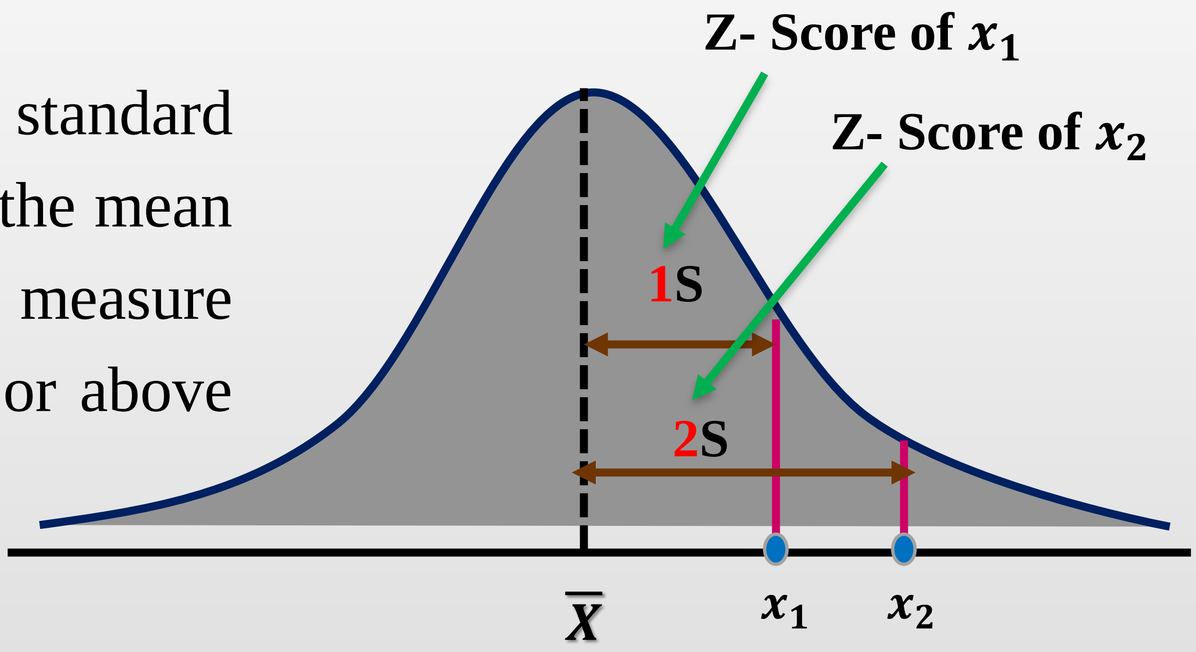
SEC " 4 "

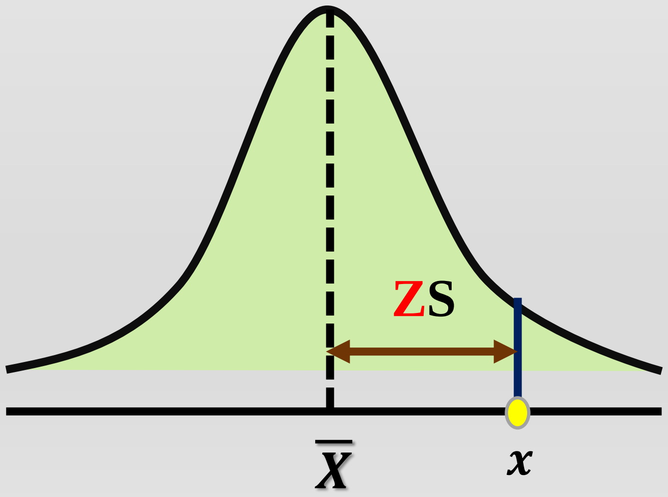
Z – SCORE

CORRELATION

Z - SCORE

- Simply put, a z-score (also called a standard score) gives you an idea of how far from the mean a data point is. but more technically it's a measure of how many standard deviations below or above the population mean a raw score is.



