Relative Frequency

Some probabilities cannot be calculated by just looking at the situation. For example, you cannot work out the probability of which school subject is someone's favourite by assuming that all subjects are equally likely to be chosen. (Try to explain why this is true.) However we can use the results from Phase 1 of *CensusAtSchool* and use these to **estimate** the probability of someone having a particular favourite subject. This is the **relative frequency** or estimated probability using data**.**

In a random sample of 10 people, 4 chose PE/sport as their favourite subject.

Writing this as a fraction, $\dfrac{4}{10} = \dfrac{2}{5}$ As a decimal = 0.4

This **isn't** the probability of PE/sport, but it is an **estimate** of that probability.

We say that the **relative frequency** of choosing PE/sport is $\dfrac{2}{5}$ or 0.4

$$\mathrm{Re}lative\ frequency = \frac{number\ of\ successes}{sample\ size}$$

We can **estimate** the probability of a particular outcome by calculating the relative frequency using the above formula.
The estimate of probability becomes more accurate the larger the sample size. So choosing samples of increasing size we get the following results:

Sample size	Number choosing PE/sport	Relative frequency
5	1	$^1/_5 = 0.2$
10	3	
20	4	
30	6	
50	9	
100	20	
200	46	
500	111	
1000	213	
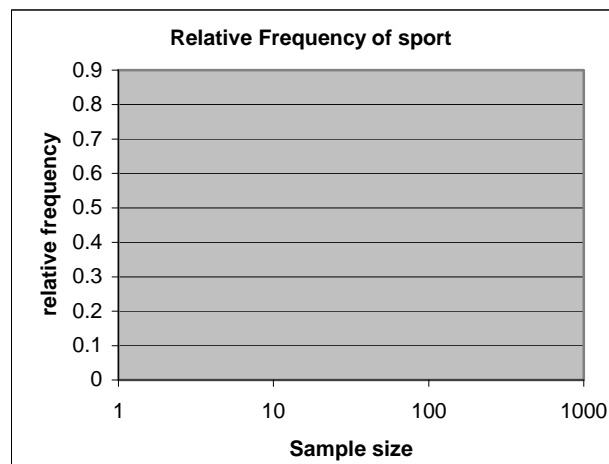
a) Calculate the relative frequencies and put these results into a line graph like the one above to see how the relative frequency behaves. Comment on your results
(Tip - the scale along the bottom is chosen so you can see the results clearly. To do this in Excel choose '*Format axis'* go to the '*Scale tab'* and tick '*logarithmic scale'*)

b) The largest sample size we can have is from the full database of 60,000. This gives $\dfrac{13377}{59392}$ or 0.225 as the relative frequency for choosing PE/Sport as the favourite subject.

c) Use the data below to find the relative frequency of choosing Mathematics as the favourite subject.

Sample size	Number choosing Maths	Relative frequency
5	1	$^1/_5 = 0.2$
10	1	
20	3	
30	4	
50	5	
100	8	
200	19	
500	44	
1000	93	

Fill the gaps in the following:-

If I had a new sample of 5000 I would expect roughly_____of them to choose Mathematics as their favourite subject.

If I had a new sample of 800 I would expect roughly_____of them to choose Mathematics as their favourite subject.

Which estimate do you consider to be the best? Will it always be from the largest sample?

d) Now choose another subject to investigate. Access your samples of data from the random data selector (http://worksheet.censusatschool.ntu.ac.uk/random2/) on the *CensusAtSchool* website. Choose data from the UK phase 1 database.

e) Investigate the relative frequencies of favourite subject for children from South Africa or New Zealand. Use data from the random data selector.

.