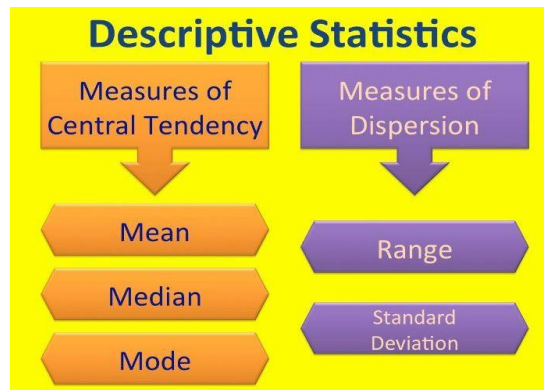# Lesson 8

Descriptive Statistics: Measures of Central Tendency and Dispersion

Learning Objectives

- All students **will** <u>define</u> key terms and <u>calculate</u> different descriptive statistics.

- All students **should** be able to <u>justify</u> when different descriptive statistics should be used.

- All students **could** <u>evaluate</u> the use of different descriptive statistics.

Key Terms
- Descriptive statistics
- Measures of central tendency
- Mean
- Median
- Mode
- Measures of dispersion
- Range
- Standard deviation

Extension activities: You won't need to calculate the standard deviation for your exam, but knowing how to do so will increase your understanding of descriptive statistics and why the standard deviation is so useful. For a practice calculation, see the extension question on page 192 of the *Green Haired Girl* digi-book.

Questions to guide your thinking ...

- Can you define and calculate the mean, median, mode and range?
- Can you state one strength and one weakness for each measure of central tendency?
- What affects whether the mean, media or mode is most appropriate when analysing a set of data?
- What is meant by the standard deviation?
- Can you evaluate both measures of dispersion?

Descriptive statistics

Once quantitative (i.e. numerical) data has been collected, it needs to be summarised and analysed to identify trends or see the 'bigger picture.' This is done using **descriptive statistics**. Descriptive statistics are a way of using numbers to **describe the data** in order to draw meaningful conclusions.

There are three types of descriptive statistics:
- Measures of central tendency
- Measures of dispersion
- Graphs

Measures of central tendency

Measures of central tendency are measures of the average value in a set of data. There are three to consider:

Measure	Definition	How calculated
Mean	The arithmetic average of a data set	Add up all the values (i.e. scores) in a set of data and divide the total by the number of values there are.
Media	The middle value of a data set	Place values in ascending order (lowest to highest) and select the middle value. If there are two values in the middle, the mean of these is calculated.
Mode	The most frequently occurring value in a data set	Do a frequency count. In some data sets there may be two modes (bi-modal) or no mode if all the scores are different.

Example:

$$3, 7, 10, 8, 31, 10, 2$$

$$\text{Mean (avg)} = \frac{3+7+10+8+31+10+2}{7} = \frac{71}{7}$$

$$\downarrow$$

$$10.14$$

$$\overbrace{}^{7 \text{ numbers}}$$

$$\text{Median} = 2, 3, 7, \underset{\underset{\text{middle}}{\uparrow}}{8}, 10, 10, 31$$

$$\downarrow$$

$$8$$

$$\text{Mode} \quad 3, 7, \boxed{10} \; 8, 31, \boxed{10} \; 2$$

$$\downarrow$$

$$10$$

2

For the AQA exam, you need to be able to **define and calculate** each measure of central tendency, but you also need to know **which should be used** with a particular data set. This is because the measures have different strengths and weaknesses.

> ➢ **Complete the evaluation table below**. Decide whether each statement is a **strength** or **weakness**, and to which **measure of central tendency** it is referring.

Statement	Strength or weakness?	Mean, median, mode?
This measure of central tendency is **not affected by extreme scores** so can be the most useful under such circumstances.		
This is a **very crude** measure. It can end up being very different from the median and mean and so not really representative of the data as a whole.		
This measure of dispersion may be **distorted by a single extreme value** as all of the data are taken into account. Therefore, it may end up being unrepresentative of the data set as a whole.		
This measure of central tendency is appropriate for **ordinal data** (i.e. data with an order e.g. first, second, third, etc.) and is easier to calculate than the mean.		
It is the **most sensitive** measure of central tendency as it takes into account **all the scores/values** in the data set. This means it is generally more representative of the data as a whole.		
This measure of central tendency is the **easiest to calculate** and it is the only measure that can use used when there is **nominal data** (data in categories).		
This measure of central tendency is **not as sensitive** as the mean as not all scores are included in the calculation.		

