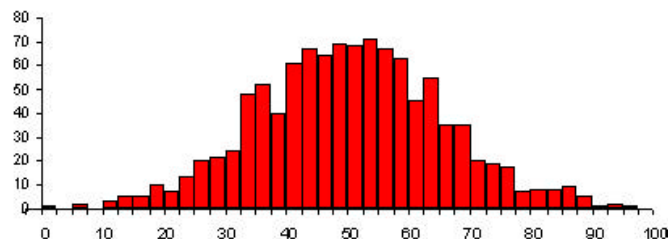


## DISCUSS Sampling Worksheet

In this activity you will use simulations to help you to understand how variation occurs within and between samples from different types of distribution and the relative strengths and weaknesses of using the sample mean, median or mid-range as estimators of the mean value of a population.

Observations from a Normal distribution.



Go to <http://www.mis.coventry.ac.uk/~nhunt/sample/>

First read about the contents of this module including the “brief introduction to how to make the best use of them”. Note that the recommended route through the module is indicated at each stage by the **red** option. After reading the introduction go to the top of the page and click on:

### Why sample?

This page gives some examples of where sampling is used.

Read it then continue on the recommended (red) route by clicking on:

### Variation

Read through this page - it explains the traps that you may fall into when you try to make judgements from data. Then click on:

### within sample

Read about how sample values may vary, then [run the simulation spreadsheet](#). This spreadsheet simulates samples from the common symmetrical distribution that is called the **normal distribution**. Work through the exercise that starts halfway down the worksheet. (You will find it useful to have a pen and paper to hand to help you count the number of items in a simulated sample that satisfy the criteria in the questions.) Enter your answers for each part of a question into the boxes provided on the spreadsheet before you show the answers.

When you have finished click on **Continue** to carry on to:

### More variation within samples (more within samples)

This page introduces two other distributions:

- an evenly spread distribution called the **uniform distribution**
  - a distribution that decreases with time called the **exponential distribution**
- [run the simulation spreadsheet](#) and work through the questions - you will explore the similarities and differences that simulated samples from these distributions have with each other and with the normal distribution you sampled from earlier.

When you have finished click on **Continue** to carry on to:

### Variation in a sample statistic (sample mean)

Read this page, [run the simulation spreadsheet](#) and work through the questions – you will explore the behaviour of the sample mean in samples from different distributions. Again you will find it useful to have a pen and paper to hand to jot down values from the simulations.

When you have finished click on **Continue** to carry on to:

### Variation between statistics (between statistics)

Read the definitions of the median and mid-range then [run the simulation spreadsheet](#) and work through the exercise to see which of the median and mid-range provides the best estimator of the population mean.

You have now completed the part of this module that is most useful for *Using and applying statistics*, but if you wish to explore further there is much more available in this module. Ask your tutor for advice.



<b>Teacher Notes</b>
----------------------

**Unit** Advanced Level, *Using and applying statistics*

**Notes** This module covers much more than is required by the specification for *Using and applying statistics*. The main parts of the module and what students are expected to learn from them are listed below. The UAS column indicates those topics that are included in the specification for *Using and applying statistics* and the Useful column indicates other sections that you may like students to use as extensions. The Omit column identifies the parts that it is recommended that you omit unless they would be useful for the students' other areas of study.

The student worksheet leads students through the UAS and Useful topics listed in the first section on Variation, but could easily be adapted if you wish students to do more or less than this.

Section	Main Points	UAS	Useful	Omit
<b>VARIATION</b>				
<i>within sample</i>	<b>Normal distribution</b> 68% within 1 s.d., 95% within 2 s.d. Mean less spread than individual items. The larger the sample, the less variable the mean. It is difficult to recognise the distribution from a sample.	√		
<i>more within samples</i>	<b>Uniform Distribution</b> Less variation in means of larger samples. Samples not always evenly spread through the range.  <b>Exponential Distribution</b> More difficult to guess what will occur in samples. Again mean less variable in larger samples.		√  √	
<i>sample mean</i>	<b>Uniform Distribution</b> Mean, max, min and range within samples. Range of means across samples smaller than range of items within the samples.  <b>Exponential Distribution</b> Comparison of the number of small values with the number of large values in samples of size 10 and 30. Shape of samples reflects that of the distribution. Estimation of mean. Mean, max and min within samples. Less spread in means than in individual items.		√  √	
<i>between statistics</i>	<b>Exponential Distribution</b> Median often below mid-range value. Mid-range over-estimates the mean. Both median and mid-range are biased estimators of the mean.  <b>Normal Distribution</b> Median and mean are better estimators of the mean, with the median less variable than the mid-range.  <b>Uniform Distribution</b> Mid-range less variable than the median, with both being unbiased.  Conclusion – no easy rules for deciding which is the best estimator.	√	√  √  √	



