Bringing British Education to You

www.nccedu.com

Foundation Mathematics

Topic 8 – Lecture 2: Understanding Dispersion

The Variance

The Standard Deviation

V1.0

Scope and Coverage

This topic will cover:

- Recognition of the variance within the distribution of data and its importance in statistics
- Recognition of the standard deviation within the distribution of data and its importance in statistics

Learning Outcomes

By the end of this topic students will be able to:

- Calculate the variance of a set of data
- Calculate the standard deviation of a set of data

The Variance - 1

- We present the difference between the individual value and the arithmetic mean in straight line brackets $|x - \bar{x}|$ called a **modulus**.
- We must remember to ensure that the product of our calculation within the modulus remains positive.
- To get around this we can square the value of x - $\bar{x}$ that is to multiply the value by itself which will always give a positive number.
- If we do this we can replace the straight line (modulus) brackets with the more conventional curved brackets ( )

The Variance - 2

- The resulting measure that is achieved when we square values of x - $\bar{x}$ is referred to as the **variance**.
- For ungrouped data it can be represented by the formula

$$\frac{\sum (x - \bar{x})^2}{n}$$

The Variance - Example

- Consider the following data set 8, 10 ,12, 14, 16, 18, 20 , 22
- The first thing to do is to calculate the arithmetic mean using $\bar{x} = \frac{\Sigma x}{n}$
- Once we have this we can apply the following formula to calculate our variance

x	$(x-\bar{x})$	$(x-\bar{x})^2$
8	-7	49
10	-5	25
12	-3	9
14	-1	1
16	1	1
18	3	9
20	5	25
22	7	49
Total		168

$$\frac{\sum(x-\bar{x})^2}{n}$$

$$= \frac{168}{8} = 21$$

The Variance – Grouped Data

- The same approach can be applied to calculating the variance for group data. This time (as before) we need to find the mid point of our class.

Hours worked	Frequency
$0<h\leq 10$	3
$10<h\leq 20$	6
$20<h\leq 30$	11
$30<h\leq 40$	15
$40<h\leq 50$	12
$50<h\leq 60$	7
$60<h\leq 70$	6

- Constructing our table we get

Midpoint x	Frequency f	fx	x^2	fx^2
5	3	15	25	75
15	6	90	225	1350
25	11	275	625	6875
35	15	525	1225	18375
45	12	540	2025	24300
55	7	385	3025	21175
65	6	390	4225	25350
	$\Sigma f = 60$	2220		$\Sigma fx^2 = 97500$

The Variance – Example Continued

- From this table we can now extract the information we need to put into our formula

$$S^2 = \frac{\sum fx^2}{\sum f} - \left[\frac{\sum fx}{\sum f}\right]^2$$

- We then get the following

$$\frac{97500}{60} - [37]^2 = 256 hours$$

- The mean is calculated using $x = \frac{\sum fx}{\sum f} = 37$

The Standard Deviation - 1

- Although the variance allows us to analyse data in an effective way, it is presenting data in squared units as the variance is expressed as s^2.
- It is necessary to present our data more often than not in single units.
- To achieve this we need to introduce the standard deviation as a way of demonstrating the relationship between the arithmetic mean and individual values.
- The calculation of the standard deviation is very simple and is the square root of the variance.

The Standard Deviation - 2

- In our previous example of how we calculated the variance of the hours worked we established that this = 256 hours
- The standard deviation of this data set is $\sqrt{256}$ or 16 hours
- The main properties of the standard deviation are as follows
  - It is based on all the values in the distribution and so is more comprehensive than dispersion measures based on quartiles
  - It is suitable for further statistical analysis

Coefficient of Variation

- The coefficient of variation compares the dispersion of two distributions and is simply calculated by the following formula

$$\text{Coefficient of variation} = \frac{\text{Standard deviation}}{\text{Mean}}$$

Topic 8 – Understanding Dispersion 2

Any Questions?

Bringing British Education to You

www.nccedu.com