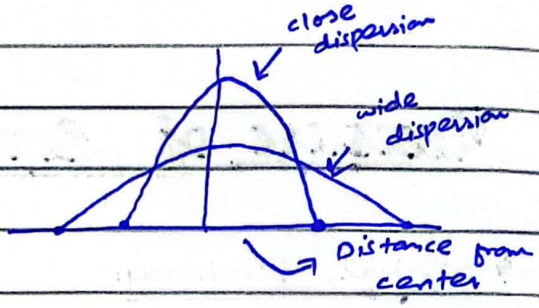
Lecture : 06

Measure of Dispersion

→ Dispersion refers to how spread out the data is.

→ It shows how far values lie from the central value (mean or median)



→ Assume median as a measure of central tendency

→ Dispersion from median

Quantiles

→ Q_2 → Divide the data into two halves (median)

→ Quantile is a general form of dividing a dataset into equal size interval

→ Every Q-Quantile have (Q-1) numbers and it divide data into Q: parts.

Quartiles (specific type of quantile where data is divided into 4 parts)

Q_1 → Represents the 25th percentile

Q_2 → Represents the 50th percentile (the median)

Q_3 → Represents the ~~75~~ 75th percentile

→ **Inter Quartile Range (IQR)** spread of data distribution

$IQR = Q_3 - Q_1$

spread of the middle 50% of observation

Small IQR: tightly clustered middle values

Large IQR: widely spread middle values

→ Five Number summary

Min → minimum

Max → maximum

Q_1 → First Quartile

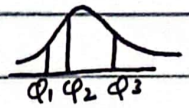
Q_2 → Second Quartile

Q_3 → Third Quartile

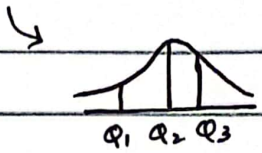
→ shape of data distribution

Symmetric: $Q_3 - Q_2 = Q_2 - Q_1$

positive skewed: $Q_3 - Q_2 > Q_2 - Q_1$



Negetive skewed: $Q_2 - Q_1 > Q_3 - Q_2$



→ Outliers

outliers are the extreme values outside the normal range

① calculate IQR

$IQR = Q_3 - Q_1$

② Lower Bound

$Q_1 - 1.5 \times IQR$

③ upper Bound

$Q_3 + 1.5 \times IQR$

④ Identify outliers

values beyond these are outliers

→ Box and whisker plot