Interview question

① measure of central tendency.

mean — average

meadian = middle value

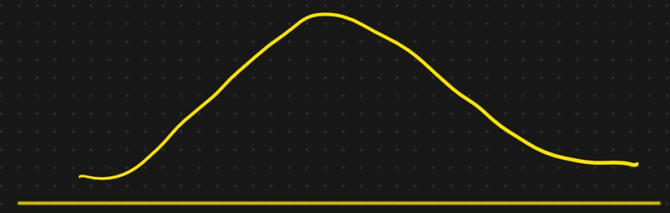
mode — freq. → 10, 20

Data = [10, 20, 30, 5, 6, 10, 20, 25]

Data Asc = 5, 6, 10, 10, 20, 20, 25, 30

$$\frac{10+20}{2} = 15$$

② (i) Right skewed → mean > median > mode.



(ii) Left " → mode > median > mean.

(iii) mean = median = mode.

③ measure of dispersion :-

① Range

② Variance

③ Standard deviation 1σ, 2σ, 3σ

④ Interquartile Range (IQR)

$$IQR = Q3 - Q1$$

Q3 → 75%

Q1 → 25%

④ How SD is different from Variance ?

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

$$\sigma = \sqrt{\sigma^2}$$

⑤ high SD (σ) what does this indicate ?

[1, 2, 3, 4] [10, 50, 40, 60, 80]

⑥ How outliers affect the mean & median ?

Mean :—

⑦ What is the IQR :

(1, 2, 3, 4) 5, 6, [7, 8] (100)

IQR → Q3 - Q1 — <Q1 — Q3> —

Q1 → $\frac{2+3}{2}$ = 2.5 → 3

Q3 → $\frac{7+8}{2}$ = 7.5

IQR = Q3 - Q1

= 7 - 3 = 4

Q3 + 1.5 IQR

= 7 + 1.5 × 4

= 13

Q1 − 1.5 IQR

= 3 − 6 = −3

$(-3, 13)$

(8)

(9) Five - Number Summary है :-

① Minimum

② Q1 (25th)

③ Median (50th)

④ Q3 (75th)

⑤ Maximum

(10) 75000 → mean
∪ → Right skewed data.
50000 → median

(11) EDA :-

① understant the data
load the dataset.
data type check
Structure of the data
missing value.

② Descriptive stats operation :-
mean
median
mode.
SD
Variance

IQR

④ Handle missing data

① Removing

② imputation.

⑭ beteet and Handle outlines.

boxplot

IQR

Z-Score.

⑤ Data Visualization:

⑥ Feature Engineering

⑦ Multicollinearity

(8) Check the distribution.

(12) diff between univariate, bivariate, multivariate.

$$Y = mx + c$$

$$Y_1 = m_1 x_1 + m_2 x_2 + c_1$$

(13) how to handle missing data during EDA?

df.isnull.sum

(14) Correlation & Causation

(15) Sales dataset :-

(16) What is a prob of getting at least one head when flipping one coin. (flipping two fair coins)

[HH, HT, TH, TT] $\left(\frac{3}{4}\right)$ =

P(at least one head) = 1 − P(no head)

= 1 − $\frac{1}{4}$ = $\frac{3}{4}$ = .75

(17) Mutually Exclusive & independent events.

(18) What is a probability of drawing Ace from a pack of card (52)

$$\frac{4}{52} \approx 0.076$$

(19) Bayes theorem :

$$P(A|B) = \left(\frac{P(B|A) \cdot P(A)}{P(B)} \right)$$

$P(A|B)$ = Prob of event A happening given that B has already happened

$P(B|A)$ = event B happening give that A is true.

$P(A)$ = Prob of event A (Prior Prob)

$P(B)$ = Prob of event (B) (total Probability of evidence)

(Q) 1 in every 1000 people are getting affected,

A medical test can detect the corona correctly by 99%

What is a Probability that i will be tested positive

$P(\text{Corona}) = \frac{1}{1000} = 0.001$

$P(\text{Positive} | \text{Corona}) = 0.99$

$P(\text{No Corona}) = 1 - 0.001 = 0.999$

$P(\text{Positive} | \text{Corona}) =$

A medical test is conducted for corona shows 99% accuracy. and for 1% it gives a false result. only 1 in 1000 can get affected by corona. give this situation what is a probblity that person actually have corona if test result shows that its positive

$$P(\text{Corona} \mid \text{Positive}) = \frac{P(\text{Positive} \mid \text{Corona}) \cdot P(\text{Corona})}{P(\text{Positive})}$$

$$= \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.01 \times 0.999}$$

$$\approx 0.091$$

$$P(\text{Positive}) = P(\text{Positive} \mid \text{Corona}) \times P(\text{Corona}) + P(\text{Positive} \mid \text{No Corona}) \times P(\text{No Corona})$$

$$= 0.99 \times 0.001 + 0.01 \times \left(1 - \frac{1}{1000}\right) \approx 0.999$$

$= (0.55 \times 0.001 + 0.01 \times 0.555$

(9.1)

(26) 80% of its product in Plant (A) & 20% in Plant (B).

The defect Rate is (5.1) & (10.1) in respective plant

If a Product is defective - what is the chance that it been manufactured in Plant A.

$$P(A \mid \text{Defective}) = \frac{P(\text{Defective} \mid A) \cdot P(A)}{P(\text{Defective})} \quad (.80)$$

$= 0.05$

$$P(D) = 0.05 \times 0.80 + 0.10 \times 0.20$$

$$= \frac{0.05 \times 0.80}{0.05 \times 0.80 + 0.10 \times 0.25} \approx 0.667$$

(28) If two dice are rolled, what is a probability that their sum is at least 10?

6 × 6 = 36

(4,6), (5,5), (5,6), (6,4), (6,5), (6,6)

$$P(\text{Sum} \geq 10) = \frac{6}{36} \quad \left(\frac{1}{6}\right)$$

29

$\left(\frac{5}{8}\right)$

(30) What do you understand by 68-95-99.7 rule in normal distribution.

$\mu \pm \sigma$

$\mu \pm 2\sigma$

$\mu \pm 3\sigma$

(31) Call center receives (10) calles in an hour of time, on an average, what is a probablity that exactly 5 calles will arrive in one hour.

$$P(K) = \frac{e^{-\lambda} \lambda^K}{K!} = \frac{e^{-10} \; 10^5}{5!}$$

$\lambda = 10$

$K = 5$

$e = 2.718$

$P(5) = 0.037$

$= 3.78\%$

(32) If coin is flipped 100 times, what is a probability of getting exactly 50 Head.

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \frac{100!}{50!(100-50)!} = \left(\frac{100!}{50! \cdot 50!}\right)$$

$n = 100$

$k = 50$

$p = (0.5)$

$$= \binom{100}{50} (0.5)^{50} (1-0.5)^{50}$$

$$= \frac{100!}{50! \cdot 50!} (0.5)^{50} \cdot (0.5)^{50}$$

$$= 0.0795$$

$$= 7.95\%$$