Extended Project Student Guide: Investigation

Handout 1: What is an Investigation?

If you look up 'investigation' or 'investigate' in a dictionary, these are some of the definitions you will find:

Definition	Source
to search or inquire into with care and accuracy	Chambers English Dictionary
the process of inquiring into or following up; research; study.	*http://en.wiktionary.org/wiki/investigation*
carry out research or study into (a subject or problem, typically one in a scientific or academic field):	http://oxforddictionaries.com/definition/english/investigate

In the context of Extended Project qualifications, the main thing that distinguishes an Investigation/Field Study project from other types of project is the collection and analysis of **primary data**. In other words, you need to make some measurements using apparatus that you have set up in a laboratory, or carry out some fieldwork that involves counting or measurement, or systematically collect information using questionnaires – or maybe a combination of two or more of those approaches.

Your own work is the **primary source** of information for your project. You will also need to refer to **secondary sources** – that is, you should review previous studies that are relevant to your project and refer to their results.

The key elements of an Investigation/Field Study Extended Project are:

- hypothesis or research question
- review of other relevant studies
- collection of primary data
- data analysis
- conclusion(s) based on data analysis
- structured report.

Resource reference

For more on information sources, see chapter 2 of the AQA Extended Project Companion, chapter 1 of the Edexcel Level 3 Extended Project Guide and the section on Using Resources in the OCR Extended Project Guide.

Planning the ways in which you will collect and analyse your data, developing your skills of lab/field work, and learning to use techniques of data analysis, are all important aspects of an Investigation Extended Project. You will first need to decide what to investigate. It is worth spending the effort to look for topic that interests you and where you have some specific ideas about the sort of data you will be able to collect. Time devoted to planning your laboratory or fieldwork will be time well spent. You might have to modify your plan if things don't turn out quite as you expect, but that's a normal part of research work, and is much better than simply making things up as you go along.

Extended Project Student Guide: Investigation

You will need to present your work in a structured report. For an Investigation/Field Study project, suitable headings would be:

- Introduction
- Literature review
- Methods
- Results and discussion
- Conclusion
- References and bibliography.

Resource reference

For more on planning your project, see chapter 1 of the AQA Extended Project Companion, chapter 3 of the Edexcel Level 3 Extended Project Guide and the section on Planning a Project in the OCR Extended Project Guide.

Extending yourself

In making your choice of project objective, remember that a really impressive Extended Project will look very different from coursework pieces you may do for your other qualifications. It may be tempting to 'play safe' and carry out your Extended Project using similar techniques and ideas to those you have already used elsewhere. However, that means you may be missing out on an opportunity to do something creative. When planning your project, look for opportunities to bring in ideas from a wide range of subject areas, learn some new techniques and develop your skills in laboratory or field work.

Ways in which you can extend an Investigation/Field Study project will vary widely according to the topic, but might include:

- Research the history of your topic and explain how ideas have changed over time. Who were the key people involved? Can changes be linked to historical events?
- Research the wider implications of your topic. Are there any ethical questions? Is there an environmental dimension? Is there any relevant legislation?
- Research the applications of your topic. Are there are economic or social factors to consider? Are there any relevant occupations or careers? Do these involve particular codes of practice?
- Learn and use some new practical skills e.g. design and make a piece of apparatus.

It is worth bearing in mind that you may want to use your Extended Project work as part of a portfolio when you are being interviewed in the future. A potential employer or admissions officer for a Higher Education course will be more impressed by evidence that you have extended yourself in new directions than by a portfolio containing two very similar pieces of work.

Resource reference

For more on carrying out your project and writing a report, see chapters 2 and 3 of the AQA Extended Project Companion, chapters 4 and 5 of the Edexcel Level 3 Extended Project Guide, and the sections on Project Formats and Developing and Realising in the OCR Extended Project Guide.

Handout 2: Investigation checklist

Resource reference

For general guidance on planning your project, see chapter 1 of the AQA Extended Project Companion, chapter 3 of the Edexcel Level 3 Extended Project Guide and the section on managing a Project in the OCR Extended Project Guide.

Use this checklist as a guide as you carry out your Investigation/Field Study.

Preliminary work

It is advisable to spend considerable time in preliminary work. Focus is all important and many pitfalls can be avoided by good thought and planning.

Your project must have:

- A clear aim – what are you trying to find out?
- A context or rationale – why is your question/topic interesting or important?

