LESSON 42

DATA HANDLING (3)

Learning Outcomes and Assessment Standards

Learning Outcome 4: Data handling and probability
Assessment Standard AS 1(a)
Calculate and represent measures of central tendency and dispersion in univariate numerical data by:

- five number summary
- box and whisker diagrams
- ogives
- variance and standard deviation.

Overview

Overview

In this lesson you will learn about:

- Variance
- Standard deviation.

Lesson

Lesson

Measures of dispersion (or spread) about the mean (ungrouped data)

There are two measures of dispersion about the mean:

- **Variance (s^2)**
- **Standard deviation** (s or σ)

These measures of dispersion about the mean are highly effective since they use all of the data values. The range is not a good measure of dispersion because it doesn't eliminate outliers, which are not normally representative of the rest of the data. The interquartile range is better in the sense that it focuses on the central half of the data. The problem with the Interquartile Range is that it is centred about the median and uses only a few of the data values. The mean is a far better measure of central tendency and the variance and standard deviation are measures of dispersion about the mean. These measures are far better than the others.

We will now illustrate these two measures of dispersion by considering an example.

Example

Example 1(without using a calculator)

A basketball team consists of 10 players. The number of points each player scored during the season are as follows:

22 33 38 39 41 52 55 61 67 75

Step 1 Calculate the **mean**.

$$\bar{x} = \frac{\sum x}{n} = \frac{22+33+38+39+41+52+55+61+67+75}{10} = \frac{483}{10} = 48{,}3$$

Step 2 Calculate the **individual deviations from the mean**. Record your results in a table.

POINTS SCORED (x)	$x - \bar{x}$
22	22 – 48,3 = –26,3
33	33 – 48,3 = –15,3
38	38 – 48,3 = –10,3
39	39 – 48,3 = –9,3
41	41 – 48,3 = –7,3
52	52 – 48,3 = –3,7
55	55 – 48,3 = 6,7
61	61 – 48,3 = 12,7
67	67 – 48,3 = 18,7
75	75 – 48,3 = 26,7
	å$(x - \bar{x} = 0$

You will probably notice that the sum of all these deviations from the mean equals 0, which is not very helpful if we are to make conclusions about how the data is spread about the mean.

Step 3 Calculate the **individual squared deviations from the mean**. Record your results in a table. Add these squared deviations.

By squaring the deviations, we are able to eliminate the negative signs so that we have more useful values.

POINTS SCORED (x)	$(x - \bar{x})$	$(x - \bar{x})^2$
22	22 – 48,3 = –26,3	691,69
33	33 – 48,3 = –15,3	234,09
38	38 – 48,3 = –10,3	106,09
39	39 – 48,3 = –9,3	86,49
41	41 – 48,3 = –7,3	53,29
52	52 – 48,3 = –3,7	13,69
55	55 – 48,3 = 6,7	44,89
61	61 – 48,3 = 12,7	161,29
67	67 – 48,3 = 18,7	349,69
75	75 – 48,3 = 26,7	712,89
	å$(x - \bar{x}) = 0$	å$(x - \bar{x})^2 = 2454,1$

Step 4 Determine the **variance** (mean of the squared deviations)

Formula 1 (Grade 11 and 12 Paper 2)

If you are working with an **entire population**, use:

$s^2 = \frac{\Sigma(x - \bar{x})^2}{n}$ to calculate the variance.

Formula 2 (Grade 12 Paper 3)

If you are working with a **random sample of a population**, use:

$s^2 = \frac{\Sigma(x - x)^2}{n - 1}$ to calculate the estimated variance of the population.

In this example, we are working with a basketball team. Therefore, we will use For-

mula 1 to calculate the variance because we are not dealing with a random sample of a larger population. The basketball team is an entire population in its own right.

variance(s^2) = $\frac{\sum(x-\bar{x})^2}{n} = \frac{2454,1}{10} = 245,41$

Step 5 Determine the **standard deviation (*s*)** by square rooting the variance.

standard deviation = s = $\sqrt{\frac{\sum(x-x)^2}{n}} = \sqrt{245,41} = 15,7$

Step 6 Determine the **standard deviation interval** for the data.

