UNIT: DATA ANALYSIS

(ii) Normal Distribution

Z- Scores

Another name for a z-score is a **standardised** score.
The z-score gives information about the **position** of a **score in relation to the MEAN** of the set of scores.

A z-score of 1, means that **the score is one standard deviation above the mean.**

A z-score of 2 means **the score is two standard deviations above the mean** and a z-score of -1 means **the score is one standard deviation below the mean.**

Example 1:
Maree sat for a test and for her mark the z-score was 3.

(a) Briefly explain the meaning of the above statement.
(b) What was Maree's mark if the mean mark was 62% and the standard deviation was 6%?

ANS:
a Maree's mark is 3 standard deviations above the mean.
b Mark=62+3x6 = 80. Maree's mark was 80%.

Example 2:
In a test the mean mark was 73% and the standard deviation was 4%. Bob's z-score was -1.
(a) What information do we get from knowing the z-score?
(b) What was Bob's mark?

ANS:
a Bob's mark is 1 standard deviation below the mean.
b Mark=73-4 = 69. Bob's mark was 69%.

Example 3:

Copy and complete the table.

Number				29	35		
z-score	-3	-2	-1	0	1	2	3

ANS: Mean is 29.

Standard deviation 35 - 29=6

Number	11	17	23	29	35	41	47
z-score	-3	-2	-1	0	1	2	3

Formula for z-scores

The formula $z = \frac{x - \bar{x}}{s}$ can be used to calculate the **z-score** for a particular score x where $\bar{x}$ is the mean and s is the standard deviation.

Example 4:
Find the z-score associated with a score of 18 if the mean is 16 and the standard deviation is 1.6.

ANS: 1.25

Example 5:
The 18 students in a class all sat for a test. The results, out of 50, are shown below.

37	41	28	44	32	35
29	45	40	36	34	31
41	38	32	42	39	36

(a) Find the mean and standard deviation, correct to two decimal places.
(b) Alan scored 39. What is his z-score? Give the answer correct to one decimal place.

ANS: (a) average= 36.67 SD= 4.88 (b) z score = 0.5

Example 6:
Find the standard deviation if a score of 78 has a z-score of 1.4 when the mean of the scores is 71.

ANS: SD = 5

Example 7:
Certain packets of noodles have a mean mass of 90 g with a standard deviation of 1.5 g. What is the mass of a packet of these noodles with z-score -2.2?

ANS: Packet has a mass of 86.7 grams

Comparing scores

Z-scores can be used to compare scores from different data sets. The **higher the z-score**, the **better** the result.

For example, suppose you sat for two tests and scored 75% in both. If the average was 70% in the first test and 80% in the second test then you did better in the first test because your mark was above average, whereas your mark was below average in the second test.

Example 8:

Cameron sat for tests in both Maths and Science. In Maths he scored 87% with a z-score of 1.7 and in Science he scored 73% with a z-score of 2.3. Which is the better result?

ANS:

Cameron did better in Science, because the z-score was higher. [In Science, Cameron's mark was 2.3 times the standard deviation higher than the mean, whereas in Maths his mark was only 1.7 times the standard deviation higher than the mean.]

Example 9:

(a) Andre received a z-score of - 1.2 for his Biology exam. What does this mean?

(b) If Andre received a z-score of -0.8 for his Geology exam, which was the better result?

ANS:

(a) Andre's mark was 1.2 times the standard deviation below the mean.

(b) Geology was the better result. [In Geology Andre's mark was 0.8 times the standard deviation below the mean. It was not as 'bad' as Biology.]

Example 10:

Friends Jake and Nicholas attend different schools and are arguing about who scored better in their half-yearly Maths exams.

(a) Jake scored 80% in his exam. The mean was 67% and the standard deviation was 6.5. What is Jake's z-score?

(b) Nicholas scored 77% in his exam. The mean was 68% and the standard deviation 7.2. What was Nicholas'z-score?

(c) Who performed better? justify your answer.

ANS: (a) z score =2 (b) z score = 1.25

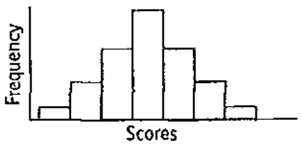
(c) Jake scored the better result. The z-score corresponding to his mark is higher than that corresponding to Nicholas's mark.

Properties of a normal distribution

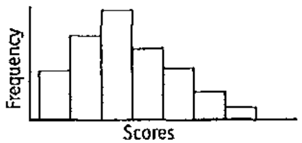
Often the properties of data are such that it displays a **uniform** spread, **symmetrical** about the mean. When this happens we say the data has a **normal distribution.** If the data is **normally distributed** the **mean, mode** and **median** are all **equal**. The frequency histogram of data that is normally distributed is **'bell-shaped'**.