Before you get down to any detailed planning of laboratory work or fieldwork, you need to:

- Discuss your ideas with your teacher/tutor
- State your rationale (reasons) for the project
- Find out about previous studies and projects
- Extend your knowledge and understanding of relevant theories
- Develop your own ideas relating to your research area (for example using a computer simulation or with reference to past projects and secondary sources)
- Pose a research question or formulate a hypothesis that you intend to test
- Consider ways in which you will extend your project beyond the collection and analysis of data

Planning data collection

When you have done some preliminary work you need to:

- plan an approach
- select suitable materials and equipment

Specifically, you should:

- list any laboratory apparatus or other equipment needed
- choose the site(s) where you will carry out any fieldwork
- record the location of any fieldwork sites on a map
- design any questionnaires you intend to use
- decide how you will distribute and collect any questionnaires
- consider data set/sample size
- consider safety and complete a risk assessment
- list the variables that you intend to investigate
- where relevant, decide on the variables that you will control

- consider the resolution/sensitivity of instruments/techniques
- think about accuracy
- think about analytical methods
- carry out a pilot or control
- adapt your plan in view of pilot tests

Collecting and recording data

Once you have fully planned your data collection then you should:

- organise your working environment
- set up experiments carefully
- collect samples methodically
- make field sketches and notes
- draw up results tables
- record data methodically
    - include units
    - take sufficient data points
    - repeat measurements
    - check for anomalies
    - record to appropriate precision
- consider the mode of display of data
- look for patterns and trends
- review plans and made modifications
- explain any changes to procedure

Presenting and analysing data

When you have collected your data you should:

- display the data graphically
    - use sensible scales
    - fully label axes on graphs and charts
- consider error bars and significance levels
- identify trends
- look for correlations in data sets
- apply statistical techniques if appropriate
- extract numerical information from graphs (e.g. gradient, intercept, area) if appropriate
- make calculations
- come to a definite conclusion based on your results

Extended Project Student Guide: Investigation

Writing the report

Resource reference
For more on writing a report, see chapter 3 of the Edexcel Level Extended Project Guide.

Having collected and analysed your data, and reached your conclusion, you will need to present your work in a report. You should:

- structure the report under main headings
- use sub-headings within long sections
- clearly state your aims, hypothesis/research question and conclusions
- explain the background to your project
- include an account of ways in which you have extended your project beyond the collection and analysis of data
- use diagrams, sketches, maps and photographs to convey information about your methods and results
- use *concise* tables, diagrams, maps and graphs to show your main results
- put any large tables of raw data, and any repetitive graphs, maps, diagrams or calculations, into an appendix
- explain specialist terms (e.g. in a glossary)
- reference all sources in a full bibliography
- evaluate the reliability of your sources
- explain the theories, models or equations that you use to analyse your results
- define any symbols or abbreviations that you use
- set out any calculations clearly
- state a conclusion
- confirm or discount your hypothesis
- give an answer to your research question
- link the conclusion to your results
- discuss limitations both qualitatively and quantitatively
- propose future/further work
- write a short summary of your work to go at the beginning of the report
- mark page numbers and produce an index
- proofread the report

Handout 3: Safety and risk assessment

With every practical activity there is an element of risk. Most hazards can be avoided or their impact minimised by prior risk assessment and appropriate action

Hazards can include:

- hazardous chemicals
- exposure to fumes or dust
- electrical hazards
- sharp objects
- large masses
- slippery, fragile or uneven surfaces
- high/low temperatures
- bright/low lighting
- high noise levels
- working at heights
- working in isolation
- working in vulnerable/confrontational situations

In general follow all warnings, notices and advice and act sensibly.

Risk Assessments should be completed for all practical activities. Your school or college will have its own risk assessment forms that should be completed in consultation with your teacher.

Start by filling in Table 1, which could be included in your project report.

Resource reference

Chapter 2 of the AQA Extended Project Companion provides a risk assessment grid.

Table 1: Investigation project safety notes

Name of investigator	*Project title*
Brief description of laboratory/field work	
Hazard	*Suggested control measure(s)*

Under 'suggested control measure' note how you will attempt to control the hazard, and/or how you will limit the harm it might cause.

As well as noting the hazards you should attempt to quantify the **risk** using the ideas of **probability** and **severity**.

A probability of 1 means something is 100% certain to occur. A probability near 0 means something is very unlikely to occur.

To quantify the risk, the probability of a problem occurring should be multiplied by the likely severity of the potential outcome. For example, past experience might suggest minor cuts occur in a design workshop on 1 in 100 occasions (relatively frequent). On a scale of 0-10 the severity of each incident can be ranked (perhaps 1 for a minor cut) thus 0.01 (1%) x severity 1 = 0.01. This is a low risk and is acceptable, especially if plasters are stored close by.

However, a major wound requiring emergency evacuation to hospital might occur 1 in 100,000 occasions, with severity 9. Again we multiply the two factors: 0.00001 x 9 = 0.00009, so really there is very little risk.

In most situations the highest ranked risk will be the most likely to occur, but this is normally the least serious and so the risk is deemed acceptable. Conversely some of the least obviously dangerous risks (e.g. bad weather) can have very serious consequences and need further risk avoidance strategies.

1. Questionnaires and surveys

Designing a questionnaire

All projects involve collecting information. Much information can be found from secondary sources (reliant on other people), but a key feature of an Investigation/Field Study Extended Project is that it must include primary data – data you collect yourself.

In investigations that involve studying people's behaviour or opinions, the questionnaire is the most frequently used source of primary data, but it can also be the least successful.

The following guidelines indicate how some common pitfalls can be avoided.

Plan
Define your research aim. This seems obvious but is often neglected. Consider carefully the exact information that you wish to find out. Do not ask something you don't really need to know.