This interval is given by: $(\bar{x} - s; \bar{x} + s)$

For the example used, the standard deviation interval is:

(48,3 – 15,7; 48,3 + 15,7)

= (32,6 ; 64)

Step 7 Draw **conclusions** about the spread of the data about the mean by establishing how many of the data values lie within or outside of the standard deviation interval.

In this example, it is clear that 7 of the 10 points lie within the standard deviation interval (33, 38, 39, 41, 52, 55 and 61). This means that most of the players performed well by scoring close to the mean points score. The teamwork was very good for this team.

Note: A small standard deviation indicates that the data items are clustered around the mean. A large standard deviation indicates that the items are more spread out.

Example

Example 1 (using a calculator)

The standard deviation can be easily calculated by using a calculator. We will use the CASIO *fx-82ES* to illustrate the method. The sequence is as follows:

Push the button: **MODE**

Push the button: **2 : STAT**

Push the button: **1 : 1 – VAR**

Enter the points: **22 = 33 = 38 = 39 = 41 = 52 = 55 = 61 = 67 = 75 =**

Push the button: **SHIFT**

Push the button: **STAT**

Push the button: **5: VAR**

Push the button: 3 **: $x\sigma n$**

Push the button: **=**

The answer will read: **15,7**

To clear calculator:

Push the buttons: **MODE**

Push the buttons: **1: COMP**

Example

Example 2

Consider the following table which shows the number of learners who obtained certain marks out of 30 for a class test.

Marks (x)	18	19	20	21	22	23	24	25	26	27
No of learners	2	2	3	4	6	9	12	6	5	3

Calculate the mean and standard deviation for this data.

Solution

Marks x	**Freq f**	**f** $\times x$	$x - \bar{x}$	$(x - \bar{x})^2$	**f** $\times (x - \bar{x})^2$
18	2	36	−5,2	27,04	54,08
19	2	38	−4,2	17,64	35,28
20	3	60	−3,2	10,24	30,72
21	4	84	−2,2	4,84	19,36
22	6	132	−1,2	1,44	8,64
23	9	207	−0,2	0,04	0,36
24	12	288	0,8	0,64	7,68
25	6	150	1,8	3,24	19,44
26	5	130	2,8	7,84	39,2
27	3	81	3,8	14,44	43,32
Total	52	$\bar{x} = \frac{1206}{52} = 23{,}2$			258,08

The mean for this data is 23,2.

$s^2 = \frac{\sum f(x - \bar{x})^2}{n} = \frac{258{,}08}{52} = 4{,}963076923$ (variance)

The standard deviation is:

$s = \frac{\sum f(x - \bar{x})^2}{n} = \sqrt{4{,}963076923} = 2{,}23$

Example 2 (using a calculator)

The standard deviation can be easily calculated by using a calculator. We will use the CASIO *fx-82ES* to illustrate the method. The sequence is as follows:

Push the button: **MODE**

Push the button: **2 : STAT**

Push the button: **1 : 1 – VAR**

Push the buttons: **SHIFT SETUP**

Scroll down and push: **3: STAT**

Push the button: **1: ON**

Enter the marks: **18 = 19 = 20 = 21 =**
22 = 23 = 24 = 25 =
26 = 27 =

Enter the frequencies: **2 = 2= 3 = 4 =**
6 = 9 = 12 = 6 =
5 = 3 =

Push the button: **SHIFT 1**

Push the button: **5: VAR**

Push the button: 3 **:** $x\sigma n$

Push the button: **=**

The answer will read: **2,23**

To clear calculator:

Push the buttons: **SHIFT SETUP**

Scroll down and push: **3: STAT**

Push the button: **2: OFF**

Push the buttons: **MODE**

Push the buttons: **1: COMP**

Activity

Activity 1

1. The number of points scored by four Formula One racing drivers over a number of races are given below:

A	1	1	1	2	6	6	8	8	8	8			
B	1	2	6	8	8	8	8	8	8				–
C	1	1	2	2	4	4	6	6	8	8		–	–
D	2	2	2	4	4	6	6	8	8				–

   (a) Calculate the mean and standard deviation for each driver without using a calculator.

   (b) Now calculate the standard deviation for each driver by using your calculator.

   (c) Discuss the performance of each driver by referring to the standard deviation for each driver.