Normal Bell curve

Example 11:
Determine whether the data could be normally distributed from the shape of the histogram.

(a)


(b)


(c)


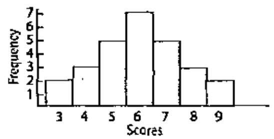
ANS:
(a) Data is normally distributed.[Bell-shaped distribution]
(b) Not normally distributed.[It is symmetrical about the mean but it is not bell-shaped.]
(c) Not normally distributed. [Positively skewed]

Example 12:
Consider the scores:

3 3 4 4 4 5 5 5 5 5 6 6 6 6
6 6 6 7 7 7 7 7 8 8 8 9 9

(a) Find the mean.
(b) What is the mode?
(c) What is the median?
(d) Draw a frequency histogram.
(e) Is the data normally distributed?

ANS: **(a) mean = 6** **(b) Mode = 6** **(c) Median = 6** **(d)**

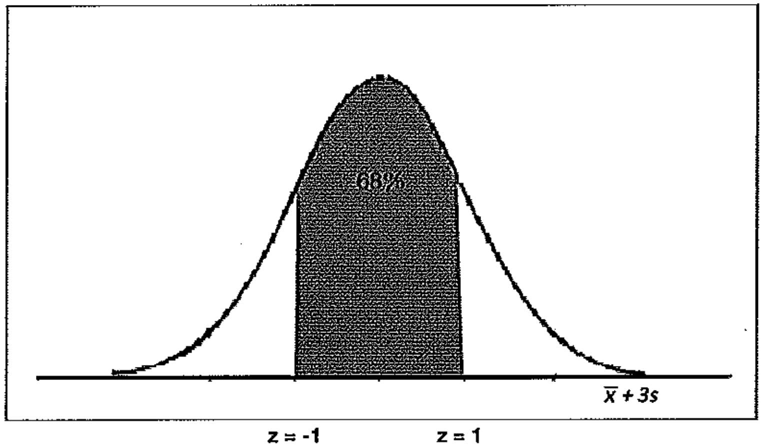


(e) Yes, the data is normally distributed. [The mean, mode and median are all 6. The histogram is 'bell-shaped'.]

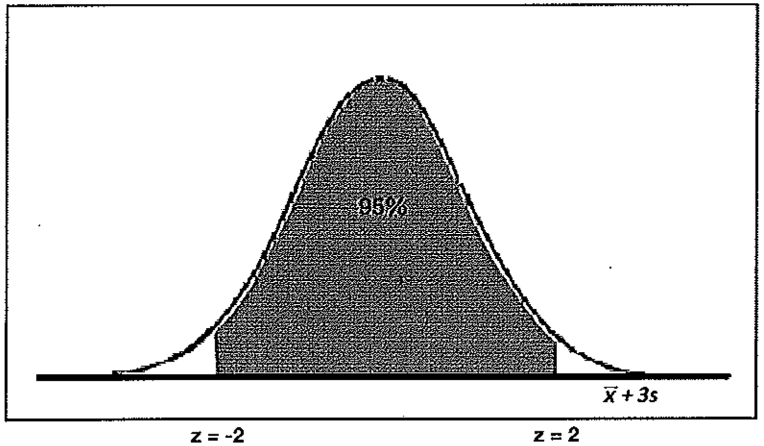
Further properties of the normal distribution

For normally distributed data:

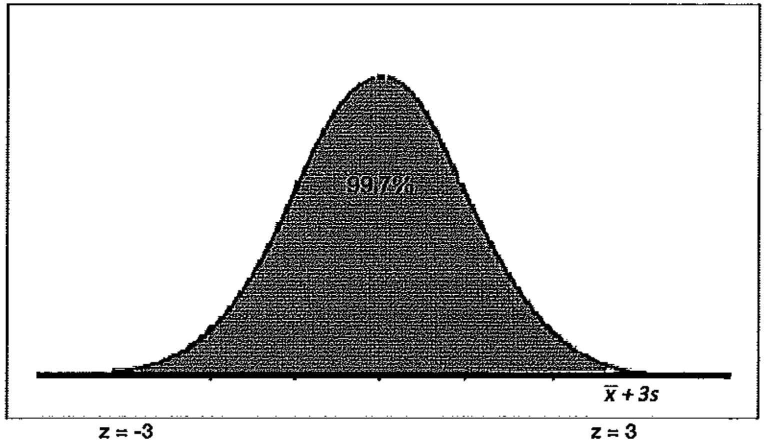
- approximately **68%** of scores lie within **1 standard deviation of the mean (z = 1)**



- approximately **95%** of scores lie within **2 standard deviations of the mean (z = 2)**



- approximately **99.7%** of scores lie within **3 standard deviations of the mean.** (z = 3)



<u>**Note:**</u>

These percentages need to be remembered. They will not be supplied in an exam.

