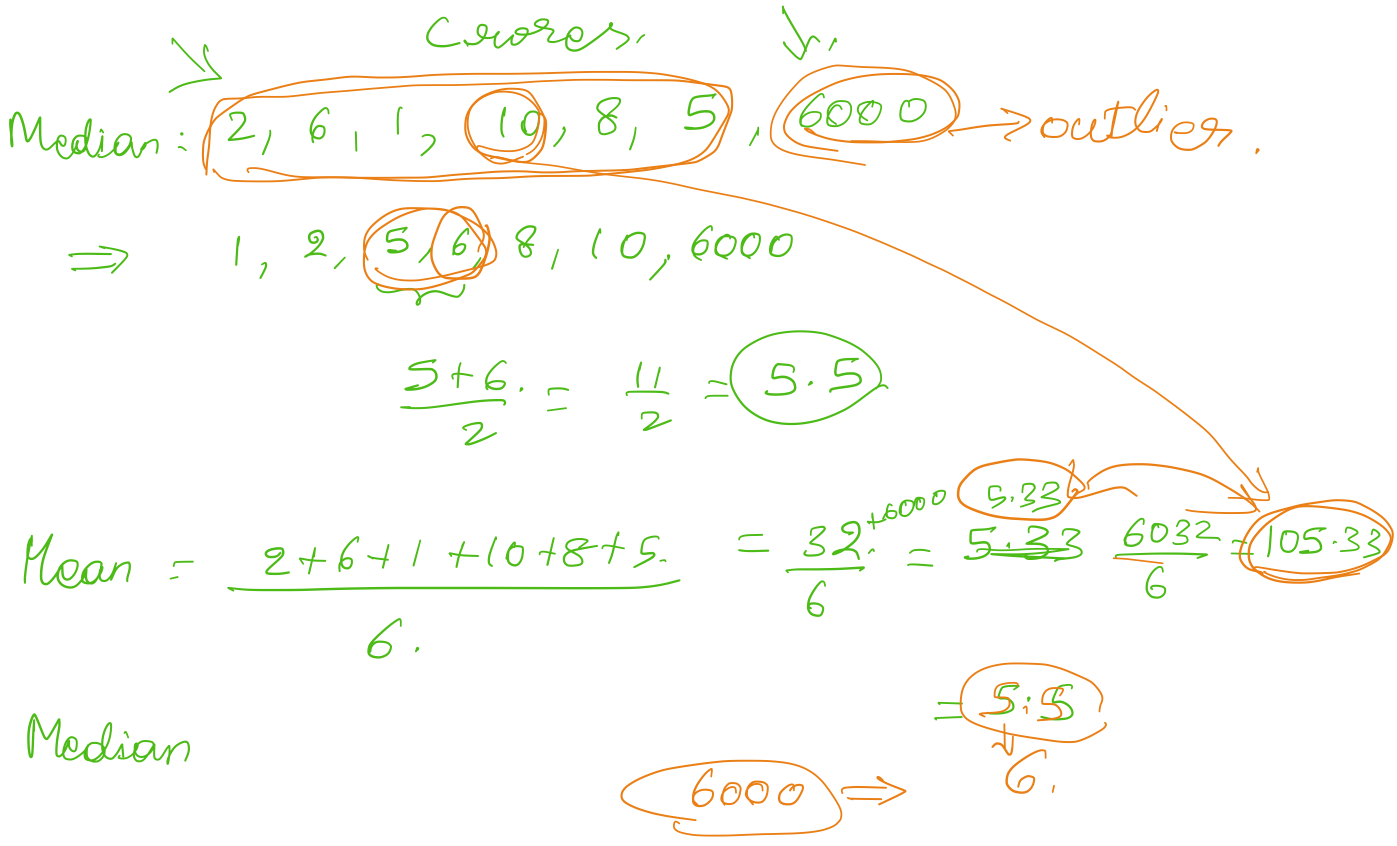
Descriptive stat.

Random variable

Central tendency.

5 number rule.

Univariate analysis.



Mean, Median, Mode.

Mean Absolute Deviation

$\sum_{i=1}^{n} \frac{(x_i - \bar{x})}{n}$ $\bar{x}$ ⇒ sample mean.

X ⇒ pop. mean.

2, 4, 6, 1 ⇒ $\frac{13}{4} = 3.25$

3.25 + 1.75 − 1.75

1, 2, 4, 6.

$$\frac{|2-3.25| + |4-3.25| + |6-3.25| + |1-3.25|}{4}$$

$$\frac{(1.25) + 0.75 + 2.75 + 2.25}{4} = \frac{7}{4}$$

MAD = 1.75

Median

$$\sum_{i=1}^{n} \frac{|x_i - m|}{n}$$ m − median.

Variance.

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

2, 4, 6, 100

$\frac{112}{4} = 28$

$(1.25)^2 + (0.75)^2 + (2.75)^2 + (100-28)^2$ ← 28. →

4.

$(2-38)^2 + (4-38)^2 + (6-38)^2 + (100-38)^2$.

Variance = $(26)^2 + (24)^2 + (32)^2 + (72)^2$

72

S.D. = $\sqrt{\text{Variance}}$

S.D. Mean.

2, 1, 4, 6, 8. = $\frac{21}{5} = 4.2 \Rightarrow$ Mean.