Time
How long do you have available to carry out the questionnaire? It is no good being so ambitious that you don't have time to complete it.

Questions
Think of the type, number and wording of your questions. Questions can be open or closed:

- A **closed question** requires a definite factual answer, which can usually be expressed by a single word or short phrase, and which is sometimes one of a relatively small number of possible answers anticipated by the questioner. (An extreme example of a closed question is one that can only be answered 'yes' or 'no'.)
- An **open question** is one that gives the respondent the opportunity to provide as much, or as little, information or opinion as they wish. Open questions usually produce much longer answers than closed questions.

The advantage of closed questions is that they allow for easy analysis but the disadvantage is that they could fail to find out some really important information.

The number of questions should be relatively few, no more than seven if the survey is to be used face to face. Respondents are unlikely to give more than a few moments of their time.

The wording must be appropriate to your respondents. Avoid using technical knowledge or specialist language in a survey of the general population, otherwise you will find yourself having to explain the question.

Order
Make sure that your questions follow a logical order both for the sake of the respondent and for yourself in collating the data. Failure to do this will mean extra time at the end. There are dangers if one question depends on the response to another – this is called branching. If there are too many branching questions the survey will become too complicated.

Sample size
The questionnaire must have enough respondents to make the work meaningful and, preferably, statistically significant. In Spearman's rank correlation a minimum of 12 pairs is normally accepted. However, if your questionnaire simply asked twelve people this would not be very impressive.

Consider how the data will be used. A survey that asked only ten people and claimed that '70% of the population stated that ...' could be misleading.

A good starting point is to aim for about 50 responses in order to have the necessary authority to draw conclusions.

Methodology
In schools and colleges, the most frequently used method is the face-to-face questionnaire. The advantage is that you will quickly gain some results.

Directly recorded interviews can be successful, particularly if arrangements are made in advance, but many respondents do not like to be recorded and certainly not filmed.

Postal questionnaires can be used in certain circumstances but tend to have a low return rate – less than 1% in some cases. To be cost-effective and successful a target audience would have to be identified. If teenage respondents were required, then text messages or email could be effective.

Whatever method you adopt for gathering your data, make sure that you abide by the relevant legislation or codes of conduct and that you follow safety procedures. If you want to ask questions to people in the street or on commercial premises, check whether you need permission e.g. from the local council or from an owner/manager. For a postal survey, there are restrictions about the use of information databases, and telephone surveys are bound by strict regulations too.

Bias
A biased sample will produce biased results. This can manifest itself in the nature of the questions or in your target group. The target group could be biased by age, gender or outlook. If you are aware of this then you must try and avoid it unless you state the aims of the survey at the beginning and the source of your answers. For example, should there be a curfew on young teenagers being out at night? It would be relevant in this example to target individual groups separately to gain a balanced result. You must avoid leading questions as this will destroy the integrity of your result.

Timing
The time of day, month and year can all have a substantial influence on your results. Shopping surveys and pedestrian counts are notorious for changing considerably during the day, and the months approaching Christmas and the sales have the highest pedestrian flows while some intermediate months in the summer have the lowest. The year could be significant as social changes in shopping habits become apparent. If possible, collect data on several occasions and look for variations over time.

Design
Use a computer to help you design your survey. Pay attention to layout and make sure it is clear and easy to follow.

Pilot
You should carry out a pilot to establish whether the questions 'work'. No research should really be done without a pilot survey. Failure to undertake a pilot will almost certainly mean that you have not asked the most appropriate questions. For example, if you find that there are questions that people mostly don't answer, or respond 'don't know', then they are probably poor questions.

Additionally, a pilot will give you some result to compare. If your main survey yields different results then ask why.

Good questions?

Here are some examples of good and bad questionnaire design:

Aim
- To establish the impact of a new out-of -town shopping centre.

Bad question. The term 'impact' is vague and does not lend itself to any well-defined measurement.

- To establish the sphere of influence of the new out-of-town shopping centre

Good question. 'Sphere of influence' is a geographical term that can measure the distance people are prepared to travel to use the facility. A questionnaire could yield quantitative information about this.

Age and gender

The first question is often about age and gender, and for a face-to-face survey you can judge this by observation. You should decide whether the information is necessary. In the shopping-centre example it would only be relevant if the work was to investigate shopping patterns further.

Questions

Where do you live? (No choices)

This is far too vague. Most people will not reveal their exact address, and even if they do you will probably not know where it is. Many answers are likely to be difficult to categorise (e.g. 'on the X estate', 'behind the bus station').

You need a more focused question. If you provide a set of options, so that responses can be recorded in tick-boxes, it will make it easier to gather information showing if shoppers are local or if they have travelled larger distances.

In the local town	Within the area (max 20 min journey from the town)	In the county	Other (please state county)

Activity 1: Good questions?
For each of the questions below, say why it is bad and suggest how it could be improved.
- Do you use the centre often?
- What mode of transport did you use to get to the centre?
- Why are you in this shopping centre?
- What type of shopping are you doing?