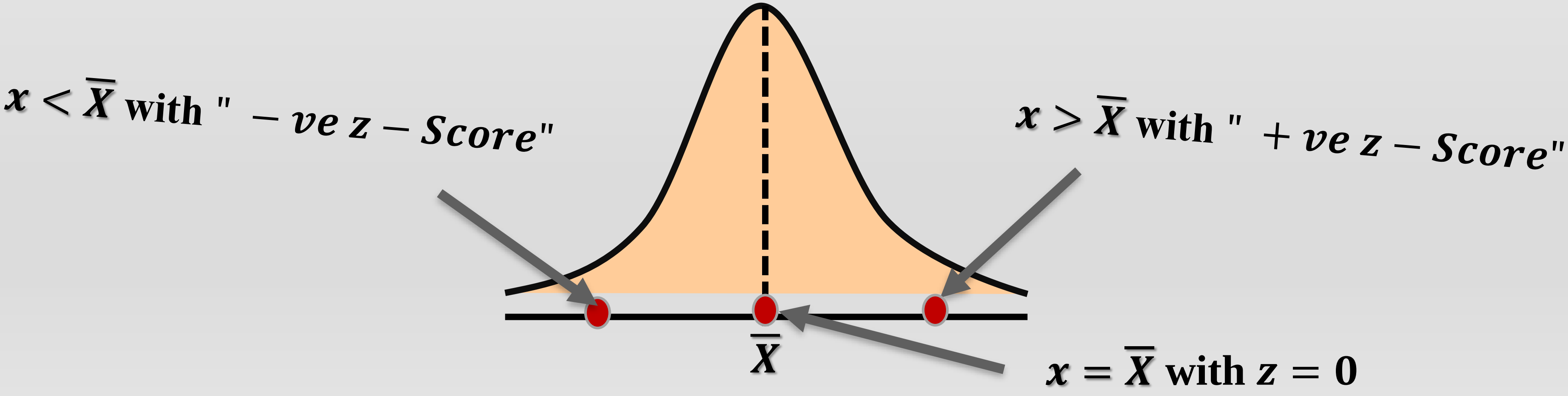


$$x = \overline{X} + Z * S$$

$$Z = \frac{x - \overline{X}}{S}$$

Z- Score

- **Positive** → It indicates that the data point's score is above the mean.
- **Zero** → It indicates that the data point's score is identical to the mean score.
- **Negative** → It indicates that the data point's score is below the mean.



Why are Z-Scores important :

Enables us to compare two scores that are from different samples (which may have different means and standard deviations).

Group A

↓

Mona

$x_M = 25$

$\bar{x}_A = 10$

$s_A = 3.5$

Group B

↓

Sara

$x_s = 30$

$\bar{x}_B = 15$

$s_B = 5.6$

$$Z_M = \frac{x_M - \bar{x}_A}{s_A} = \frac{25 - 10}{3.5} = 4.29$$

$$Z_s = \frac{x_M - \bar{x}_A}{s_A} = \frac{30 - 15}{5.6} = 2.68$$

$$Z_M > Z_s$$

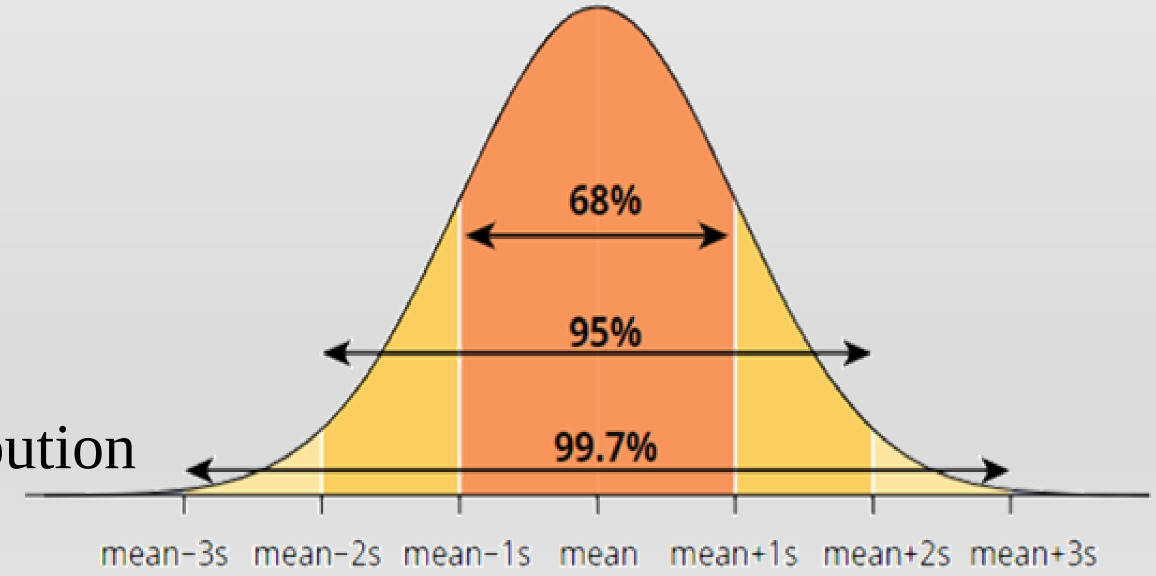
Mona is better than Sarah

Empirical Rule :

Is a statistical rule which states that for a normal distribution, almost all observed data will fall within three standard deviations of the mean or average.