2. The following table contains the number of learners who obtained certain marks on a class test out of 30.

Marks	20	21	22	23	24	25	26	27	28	29
No of learners	3	3	4	5	7	10	13	5	4	2

Calculate the standard deviation for this data by means of drawing a frequency table.

Lesson

Lesson

Measures of dispersion about the mean (grouped data)

Example

Example 3 (without using a calculator)

Consider the example of the Health and Fitness group. Calculate the standard deviation for this set of data.

Calculate the estimated mean:

Class interval	Frequency	Midpoint class interval	Frequency × Midpoint
0 – 9 years	0	–	–
10 – 19 years	11	14,5	159,5
20 – 29 years	14	24,5	343
30 – 39 years	17	34,5	586,5
40 – 49 years	13	44,5	578,5
50 – 59 years	7	54,5	381,5
60 – 69 years	6	64,5	387
70 – 79 years	5	74,5	372,5
80 – 89 years	4	84,5	338
90 – 99 years	0	–	–
Totals	77	–	3146,5

Estimated mean = $\frac{3146,5}{77} = 40,8$

Calculate the squared deviations from the mean:

Class interval	f	Midp (m)	f × m	$m - \bar{x}$	$(m - \bar{x})^2$	$f \cdot (m - \bar{x})^2$
0-9 years	0	–	–	–	–	–
10-19 years	11	14,5	159,5	14,5 – 40,8 = –26,3	691,69	7608,59
20-29 years	14	24,5	343	24,5 – 40,8 = –16,3	265,69	3719,66
30-39 years	17	34,5	586,5	34,5 – 40,8 = –6,3	39,69	674,73
40-49 years	13	44,5	578,5	44,5 – 40,8 = 3,7	13,69	177,97
50-59 years	7	54,5	381,5	54,5 – 40,8 = 13,7	187,69	1313,83
60-69 years	6	64,5	387	64,5 – 40,8 = 23,7	561,69	3370,14
70-79 years	5	74,5	372,5	74,5 – 40,8 = 33,7	1135,69	5678,45
80-89 years	4	84,5	338	84,5 – 40,8 = 43,7	1909,69	7638,76
90-99 years	0	–	–	–	–	–
Totals	77	–				30182,13

Calculate the variance:

$$s^2 = \frac{\sum f \cdot (m - \bar{x})^2}{n} = \frac{30182,13}{77} = 391,9757143$$

Calculate the standard deviation:

$$s^2 = \sqrt{\frac{\sum f \cdot (m - \bar{x})^2}{n}} = \sqrt{391,9757143} = 19,8$$

Example

Example 3 (using a calculator)

The standard deviation can be easily calculated by using a calculator. We will use the CASIO *fx-82ES* to illustrate the method. The sequence is as follows:

Push the button: **MODE**

Push the button: **2 : STAT**

Push the button: **1 : 1 – VAR**

Push the buttons: **SHIFT SETUP**

Scroll down and push: **3: STAT**

Push the button: **1: ON**

Enter the midpoints: **14,5 = 24,5 = 34,5 = 44,5 = 54,5 = 64,5 = 74,5 = 84,5 =**

Enter the frequencies: **11 = 14 = 17 = 13 = 7 = 6 = 5 = 4 =**

Push the button: **SHIFT 1**

Push the button: **5: VAR**

Push the button: **3 : $x\sigma n$**

Push the button: **=**

The answer will read: **19,8**

To clear calculator:

Push the buttons: **SHIFT SETUP**

Scroll down and push: **3: STAT**

Push the button: **2: OFF**

Push the buttons: **MODE**

Push the buttons: **1: COMP**

Activity

Activity 2

1. The table represents the percentage of income spent on recreation by 50 people.

   (a) Complete the following table for this data.

Percentage	**f**	**Midp (m)**	**f × m**	**m – $\bar{x}$**	**(m – $\bar{x}$)²**	**f × (m – $\bar{x}$)²**
12 < p £ 18	8					
18 < p £ 24	20					
24 < p £ 30	12					
30 < p £ 36	8					
36 < p £ 42	2					
Total			$\bar{x}$ =			

   (b) Calculate the mean for this data.

   (c) Now calculate the standard deviation for this data.

   (d) Now verify your answer by using a calculator.