Surveying the market
In 2003 Spelthorne Borough Council conducted some market research in their town centre. Their questionnaire was quite sophisticated and included several branching questions, making it quite hard to administer. Here are the opening questions.

ASK ALL:

Q.1 Are you visiting and intending to use any facilities in this town centre today?
(1)

Yes, visiting shops	1	GO TO Q.2
Yes, but not shops	2	GO TO Q.2
No	3	CLOSE INTERVIEW

ASK THOSE WHO SAID 'YES, VISITING SHOPS' OR 'YES, BUT NOT SHOPS' AT Q.1, OTHERS CLOSE INTERVIEW

Q.2 Did you come here directly from home, your place of work, another shopping centre, or somewhere else? (2)

Home	1	GO TO Q.4
Work	2	GO TO Q.3
Another shopping centre	3	GO TO Q.3
Somewhere else	4	GO TO Q.3

You can see the full questionnaire at the Spelthorne website:

http://www.spelthorne.gov.uk/CHttpHandler.ashx?id=2173&p=0

In writing

Sometimes it is appropriate to use a written questionnaire that you ask respondents to fill in themselves. This can involve questions and answers with more complexity and depth than is possible when you just stop people for a few minutes in the street.

A postal survey can allow you to collect data from people spread over a large area, but the problem is to persuade the respondents to answer. The response rate improves if you can make individual contact with people, but this is usually not possible. You must expect an extremely low response rate if your survey is simply sent out at random.

Activity 2: Surveying the market

Suggest reasons why the Spelthorne questionnaire would be hard to use. Discuss the advantages and disadvantages of using branching questions.

2-3. Describing and displaying data

Various variables

Continuous variables are those (such as the length and period of a pendulum) which can take any of a continuous range of values.

A **discrete variable** is one that can only take certain numerical values. For example, if you are counting the number of people, N, giving a particular response to a survey question, N is a discrete variable as it can only be a whole number.

Sometimes you might be dealing with data which are grouped into categories that have no numerical meaning. For example, if you survey the speeds at which different colour cars are driven, then the colour is a **categoric variable**.

Descriptive statistics

Many Investigation projects involve gathering data that need to be displayed and analysed in such a way as to provide statistical information. **Descriptive statistics** are numbers that help describe and analyse a sample or a population. They provide a way of summarising information.

The simplest descriptive statistics that you might use are the mean, mode and median. Loosely speaking, each of these is an 'average', but each conveys distinctly different information.

Extended Project Student Guide: Investigation

Mean

The **mean** of a set of data is the most common type of average in everyday use. The mean value of a measurement *x*, denoted *x-bar*, is found by adding up all the different measurements of *x* (e.g. age someone is when they marry) and dividing by the number, *n*, of items in the sample (e.g. the number of people surveyed).

$$x\text{-bar} = \sum x/n$$

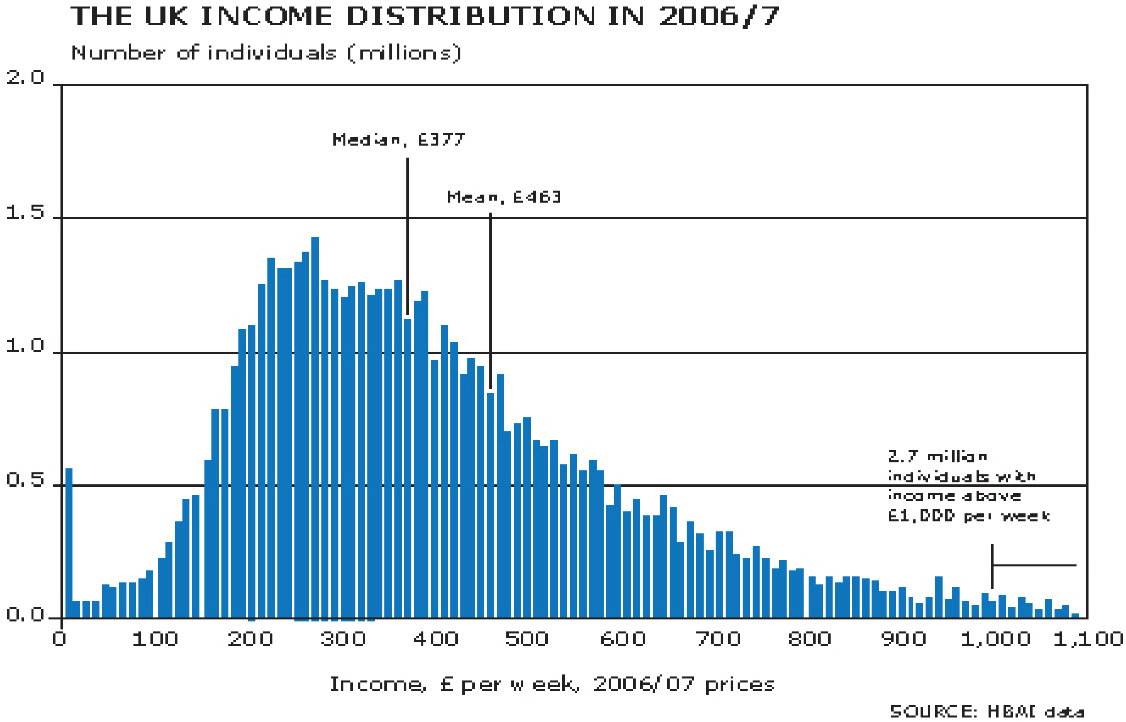
The mean can convey interesting information: for example, the average age of marriage in the UK is 30.5 for men and 28.2 for women (national governmental statistics 2001).