- In particular, the empirical rule predicts that **68%** of observations falls within the first standard deviation **($\bar{X}$ ± 1S)**, **95%** within the first two standard deviations **($\bar{X}$ ± 2S)**, and 99.7% within the first three standard deviations **($\bar{X}$ ± 3S).**

Normal distribution



Sheet (2)

14. Student grades on a chemistry exam were:

{77, 78, 76, 81, 86, 51, 79, 82, 84, 99}

I. Are there any outliers? If so, which scores are they?

II. What would be the score that exceeds the mean by 1.5 standard deviations?

Solution:

I. $n = 10$

- sorted data: 51 76 77 78 79 | 81 82 84 86 99

- Median $= Q_2 = \frac{x_5 + x_6}{2} = \frac{79+81}{2} = 80$, $Q_1 = 77$, $Q_3 = 84$.
- $\text{IQR} = Q_3 - Q_1 = 84 - 77 = 7$.
- $1.5 * \text{IQR} = 1.5 * 7 = 10.5$.
- Any value $> Q_3 + 1.5 * \text{IQR} = 84 + 10.5 = 94.5$. → 99
- Any value $< Q_1 - 1.5 * \text{IQR} = 77 - 10.5 = 66.5$. → 51

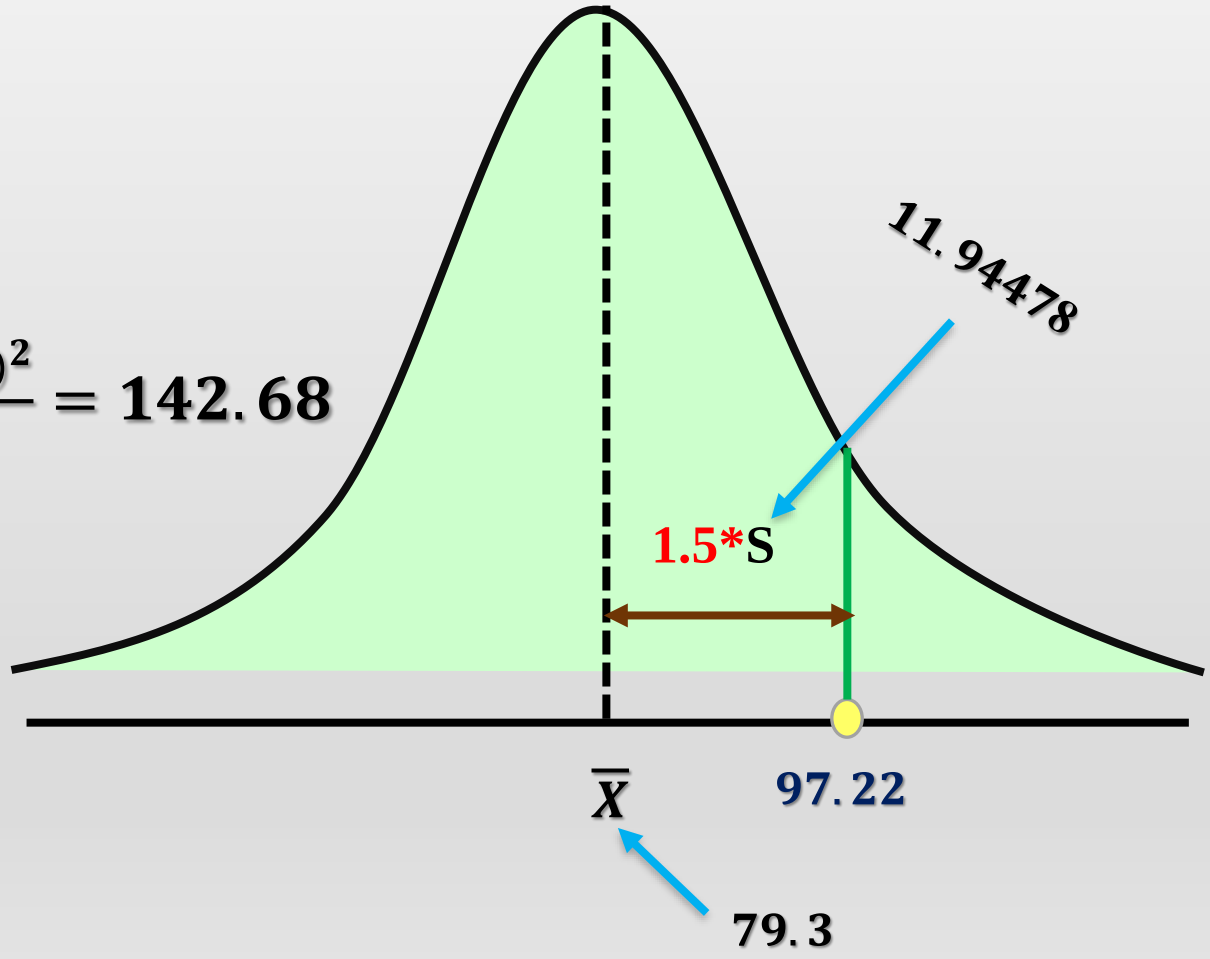
Outliers = 51, 99

II. What would be the score that exceeds the mean by 1.5 standard deviations?

Solution:

- $x = \overline{X} + 1.5 * S$
- $\overline{X} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{77+78+\cdots+99}{10} = 79.3$
- $S^2 = \frac{\sum_{i=1}^{10}(x_i-\overline{x})^2}{9} = \frac{(77-79.3)^2+\cdots+(99-79.3)^2}{9} = 142.68$
- $S = \sqrt{142.68} = 11.94478$

$x = 79.3 + 1.5 * 11.94478 = 97.22$



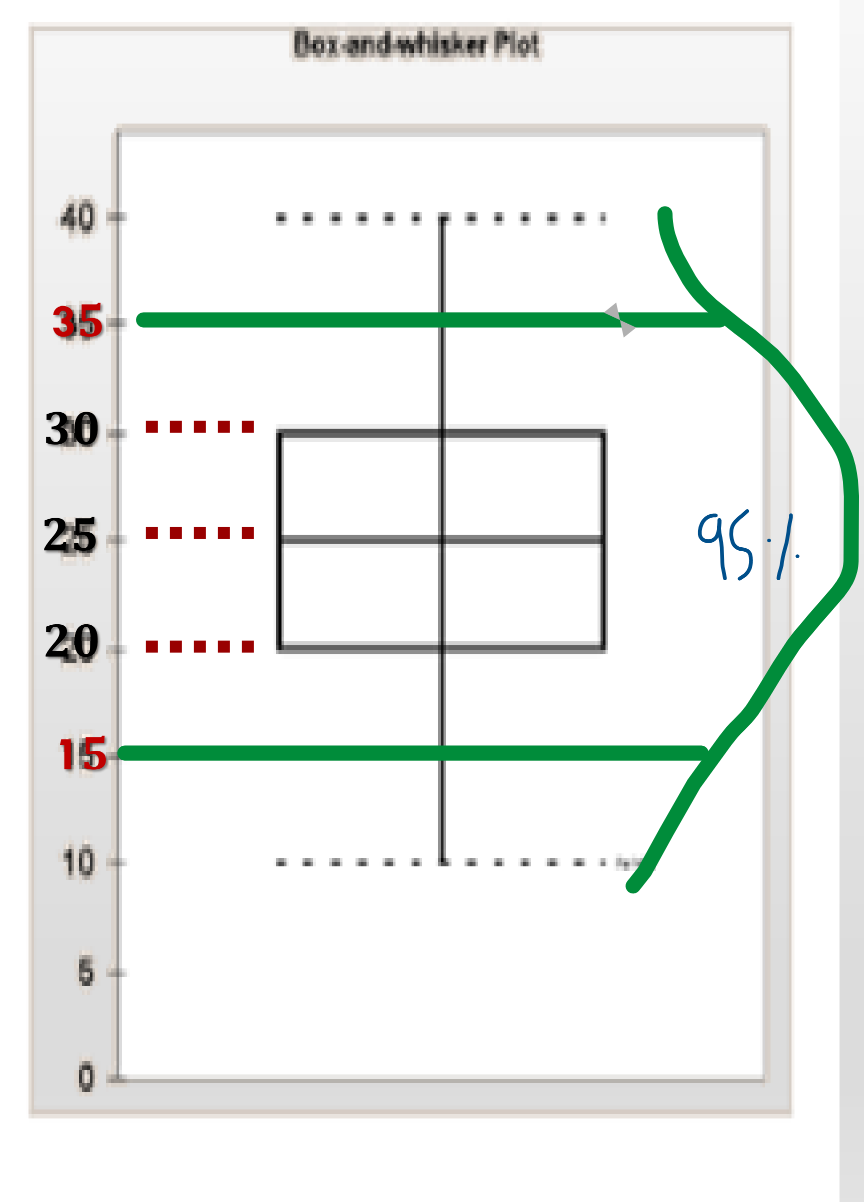
Sheet (2)