Example 13:

An exam was given to Year 12 students and the results were normally distributed. The table below shows the marks associated with certain z-scores.

Mark	41	48	55	62	69	76	83
z-score	-3	-2	-1	0	1	2	3

(a) What is the mean?
(b) What is the standard deviation?
(c) Approximately what percentage of scores lie between 55 and 69?
(d) Approximately what percentage of scores lie between 48 and 76?
(e) Approximately what percentage of scores lie between 41 and 83?

ANS:
(a) Mean = 62 [$z = 0$]
(b) $s = 69 - 62 = 7$
(c) 68% of scores [$-1 < z < 1$]
(d) 95% of scores [$-2 < z < 2$]
(e) 99.7% of scores [$-3 < z < 3$]

Example 14:

The number of oranges in boxes sent to market by a particular orchardist is normally distributed with mean 125 and standard deviation 8.
(a) Between what numbers will the contents of 68% of the boxes lie?
(b) Approximately what percentage of boxes will have between 101 and 149 oranges?

ANS:
(a) 68% lie within 1 standard deviation of the mean. $\bar{x} + s = 125 + 8 = 133$ $\bar{x} - s = 125 - 8 = 117$
$\therefore$ 68% of boxes will have between 117 and 133 oranges.

(b) $x = 101$, $\bar{x} = 125$, $s = 8$, $z = \frac{x - \bar{x}}{s}$, $z = 3$. 99.7% lie within 3 standard deviations of the mean.

$\therefore$ Approximately 99.7% of boxes will have between 101 and 149 oranges.

IMPORTANT

In a **normal distribution, nearly all scores (99.7%)** lie within **three standard deviations** of the mean. We say that a score **'almost certainly lies'** within three standard deviations of the mean.

Most scores (95%) lie within **two standard deviations** of the mean. We say that a score **'most probably'** lies within two standard deviations of the mean.

Example 15:

The marks in an exam are normally distributed with standard deviation 12.5% and mean 61%.

(a) Between what marks will a score most probably lie?

(b) Between what marks will a score almost certainly lie?

ANS:

A score will most probably lie between 36% and 86%.

A score will almost certainly lie between 23.5% and 98.5%.

It may be necessary to do some calculations to find required percentages. The answers will not always be 68% or 95% or 99.7%.

Remember:

Because the distribution is symmetrical half of the scores lie on each side of the mean.

Example 16:

The mass of certain packets of dog feed are normally distributed with a mean of 21 kg and a standard deviation of 500 g.

(a) Draw up a table to show the masses corresponding to z-scores of

-3, -2, -1, 0, 1, 2 and 3.

(b) What percentage of packets have masses

i. between 20 kg and 22 kg?
ii. between 21 kg and 21.5 kg?
iii. more than 21 kg?
iv. less than 20.5 kg?
v. between 19.5 kg and 20 kg?

ANS:

(a)

Mass (kg)	19.5	20	20.5	21	21.5	22	22.5
z-score	-3	-2	-1	0	1	2	3

(b)

I Between 20 kg and 22 kg is within two standard deviations of the mean. 95% of packets will be between 20 kg and 22 kg.

ii 68% of scores lie within 1 standard deviation of the mean. Half of these will lie between the mean and 1 standard deviation above the mean. 34% of packets will have masses between 21 kg and 21.5 kg.

iii The mean is 21 kg. Half of the packets will be heavier than the mean. 50% of packets will have masses greater than 21 kg.

iv 50% of packets have masses less than 21 kg.
34% of packets have masses between 20.5 kg and 21 kg.
Percentage less than 20.5 kg = 50% - 34% = 16%
16% of packets will have masses less than 20.5 kg.

v 99.7% of packets have masses within 3 standard deviations of the mean and 95% have masses within 2 standard deviations of the mean.

Difference = 99.7% - 95% = 4.7%
4.7% of packets have masses between two and three standard deviations of the mean.
Half of these will be between 19.5 kg and 20 kg.
2.35% of packets will be between 19.5 kg and 20 kg.