The mean is most meaningful when there is a large sample and/or values of *x* are fairly close to one another. If you have a small sample of widely scattered values, the mean might not represent a typical case. If you are using a survey, or taking samples, ensure you collect plenty of responses or measurements.

Median

The **median** is the middle value of a set of data. Half the data are smaller than the median and half larger. To find the median, put the data in rank order and identify the middle. For example, in Figure 1, the number of people with a weekly income below the median (£377) is the same as the number with an income above the median.

Figure 1: Median and mean income



Using the median tends to minimise the affect of rogue numbers, so if the values are widely scattered between high and low then the median is more useful than the mean. For example, suppose 5 samples yielded the values 1, 2, 3, 4, 100. The median is 3 but the mean would be 22 which would be misleading.

The median is always worth displaying alongside the calculated mean. In the income data shown in Figure 1, the median weekly income is £377 whereas the mean is £463. The difference is considerable for the family budget. The presence of a few people receiving over £1000 per week affects the mean value but does not contribute much to the median. The median therefore gives a better estimate of ‘typical’ weekly income.

Mode

The **mode** is the most frequently-occurring value and can be useful when displaying the distribution in categories. Figure 2 shows the results of a field study measuring the sizes of a certain organism. For both males and females, the mode is 4.1-4.7 cm.

Figure 2: The mode is the value that occurs most frequently

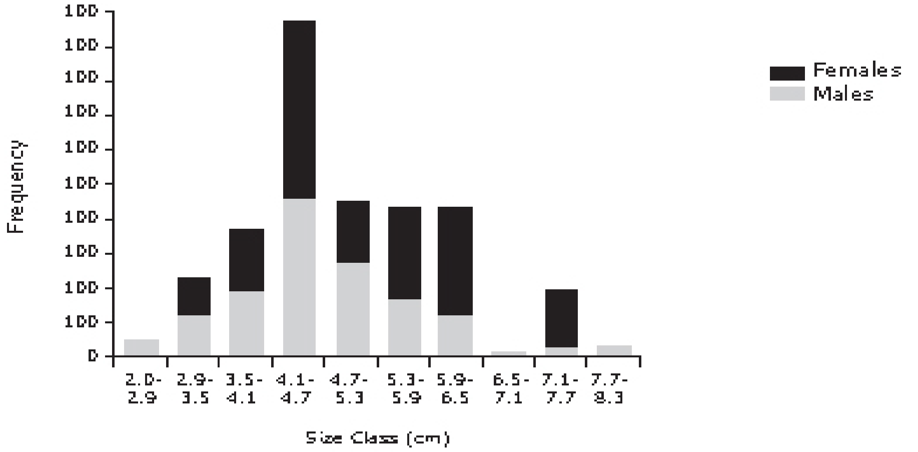
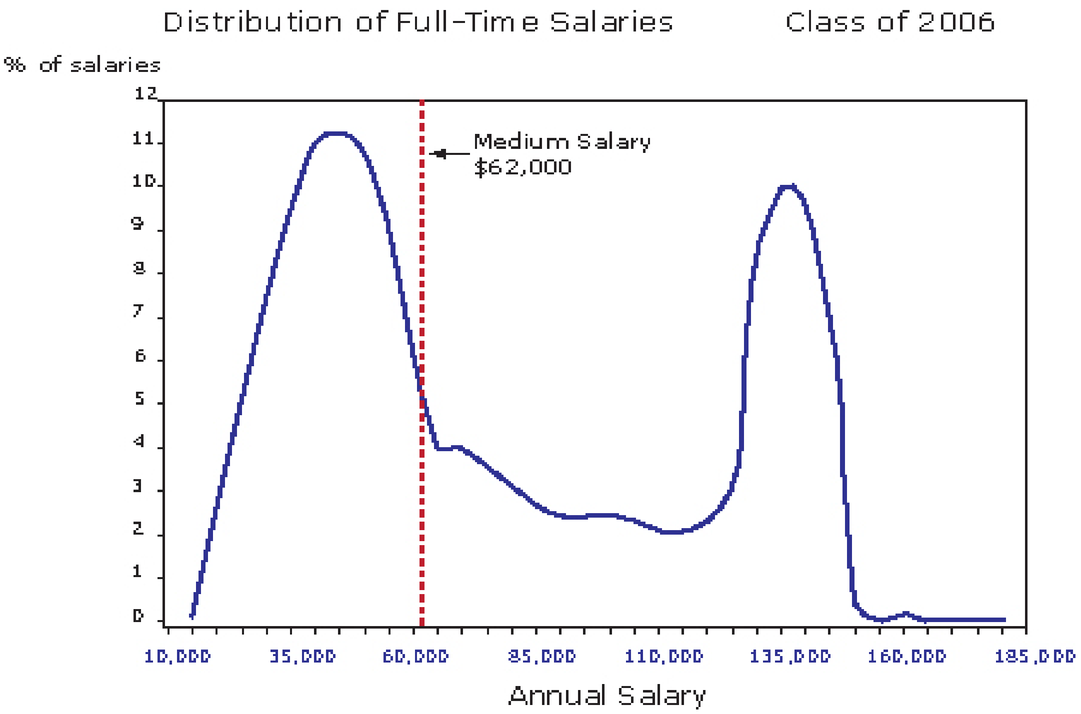