15. **Consider the opposite boxplot. Suppose 95% of the data falls between 15 and 35. What is the standard deviation of this sample of data?**

Solution:

- $Q_1 = 20, \quad Q_2 = 25, \quad Q_3 = 30$
- $Q_2 - Q_1 = 25 - 20 = 5,\ Q_3 - Q_2 = 30 - 25 = 5$
- $Q_2 - Q_1 = Q_3 - Q_2$, **the data is symmetric "normal distribution".**
- **95% falls between 15 and 35, so this range constitutes** $(\overline{X} - 2S\ ,\ \overline{X} + 2S\ )$
- **The mean** $\overline{X}$ **is equal to the median (Q2) as the distribution is symmetric. So,** $\overline{X} = 25$
- $\overline{X} + 2S = 35 \longrightarrow 25 + 2S = 35$

  $\longrightarrow \quad S = \frac{35-25}{2} = 5.$



Sheet (2)

16. Mensa, the largest high-IQ society, accepts SAT scores as indicating intelligence. Assume that the mean combined SAT score is 1500, with standard deviation 300. Jacinto scored a combined 2070. Maria took a traditional IQ test and scored 129. On that test, the mean is 100 and the standard deviation is 15. From the test scores, who is more intelligent? Explain.

Solution:

Jacinto

$x = 2070$

$\bar{x} = 1500$

$s = 300$

$$z_J = \frac{2070 - 1500}{300} = 1.9$$

Maria

$x = 129$

$\bar{x} = 100$

$s = 15$

$$z_M = \frac{129 - 100}{15} = 1.933$$

Maria is more intelligent as her Z - score is higher than Jacinto. "$z_M > z_J$"

Sheet (2)

18. The average for the statistics exam was 75 and the standard deviation was 8. Andrey was told by the instructor that he scored 1.5 standard deviations below the mean. What was Andrey's exam score?

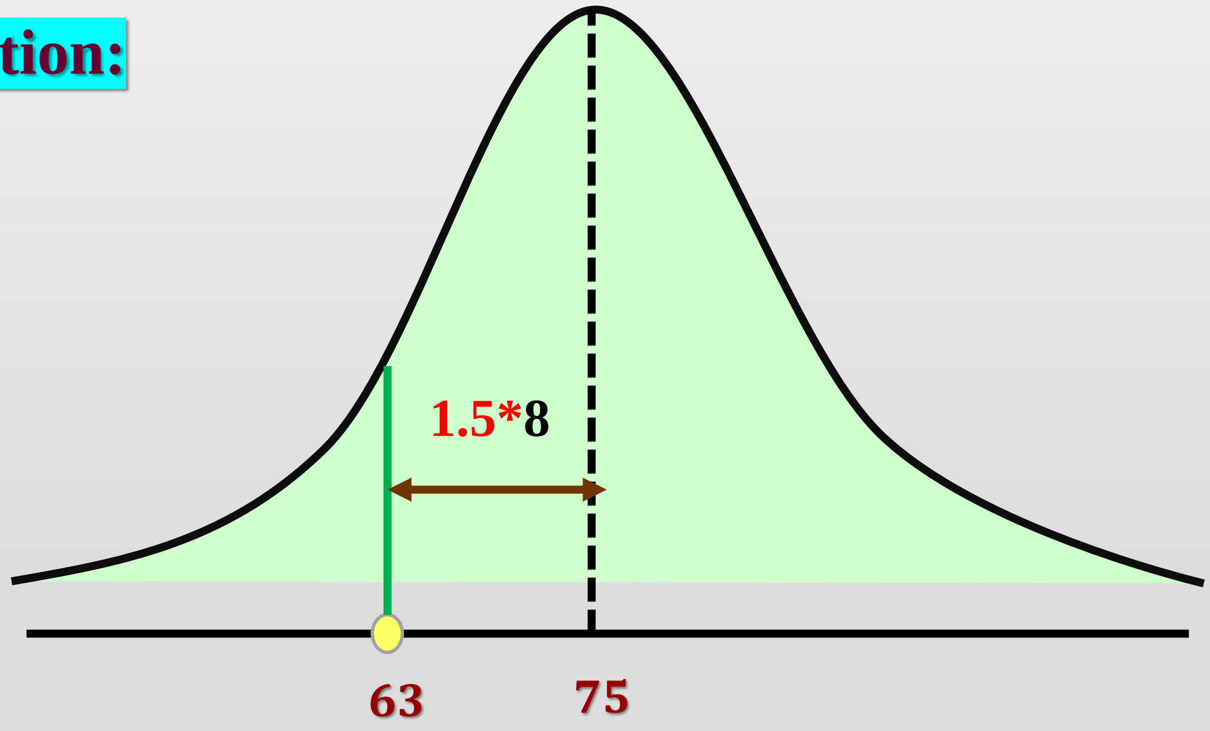
Solution:

$$\bar{X} = 75, \quad s = 8, \quad z = -1.5$$

$$x = \bar{X} + z * S$$

$$= 75 - 1.5 * 8$$

$$= 63$$



Sheet (2)

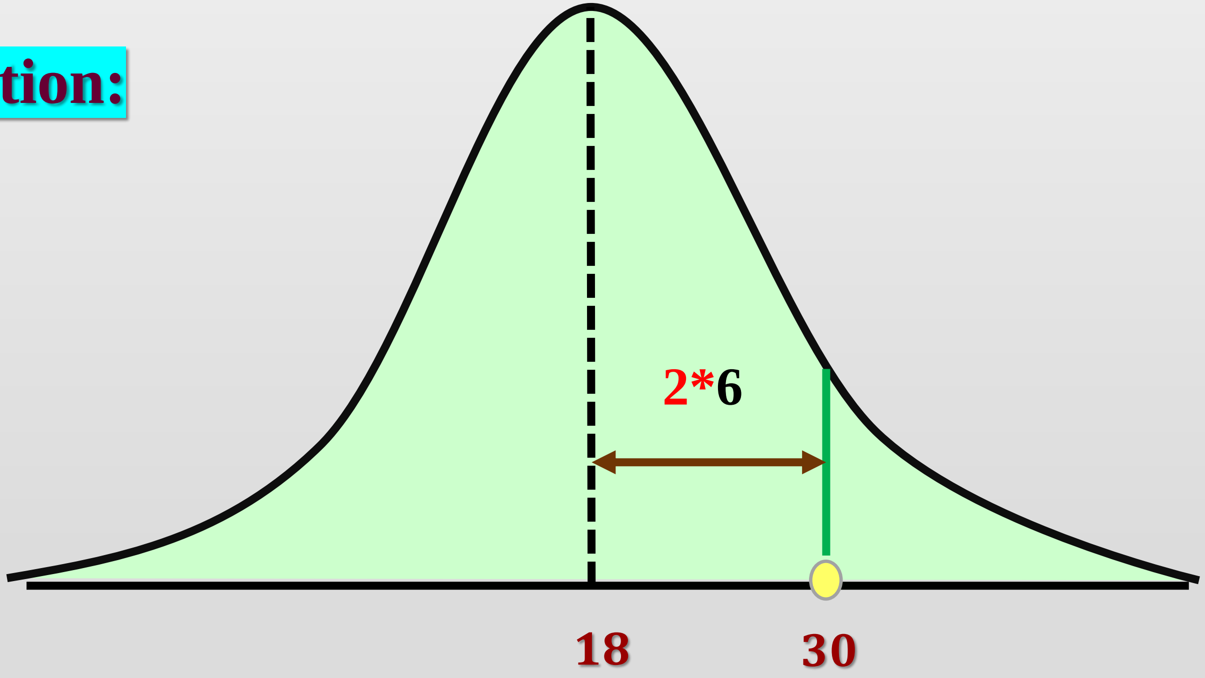
20. A highly selective university will only admit students who place at least 2 z-scores above the mean on the ACT test that has a mean of 18 and a standard deviation of 6. What is the minimum score that the applicant must obtain to be admitted to the university?

