3.3 ** Measure of Distribution shape

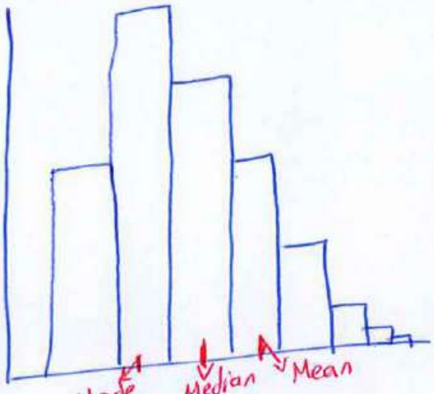
** Relative Location

** Detecting Outliers

Measure of Distribution Shape

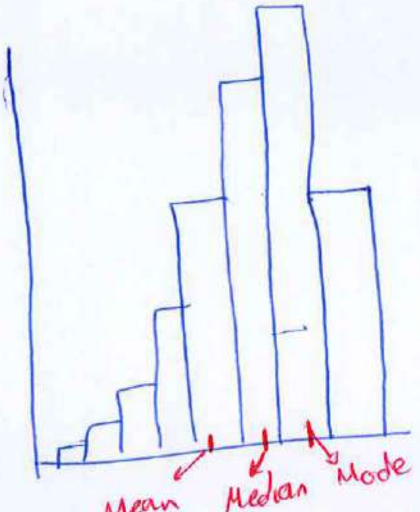
Skewness :- is a measure of Distribution Shape

Recall :- Main



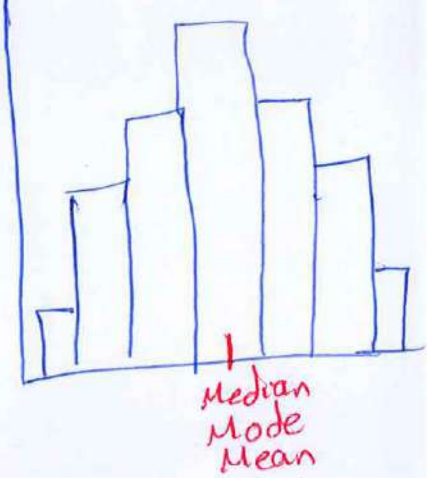
Right Skewness

Skewness > 0



Negative Skewness

Skewness < 0



- Normal Dist
- Bell shaped
- Symmetric
- Skewness = 0

* when data are highly Skewed Median is better than Mean

Bell-Shaped (Symmetric) mean = Median = Mode (at center)

Right Skewness Mean > Median

left Skewness Mean < Median

** Relative Location :- (The Standarized Value)

A Z-score specifies the location of each x-value within a distribution

Given a sample $x_1, x_2, \_\_\_\_\_ x_n$ with Mean = $\bar{x}$ and standard deviation = S

** each value x_i has Zscore Z_i

$$\boxed{Z_i = Z_{x_i} = \frac{x_i - \bar{x}}{S}}$$

* Z score : How many standard deviation the data Value X_i is above or below the Mean

Exp:- A sample of grades has $\bar{x} = 60$ and $S = 10$

(1) Find the Z-score of the grade 70

Sol:- $Z(70) = \frac{70-60}{10} = 1$

→ The value 70 is greater (above) the mean by one ~~std~~ standard deviation

(2) Find the Z-score of the grade 40

Sol. $Z(40) = \frac{40-60}{10} = -2$

The value 40 is smaller (below) the Mean by two standard deviation

③ Find the Z-score of the grade 60

Sol. $Z(60) = \dfrac{60-60}{10} = 0$

Note:-

$x_i > \bar{x} \longrightarrow Z_i > 0$

$x_i < \bar{x} \longrightarrow Z_i < 0$

$x_i = \bar{x} \longrightarrow Z_i = 0$

Note

$Z_i = \dfrac{x_i - \bar{x}}{S}$

$x_i = Z_i S + \bar{x}$

Exp: If Mean = 44.81 and ~~Stad~~ Standard deviation = 5.41

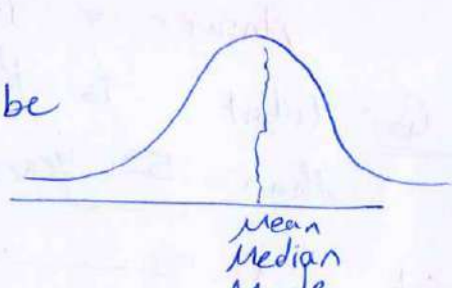
Find the data value whose Z score is −2.01