Figure 4. Overall size frquency distribution of Macrobrachium dux at Abavo, Orogodo River study area (March-August, 1995).

Sometimes statistics have a **bimodal distribution**, as in Figure 3, which shows data from a survey of salaries in American law firms. In such cases the best way to describe the data is to quote both modal values; neither the mean nor the median would be very meaningful. In Figure 3, the two modal salaries are about $42000 and $140000; the median of $62000 on its own would give a misleading picture.

Figure 3: Bimodal salary distribution



Activity 3: Describing data
Collect the following data from members of your class:
- height
- shoe size
- number of brothers and sisters
- method of travel from home to school/college (walk, cycle, bus, car, train)

For each set of data, decide whether it would be possible, or meaningful, to use a mean, median or modal value, and then use the appropriate statistics to describe each set.

Using Excel
In any Investigation project you will need to display and manipulate your data in order to extract information and draw conclusions. One common way to do this is to use Excel spreadsheets and their associated graphical capabilities. Excel is straightforward to use and makes manipulation of large datasets easy.

However, it comes with a warning; when using Excel one should be aware of its limitations. A computer will only do what you tell it to do and it will not know the specifics of the experiment or fieldwork which has generated the data. It will not know how to identify irregular results or the presence of systematic or human error. It can only work with the numbers, and often these will be displayed to an unrealistic level of precision. It is rarely useful to use the computer-generated mathematical regression line fit. Spurious and often intricate relationships can be mistakenly identified. If a power law is suspected for example it is better to check by plotting an appropriate log graph. You should always to think through the result you expect and judge trend lines by eye rather than relying on the computer software.

Graphs and charts are a very effective way of communicating trends or relationships; however this is only the case if they are properly labelled and not overcomplicated. Individual points should be readily identifiable and irregular readings commented on. Appropriate scales are very important and beware of 'zooming in' by breaking the axes and ignoring the origin.

Plotting a graph
Excel allows you to plot a variety of graphs and charts. For displaying data from an Investigation, the most useful are bar and column charts, *x-y* scatter graphs and pie charts. Bar and column charts essentially do the same thing, using either horizontal bars or vertical columns to represent the data.

Discrete and categoric variables are best displayed using a bar/column chart or a pie chart. A pie chart is useful only if you need to show how a total number or amount of something is divided up between categories.

An *x-y* scatter graph is the most commonly used for experimental data involving continuous variables. You can use Excel to add trend lines, but be cautious – often a hand-drawn line added 'by eye' is more appropriate.

Error bars
To asses the significance of any results, and the reliability of the trend line chosen, **error bars** should be included to show the uncertainty in each measurement. These should be added from a position of knowledge of the data collection process itself.

Uncertainty in x may well be different from uncertainty in y.

Uncertainty might arise from the measuring instrument itself. For example, if a ruler is graduated in mm, the measurement should only be quoted to the nearest mm and the uncertainty will be ±0.5 mm.

(This uncertainty is sometimes referred to as the **sensitivity** or **resolution** of the measuring instrument.)

If repeated measurements of the same quantity give different results (i.e. results that differ by more than the limitations of the instrument), you will need to calculate and plot the mean value. The uncertainty can be estimated by looking at the range of values above and below the mean, or (if you have a lot of repeated measurements) can be calculated as the standard deviation in the mean.

Using Excel, you can add error bars to an Excel *x-y* scatter graph.

Gradients

The **gradient** of a graph can often yield useful information. Gradients are best measured by selecting suitable points on the trend line by eye, rather than using the computer to calculate from the whole range of data points plotted. Choose points that *lie on the trend line*, rather than plotted data points, and choose points that are as far apart as possible.

Remember that gradients will often have units and any prefixes such as milli, mega etc in the variables must be taken into account.

Activity 4: Displaying data
Decide how best to display each set of data that you collected in Activity 3.
Imagine you wanted to explore the relationship between height and shoe size. What type of graph or chart would be suitable for displaying these two sets of data together?

Use Excel to produce graphs and charts of the data from Activity 3.

4-5. Testing a hypothesis

Defining terms

A **hypothesis** (plural: hypotheses) is a statement of what you believe might be the outcome of experimental or observational exploration, which you put forward for further investigation. Often a hypothesis is an 'educated guess' based on previous knowledge.