$\Rightarrow$ 1, 2, 4, 6, 8

↓ median.

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

$\Rightarrow$ MSE.
RMSE.
MAE.

2, 1, 4, 6, 8 $\Rightarrow$ Mean $\Rightarrow$ 4.2.

S.D. = $\sqrt{\text{Variance}}$ =

Actual Pred.

5 ⟷ 4 : 1

6 ⟷ 7.6

7 ⟷ 5 : 4

$(10)^2$

Mean absolute Deviation

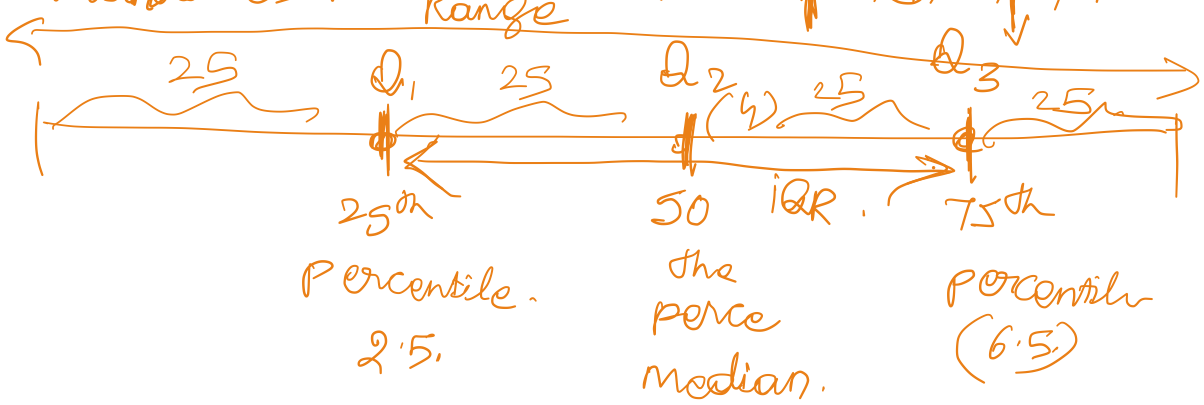
Median " "

Percentile.

2, 4, 8, 1, 9, 3, 10.

⇒ 1, 2, 3, 4, 5, 8, 9.

Quartiles:

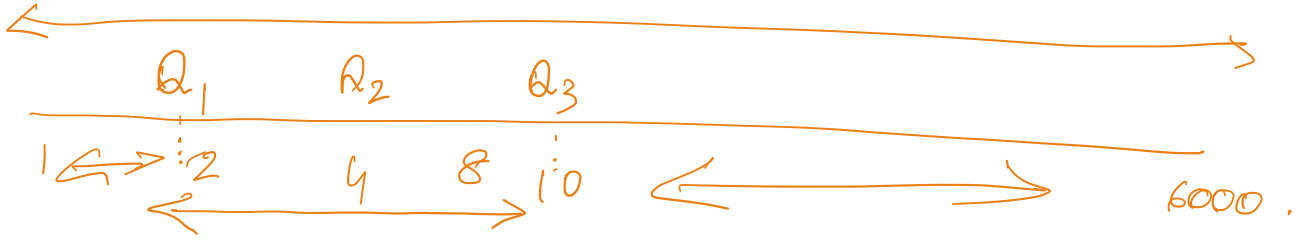


IQR → interquartile range.

Range = Max − Min ⇒ 10 − 1

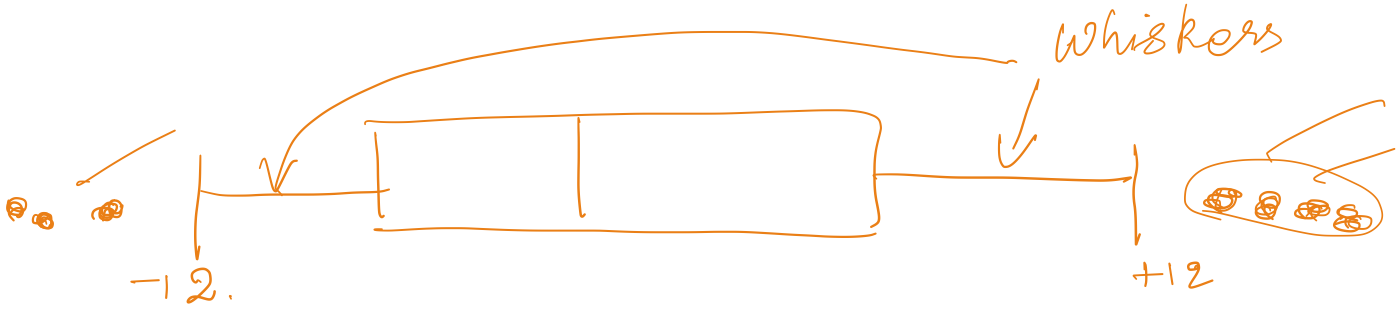
⇒ IQR = $Q_3 - Q$ ⇒ 50%

→ 2, 4, 1, 8, 6000.



Range = 6000 − 1 = 5999.

IQR = 10 − 2 = 8 = 1.5 × 8 = 12.



5-number ⇒ Min, Q_1, Q_2 (Median), Q_3, Max.

1.5 * IQR ⇒ outlier.

⇒ ±12.