Solution:

$$\overline{X} = 18, \qquad s = 6,$$

The minimum score with $z = +2$

$$x = \overline{X} + z * S$$
$$= 18 + 2 * 6$$
$$= 30$$



CORRELATION

- **Correlation shows the strength and the direction of a relationship between two variables and is expressed numerically by the correlation coefficient “ r ”.**
- **A correlation is assumed to be linear (following a line)**
- $-1 \leq r \leq 1.$
- $r = 0$, **is no correlation (the values don't seem linked at all)**

x	y
-	-
-	-

$r =$ **sign** **value**

- **sign**
  - **+ve**: when the values **increase** or **decrease** together.
  - **-ve**: when one value **decreases** as the other **increases**
- **value**
  - **Weak** → **0.1, 0.2, 0.3**
  - **Moderate** → **0.4, 0.5, 0.6**
  - **Strong** → **0.7, 0.8, 0.9**
  - **Perfect** → **1**

$0.6898 \approx 0.7$

$0.96987 \approx 1$

$$z_x = \frac{x - \bar{x}}{s_x}$$

$$z_y = \frac{y - \bar{y}}{s_y}$$

$$r = \frac{\sum z_x z_y}{n - 1}$$

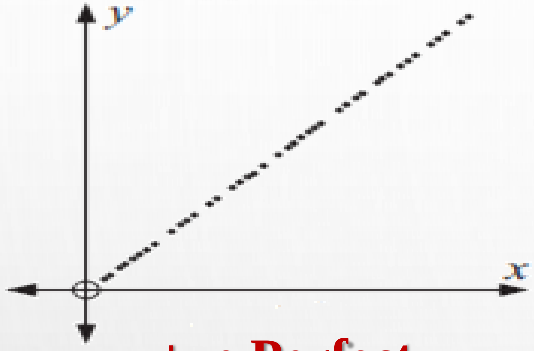
$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}},$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}, \quad s_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$$

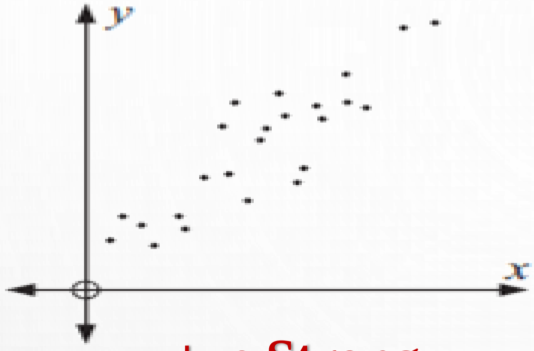
x	z_x	y	z_y	$z_x z_y$

$$\sum z_x z_y$$

Scatter diagram for positive and negative correlation :