In some types of Investigation, it is appropriate to begin by stating a hypothesis, and the purpose of your laboratory work or field work is then to test whether your hypothesis is confirmed or rejected.

Formally, you should state a basic or **null hypothesis** (H_0), and an alternative hypothesis (H_1).

The hypotheses are tested to see if they can be accepted or rejected at a **significance level** (normally 5% in effect with 95% certainty), i.e. the results do not lie at the far end/s of some distribution of probability

The significance level is usually stated as a percentage, which demonstrates at what percentage the null hypothesis is rejected and the alternative hypothesis accepted. The regions in which the null hypothesis is rejected are called **critical regions**.

Throwing a coin

Imagine you want to investigate whether heads are obtained more often than tails when throwing a given coin. If this is the case, then the probability of throwing heads is greater than 0.5. (If $p = 0.5$, then 50% of throws yield heads.) We state the null and alternative hypotheses:

H_0: $p = 0.5$

H_1: $p > 0.5$

We decide to test them at 5% (this is the significance level). The critical region is $R(h) > 0.95$.

We throw coin A 100 times and obtain heads 61 times. $R(h) = 0.61$

$R(h) < 0.95$. This is not in the critical region so we accept the null hypothesis: we can be 95% confident that coin A does not give more heads than tails.

We then throw coin B 100 times and obtain heads 96 times. $R(h) = 0.96$

Now $R(h) > 0.95$. This lies in the critical region so for coin B we must accept the alternative hypothesis: we are 9% confident that coin B does give more heads than tails.

Correlation

Correlation is often used in data analysis to help establish a relationship or association between two variables.

The **Spearman Rank correlation** establishes a possible association by dealing with data placed in rank order (see later for a worked example). It does not matter whether the data follow a 'normal distribution' so the Spearman test can be used in a wide variety of situations, making it a commonly used test.

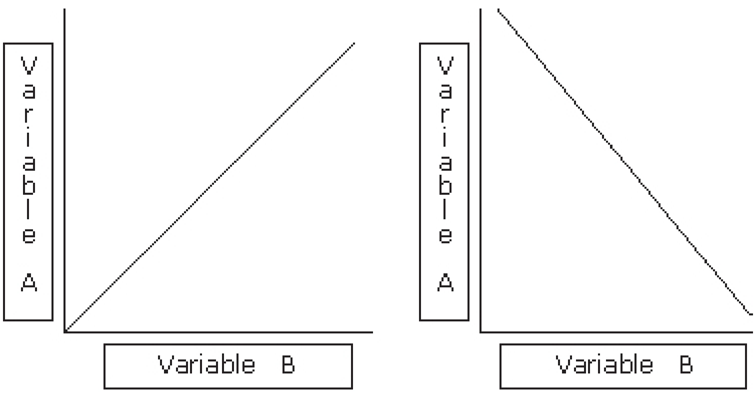
A **correlation coefficient** is calculated to ascertain whether there is a positive or negative association between two sets of data. Using the Spearman test, the coefficient is usually given the symbol r_s.

A rank correlation coefficient always has to be between +1 and -1 (if your calculated answer is outside this range then you have made mistake).

A positive correlation means that if one variable increases so does the other (Figure 4 - a). For example, there is usually a positive correlation between life expectancy and GNP – life expectancy usually increases if a country's GNP increases.

A negative correlation means that where one variable increases the other decreases (Figure 4 - b). For example, there is usually a negative correlation between altitude and the number of plant species.

Figure 4: Positive and negative correlation



(a) $Rs = 1$ (b) $Rs = -1$

Figure 4 shows idealised situations with all points lying on a straight line. In practice points will be scattered above and below the line so the relationship will not be perfect.
If $r_s = 0$, there is no relationship between the variables. For intermediate values, you need to consider whether the correlation is significant.

Hypothesis and significance

At the start of your statistical test you will have an alternative hypothesis and a null hypothesis, H_1 and H_0. The hypothesis H_1 states that there is a relationship between two variables and H_0 states that there is no relationship between the two.

Your calculated r_s must be bigger than the one given in a set of significance tables (e.g. Table 1) if you are to accept H_1. If it is not, then you have to accept your null hypothesis. For example, if you had 11 ranked pairs and calculated r_s = 0.6, your hypothesis H_1 could be accepted at the level of 95% certainty, but not 99%.

Table 1 Significance levels

Number of data pairs	*Minimum r_s for correlation to be significant at the stated level*	
	Significance 0.05 (95%)	*Significance 0.01 (99%)*
8	0.643	0.833
9	0.600	0.783
10	0.564	0.745
11	0.523	0.736
12	0.497	0.703

Correlation?

We will illustrate the procedure with an example.

Harry and Hermione were studying Hagrid's flop-eared snaffles. They noticed that the bigger creatures seemed to have floppier ears, and they wondered whether there was a correlation between body mass and ear length.

In order to test for a correlation they measured both variables and ranked the pairs in increasing order of body mass.

For each of the snaffles they calculated the difference, *d*, between the rankings of body mass and ear length. Table 2 shows their data and calculations.

