

MEAN, MEDIAN AND MODE

INVESTIGATION

SURNAME SURVEY

LET'S SET THE SCENE!

A stationery company has approached you as a data analysis expert to gather information for a new range of pencil cases that they would like to produce. The pencil cases are the types that have the plastic slots on the front for placing in the letters of the owner's surname.

Your task is to obtain a random sample of surnames, and analyse the length of the surnames. This data will help the company decide on the most suitable number of plastic slots to include on their pencil cases.



HOW DO WE DO THIS?

Five students, selected by the teacher, will survey different classes in the high school, until you have obtained at least 30 surnames and lengths.

The remaining students will need to get into groups of 3 or 4 and share telephone books. You can use one telephone book per group. You can either use the 'RANDOM' function on your calculator to select pages, or you can simply open up to a random page. You will select names at random by closing your eyes, and carefully placing the tip of your pen/pencil on the open page. Record the surname and the number of letters in the surname. Repeat for 30 surnames using different pages of the book. Each person in the group will collect a different set of names.

Remember: You cannot deliberately choose long or short surnames. It must be random! Record the surnames and lengths using the following table.

Name	Length	Name	Length	Name	Length
1		11		21	
2		12		22	
3		13		23	
4		14		24	
5		15		25	
6		16		26	
7		17		27	
8		18		28	
9		19		29	
10		20		30	

NOW WHAT?

Complete the following frequency distribution table using your data.

Score (x)	Tally	Frequency	fx	Cumulative Frequency
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
		$\Sigma f =$	$\Sigma fx =$	

Now calculate the following:

$$\text{Mean} = \Sigma fx \div \Sigma f =$$

Mode = most frequently occurring score =

$$\text{Median} = \left(\frac{n+1}{2} \right)^{\text{th}} \text{ term} =$$

Range = difference between highest and lowest score =

Your task now is to combine your results together with the other people in your class. The teacher will now project a large table onto the whiteboard. As you finish finding the measures of central tendency, transfer your data onto the whiteboard. You'll need to fill in your own copy of the following frequency distribution table after everyone in your class has copied their data onto the board (Note that there is no need to do a tally column again).

Complete the following frequency distribution table using the combined class data.

Score (x)	Frequency (f)	fx	Cumulative Frequency
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
	$\Sigma f =$	$\Sigma fx =$	

Now calculate the following:

$$\text{Mean} = \Sigma fx \div \Sigma f =$$

Mode = most frequently occurring score =

$$\text{Median} = \left(\frac{n+1}{2} \right)^{\text{th}} \text{ term} =$$

Range = difference between highest and lowest score =

Compare your individual results with the combined class results, and make some comments about your findings below:

Compare your individual results with the results of the whole school surname results.

THINK YOU'VE FINISHED?

The stationery company is also interested in gathering data about the letters themselves. Using the data that you gathered individually, you need to collate the results of the number of letters in the surnames as shown in the table below.

Individual Data

Letters of the Alphabet												
A	B	C	D	E	F	G	H	I	J	K	L	M
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Again, combine your data as a whole group using the table shown on the whiteboard. Copy the finished data into the table below.

Combined Class Data

Letters of the Alphabet												
A	B	C	D	E	F	G	H	I	J	K	L	M
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Finally, you need to graph the class results on a separate sheet of paper (don't forget to put your name at the top) using a bar graph, showing the occurrence of each letter as a percentage of the total number of letters.

QUESTIONS

- Q1. Recommend to the stationery company how many slots should be included on their pencil cases, justifying your reasons why.
- Q2. Which measure of central tendency do you believe is the most effective to accurately represent this data?
- Q3. What is the modal letter in your data? What was the modal letter in the class data?
- Q4. What should be the 5 most manufactured letters? Did you find they were all vowels?
- Q5. If the company were to expand their business to include Europe, would you expect the recommended number of slots should stay the same? Why or why not?
- Q6. Is this an effective method for obtaining such data? Could it be more accurate in any way? (Refer to the comparison between your individual data and the whole class data.)