Remember: If you are asked which measure of central tendency should be used with a particular data set, consider whether there are any extreme scores – a score that is much lower or higher than the others.

If there are: The median is the most suitable as the mean would become distorted by the extreme score(s).

If not: The mean is the best option as it is the most sensitive, taking all scores into account.

The mode is never the best option, except if the data are in categories.

TOP TIPS

> **Check your understanding** of measures of central tendency:

1. The following scores were obtained on a questionnaire measuring involvement in health-promoting behaviour (such as taking regular exercise). The higher the score, the more health-promoting activities the participant engaged in. The scores for ten participants are:

17, 19, 16, 21, 15, 24, 20, 22, 49, 18

State the best measure of central tendency for these scores, giving your reasons. (2 marks)

2. A parent decides her children's pocket money by using the modal amount of pocket money which members of their class at school are given. How would this be calculated?

What advantage does it have over using the mean – refer to this particular case when answering the question. (3 marks)

3. A developmental psychologist classifies children as securely attached, insecure-resistant and insecure-anxious. What measure of central tendency would he use to record this data and why? (2 marks)

4. You are a psychology teacher giving advice to students on which measure of central tendency to use on their coursework in which they have investigated learning word lists under different conditions (such as in a quiet or noisy environment). Each student group has chosen their conditions. What comprehensive advice would you give them? (4 marks)

Measures of Dispersion

Measures of central tendency are based on the **spread of scores**, i.e. how far scores vary and differ from one another.
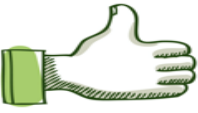
You need to know about two of these measures: The range and standard deviation.

|	
Range – A simple calculation of the dispersion in a set of scores.	
It is calculated by subtracting the lowest score from the highest score and adding 1. (Adding 1 is a mathematical correction that allows for the fact that raw scores are often rounded up or down).	
👍	Easy to calculate - much simpler than the standard deviation.
👎	Does not account for the distribution of scores – it does not indicate whether scores are closely grouped around the mean or spread out evenly (unlike the standard deviation, which does show this).

Example:

(28 – 14) + 1 = 15

Standard deviation – A measure of the average spread around the mean. The larger the standard deviation, the more spread out the data are.	
It is calculated by finding the difference between the mean and each score. All the differences are added up and then divided by the number of scores. This gives the variance. The standard deviation is the square root of the variance.	
👍	More precise than the range as it includes all the scores in the data set. This gives a more accurate picture of the overall distribution of data.
👎	It can be misleading as it can 'hide' some of the characteristics of the data set – extreme values may not be revealed, unlike with the range.

Ideally, you want the standard deviation to be as low as possible (which shows you that the numbers are close to the mean i.e. a low spread of data).

Small standard deviation

Large standard deviation
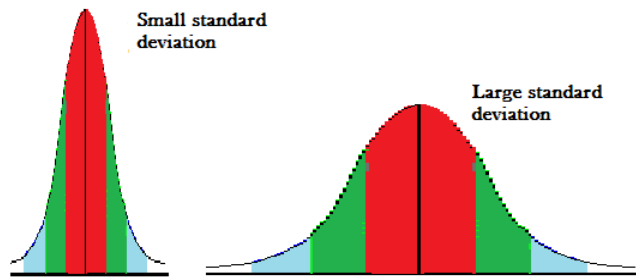
> ➢ **Check your understanding** of measures of dispersion:

1. Felix wants to find a clown to entertain some children at their Christmas Party. He finds two local clowns and looks at their online ratings, each scored out of 10. The top two clowns are Krusty and Bozo. The table shows their ratings and the spread of the scores:

	Mean	Range	Standard Deviation
Krusty	5.6	9	5.9
Bozo	5.5	3	2.3

Which clown should he choose?

a. The mean score for each clown is very similar but the range and standard deviation differ widely. What does this tell us about the ratings for Krusty as compared to the ratings for Bozo? Give as much information as you can. (3 marks)

b. If you wanted to be confident that the clown would not hugely disappoint, which one would you choose and why? (2 marks)

c. If you had to select one of these clowns for a national competition for the funniest clown, which would it be and why? (2 marks).