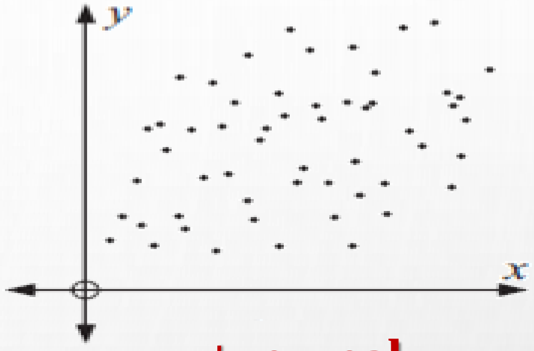


+ve Perfect

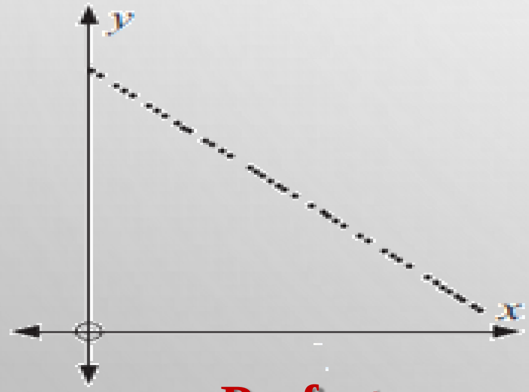


+ve Strong

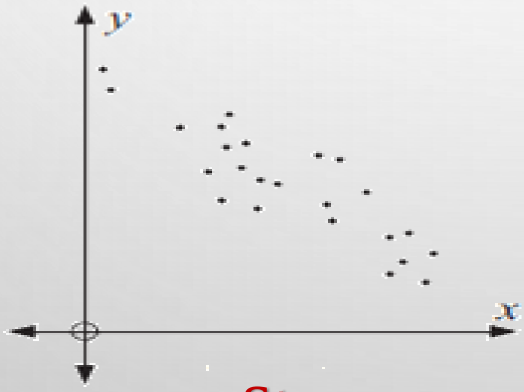
+ve Moderate



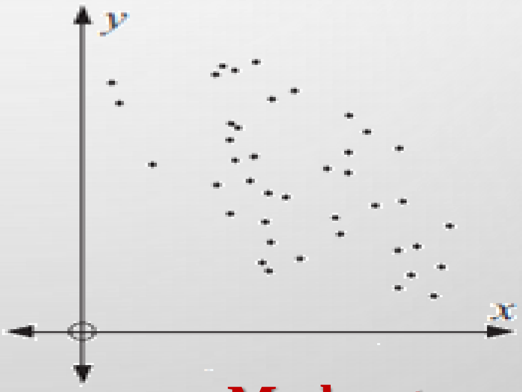
+ve weak



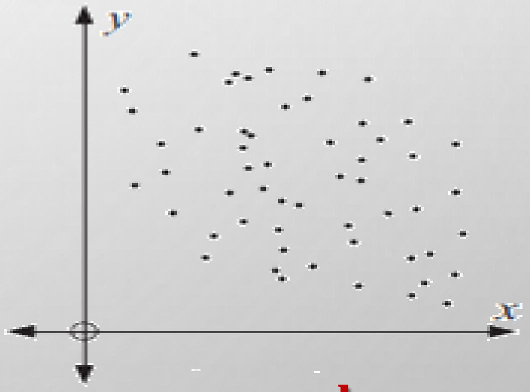
-ve Perfect



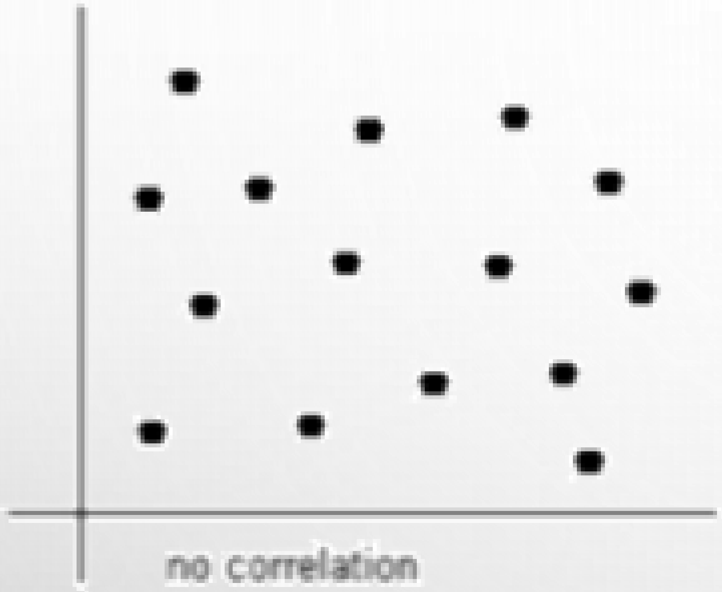
-ve Strong



-ve Moderate



-ve weak



no correlation

non-linear correlation

Sheet (3)

3. Calculate the correlation coefficient for the following data:

x	Y
12	1
8	7
5	4
3	6
2	4
0	2

Solution:

$$n = 6$$

$$\bar{x} = \frac{12 + \cdots + 0}{6} = 5$$

$$S_x = \sqrt{\frac{(12-5)^2 + \cdots + (0-5)^2}{5}} = 4.3818$$

$$\bar{y} = \frac{1 + \cdots + 2}{6} = 4$$

$$S_x = \sqrt{\frac{(1-4)^2 + \cdots + (2-4)^2}{5}} = 2.2804$$

x	$z_x = \frac{x-5}{4.3818}$	Y	$z_y = \frac{y-4}{2.2804}$	$z_x z_y$
12	$\frac{12-5}{4.3818} = 1.5975$	1	$\frac{1-4}{2.2804} = -1.3156$	-2.1017
8	$\frac{8-5}{4.3818} = 0.6847$	7	$\frac{7-4}{2.2804} = 1.3156$	0.9008
5	$\frac{5-5}{4.3818} = 0$	4	$\frac{4-4}{2.2804} = 0$	0
3	$\frac{3-5}{4.3818} = -0.4564$	6	$\frac{6-4}{2.2804} = 0.8770$	-0.4003
2	$\frac{2-5}{4.3818} = -0.6847$	4	$\frac{4-4}{2.2804} = 0$	0
0	$\frac{0-5}{4.3818} = -1.1411$	2	$\frac{2-4}{2.2804} = -0.8770$	1.0007
			$\sum z_x z_y$	-0.6005

$$r = \frac{\sum z_x z_y}{n-1} = \frac{-0.6005}{5} = -0.1201 \approx -0.1$$

-ve weak correlation.