Section	Main Points	UAS	Useful	Omit
<b>DISTRIBUTIONS</b>				
<b>Normal</b>	In context - speed of cars on by-pass with $\mu = 50$ , $\sigma = 15$ miles per hour. As size of sample increases, sample mean $\rightarrow 50$ , sample standard deviation $\rightarrow 15$ , shape $\rightarrow$ Normal. Distribution of sample means leading to mean of $\bar{x} = \mu$ , standard deviation of $\bar{x} = \frac{\sigma}{\sqrt{n}}$ , shape Normal.		√	
<b>Uniform</b>	Using uniform population over range 0 to 100. As size of sample increases, sample mean $\rightarrow 50$ , sample standard deviation $\rightarrow 29$ , shape $\rightarrow$ uniform. Distribution of sample means leading to mean of $\bar{x} = \mu$ , standard deviation of $\bar{x} = \frac{\sigma}{\sqrt{n}}$ , shape Normal for large $n$ .		√	
<b>Exponential</b>	In context – time gap between cars with $\mu = 20$ , $\sigma = 20$ seconds. As size of sample increases, sample mean $\rightarrow 20$ , sample standard deviation $\rightarrow 20$ , shape $\rightarrow$ exponential. Distribution of sample means leading to mean of $\bar{x} = \mu$ , standard deviation of $\bar{x} = \frac{\sigma}{\sqrt{n}}$ , shape Normal for large $n$ .		√	
<b>Proportion</b>	Once proportion is set, the mean of simulated samples can only take particular values, giving gaps in the graph. Distribution of sample proportions tends to Normal for large values of $n$ , with peak at population proportion. Takes longer to settle to a Normal shape when $p$ is not near 50%.		√	
<b>Flexible</b>	Allows shape of underlying distribution to be altered. Distribution of sample means giving mean of $\bar{x} = \mu$ , standard deviation of $\bar{x} = \frac{\sigma}{\sqrt{n}}$ , shape Normal for large $n$ . Investigation of distributions of sample median and sample mid-range value. Where population is symmetrical these may be unbiased estimators of the mean, but for skewed or irregular distributions they are generally biased. Guessing game – to find shape of underlying distribution.		√	

Section	Main Points	UAS	Useful	Omit
<b>ESTIMATORS</b>				
<b>Normal</b>	Sample mean, median and mid-range are all unbiased estimators of the population mean. The sample mean is the best estimator because it is less variable i.e. has a smaller standard deviation. Efficiency as ratio of variances.			√
<b>Uniform</b>	Mid-range is best estimator, having a lower standard deviation than the sample mean and median.			√
<b>Exponential</b>	The sample mean is the best estimator because it is unbiased, whereas both the median and mid-range are biased.			√



Section	Main Points	UAS	Useful	Omit
<b>CONFIDENCE</b>				
<b>Means</b>	<p>Effect of changing <math>\bar{x}</math>, confidence level, <math>n</math>, on the confidence interval for <math>\mu</math>.</p> <p>Finding sample size needed to give a particular margin of error.</p> <p>When the sample size is small, increasing the sample size reduces the margin of error more than when the sample is large.</p> <p>The more variable the population, the larger the sample needed to give precision.</p>			✓
<b>Proportions</b>	<p>Set in contexts – holiday company, car colours, polls</p> <p>The higher the % confidence level, the wider the confidence interval.</p> <p>As the sample proportion moves away from 0.5, a larger sample is needed to use a Normal approximation to the Binomial distribution.</p> <p>The width of confidence intervals is the same for proportions that add to 1.</p> <p>Finding size of sample needed for particular margin of errors. The lower the level of confidence, the smaller the sample size required.</p> <p>Quadratic relationship between <math>n</math> and <math>p</math>.</p>			✓
<b>What is a CI?</b>	<p><b>Normal Distribution only</b></p> <p>Confidence interval does not always contain the mean.</p> <p>Width of confidence interval increases when the confidence level increases.</p> <p>Width of confidence interval decreases when the sample size increases, in proportion to <math>\sqrt{n}</math>.</p> <p>Problem when standard deviation is not known and necessity of using Student's <math>t</math> distribution.</p> <p>or</p> <p><b><i>t</i> distribution as well as Normal</b></p> <p>As above, then:</p> <p>Width of confidence interval varies from one sample to another because the estimate of the standard deviation varies. For small values of <math>n</math> the s.d. estimates are very variable and a higher value of <math>t</math> is needed to allow for this uncertainty.</p> <p>For a particular confidence level the value of <math>t</math> decreases as <math>n</math> increases converging to the corresponding value of <math>z</math>.</p>			✓

## LINKS

Most are useful for *Using and applying statistics*.