Sol: $x_i = Z_i S + \bar{x} = 33.94$

** Empirical Rule :- used only for Bell shaped

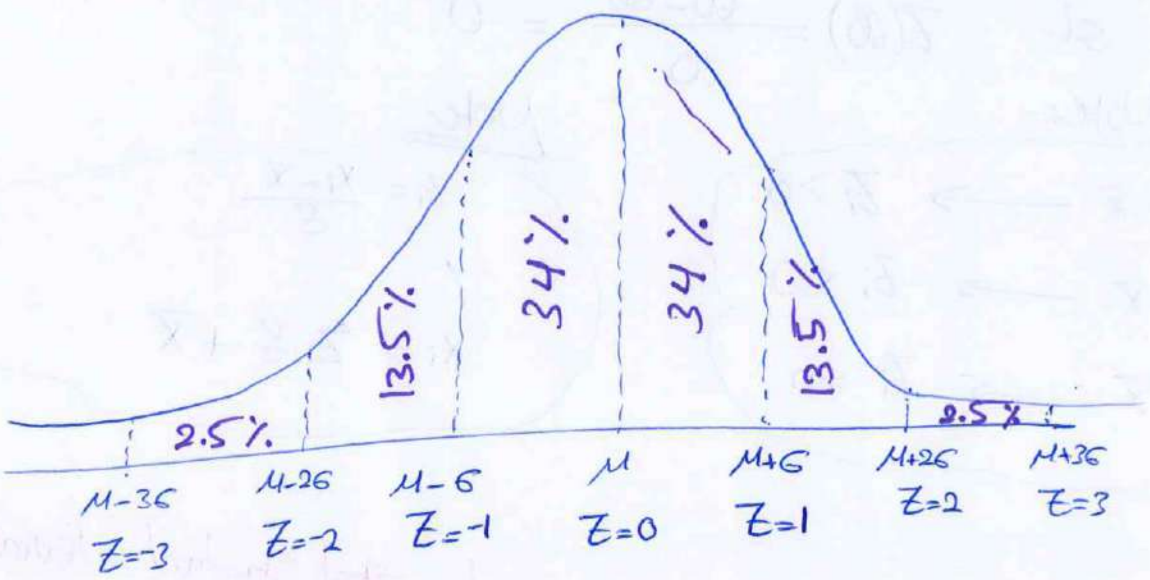
•• Assume a group of data has a bell-shaped distribution with Mean μ and std deviation = σ

① About 68% of the data values will be within one standard deviation of the Mean



② About 95% of the data values will be within two standard deviation of the Mean

③ Almost all data will be within three std. deviation of the Mean.



Exp: Assume the age of employees in BZU have a bell-Shaped distribution with $\mu = 44$ and $\sigma = 8$

Q1 What is the percentage of employee with ages between 28 and 68 years:

Sol: $X = 28 \longrightarrow Z = -2$

$X = 68 \longrightarrow Z = 3$

Answer = 97.5%

Q2. What is the percentage of employees with age less than 52 years

Sol. $X = 52 \longrightarrow Z = 1$

Answer = 84%.

Q3 what is the percentage of employees between 36 and 60 years

$X = 36 \longrightarrow Z = -1$

$X = 60 \longrightarrow Z = 2$

Answer = 81.5 %

* Detecting Outliers:

If the data has a bell-shaped distribution, we can use the empirical role to detect outliers

How?!

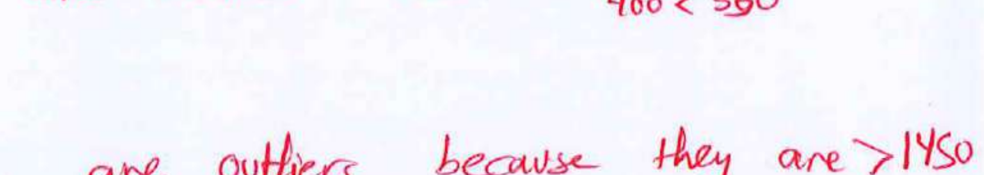
Any value $> \mu + 3\sigma$ or any value $< \mu - 3\sigma$ is an outliers

Any value has $Z > 3$ or $Z < -3$ is an outliers

Exp: Given a sample of salaries 200, 1800, 400, 700, 900, 1700, 1500, 1000

Assume this sample is taken from a bell-shaped dist. with $\mu = 1000$ and $\sigma = 150$. Find the outliers, if any?

Sol: $\mu - 3\sigma = 550$
$\mu + 3\sigma = 1450$



→ 200 and 400 are outliers because $200 < 550$, $400 < 550$

→ 1500, 1700, 1800 are outliers because they are > 1450