2

a. What is the advantage of using the range rather than the standard deviation? (1 mark)

b. What is the advantage of using standard deviation rather than range? (1 mark)

Exam Practice: Descriptive Statistics: Measures of Central Tendency and Dispersion

AQA A Unit 1 June 2009

Some psychology students read about an experiment which suggested that organisation is a useful strategy for improving memory. The students carried out an experiment to investigate the effects of organisation on word recall. They made up a list of 50 items that could be bought in a supermarket. The participants were teachers at their school. One group of participants saw the words organised into categories such as fruit, vegetables, dairy products and cleaning materials. The other group saw the same words presented randomly. The results are given in **Table 1** below.

Table 1 The number of words correctly recalled by participants who saw the organised list and participants who saw the random list

	Organised list	**Random list**
	20	15
	15	13
	18	19
	45	14
	24	20
	23	10
	28	21
	21	6
	25	22
	30	25
Measure of central tendency		
Measure of dispersion		

3(a) Identify a suitable measure of central tendency that could be used with these data. Justify your answer. Then calculate this measure of central tendency. Show your calculations (4 marks)

AQA A Unit 1 January 2012

A researcher investigated the effect of age of starting day care on levels of aggression. Four-year-old children attending a day nursery were used. Each child was assessed by the researcher and given an aggression score. A high score indicated a high level of aggression. A low score indicated a low level of aggression. The maximum score was 50.

Table 1 Mean aggression scores for four-year- old children who had started day care before the age of two or after the age of two

	Started day care before the age of two	**Started day care after the age of two**
Mean score	25	23

4(b) What do the mean scores in **Table 1** suggest about the effect of age at which children started day care on children's aggression? (2 marks)

Specimen AS Paper 2 second set

A cognitive psychologist investigating how memory works gave participants the same word list to recall in one of two conditions. All the words were of equal difficulty.

Condition 1: Ten participants recalled the words in the same room in which they had learned the words.
Condition 2: Ten different participants recalled the words in a room that was not the same room as that in which they had learned the words.
The following results were obtained:

Table 1: Mean values and standard deviations for Condition 1 and Condition 2 in a memory experiment.

	Condition 1	Condition 2
Mean	15.9	10.6
Standard deviation	3.78	1.04

3 Why are the standard deviation values found in the study above useful descriptive statistics for the cognitive psychologist? (2 marks)

AQA B Unit 1 June 2013

A psychologist wanted to see whether or not there is a difference in the expectations that men and women have of their own numeracy skills. She obtained a sample of 15 men and 15 women from a factory. She conducted her study in two parts. In the **first part** of the study, the psychologist said to each participant: "I want you to estimate how many marks you think you will get on a maths test that is suitable for 14-year-old children. If the test has a maximum score of 50, what mark do you think you will get?" The psychologist recorded the estimate given by each participant and calculated the median estimates for the men and for the women. The results of the study are given in **Table 2**.

Table 2: Median estimated maths test scores for men and women

	Median estimated maths test score
Men	31
Women	19

3(a) Explain how a median score is calculated. (1 mark)

3(f) What conclusion could the psychologist draw from the median estimated scores in **Table 2**? Justify your answer. (2 marks)

In the **second part** of the study, each participant took a 30-minute maths test suitable for 14-year-old children. The test took place under examination conditions. The psychologist marked the test. The maximum mark was 50. The results of the maths test are given in **Table 3**.

Table 3: Median maths test scores for men and women

	Median maths test score
Men	25
Women	25

3 (g) Taking the results from **both** parts of the study (Table 2 and Table 3), what can the psychologist now conclude? (3 marks)