Table 2 Body mass and ear length

Snaffle	*Body mass/kg*	*Body-mass rank*	*Ear length/cm*	*Ear-length rank*	*d_i*	*d_i^2*
A	50	2	24	1	1.0	1.00
B	50	2	26	2	0.0	0.00
C	50	2	30	3.5	1.5	2.25
D	56	4	32	6.5	2.5	6.25
E	59	5.5	30	3.5	2.0	4.00
F	59	5.5	32	6.5	1.0	1.00
G	60	7	31	5	2.0	4.00
H	66	8	35	8	0.0	0.00
					Total	18.50

Note the procedure for dealing with two or more items that have the same value. Snaffles A, B and C all had the same body mass and occupied ranks 1, 2 and 3; they were all assigned to rank 2. Snaffles E and F occupied ranks 5 and 6 so both were given rank 5.5.

Activity 5: Look at the data
Draw an *x*-*y* scatter graph using the data from Table 2. Plot body mass along the *x*-axis and ear length up the *y*-axis (do not join up the points). Discuss whether there appears to be a correlation between the variables.

Harry and Hermione then used the formula for r_s:

$$r_s = 1 - \frac{6(d_1^2 + d_2^2 + \dots + d_n^2)}{n(n^2-1)} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

where n is the number of data pairs.
Their hypotheses were:
H_0: $r_s = 0$. There is no correlation between ear length and body mass
H_1: $r_s > 0$. There is positive correlation between ear length and body mass. .
Here, $n = 8$ as there were 8 snaffles, so

$$r_s = 1 - \frac{6 \times 18.5}{8(8^2-1)} = 1 - \frac{111}{504} \approx 0.780 \text{ (3 s. f.)}$$

From statistical tables (Table 1), the critical value of r_s at this significance level for $n = 8$ is 0.643. They therefore rejected H_0 and accepted H_1: at a significance level of 0.05 (95% confidence) there is a correlation between ear length and body mass for flop-eared snaffles.

Harry and Hermione wondered whether there might also be a correlation between body mass and tail length. Their data are listed in Table 3.

Table 3 Body mass and tail length

Snaffle	*Body mass/kg*	*Body-mass rank*	*Length of tail/cm*	*Tail-length rank*	d_i	d_i^2
A	50	2	14			
B	50	2	18			
C	50	2	18			
D	56	4	17			
E	59	5.5	13			
F	59	5.5	15			
G	60	7	19			
H	66	8	11			
					Total	

Activity 6: Look – more data!
Add another vertical axis to your chart from Activity 5. Label this axis 'tail length/cm' and use the data from Table 3 to plot points showing body mass and tail length (use a different colour to distinguish them from the body-mass points). Discuss whether there appears to be a correlation.

Activity 7: Correlation?
Complete Table 3. Calculate r_s for body mass and tail length and hence decide whether there is a correlation between those variables. State your conclusion in terms of hypotheses and significance levels.

Activity 8: Join the dots?
Discuss whether joining the dots or drawing a trend line on your chart is valid in the case of (a) body mass and ear length (b) body mass and tail length.

6. In the river

In the river

Anne saw a TV news report about a factory that had accidentally discharged some waste material into a river, killing the fish and probably having a long-term effect on the biodiversity. She knew that a factory on the outskirts of her town regularly discharged waste into the river, and she wondered how even this routine legal discharge was affecting the river's ecology.

For her project, she decided to investigate the biodiversity in the river upstream and downstream of the local factory outfall. She discussed with her teacher how she could write a research question that would include the exploration of wider issues, while still putting the collection and analysis of primary data at the heart of her project. Her research question was: To what extent is current UK legislation on factory discharge effective at safeguarding the biodiversity in rivers?

Project springboard
An environmental concern can provide a good focus for an Investigation/Field Study project. TV and newspaper reports can be a good source of ideas.

Having decided on her main research question, Anne wrote down some specific questions that would help her plan and carry out her project. As the project developed, she added more questions so that eventually she had a long list:

- How can biodiversity be defined and measured?
- Given that the null hypothesis is that the factory outfall has no effect, what alternative hypotheses might be investigated, and how?
- How can the river best be sampled?
- What measurements and tests should be made on the samples?
- How can water quality be analysed chemically?
- Would sampling the river at different times of year provide useful additional data?
- Would a study of the ecology of the river banks provide useful additional data?

- How might the river-bank ecology be studied?
- What are the health and safety issues that need to be considered in carrying out the fieldwork?
- What would be the best way(s) to analyse the data?
- What statistical tests could be used?
- How can the data and/or the main results be presented using graphs and charts?
- Are there any significant differences in the biodiversity or water quality above and below the outfall?
- What other studies of river biodiversity and water quality have been carried out?
- Are there any other studies that relate specifically to factory discharge into rivers?
- What substances does the local factory discharge?
- What limits does current legislation place on factory discharge?
- Does the legislation mention particular substances?
- What are the limits on quantity or frequency of discharge?
- How is the legislation enforced?
- What role do factory inspectors and environmental protection officers play?
- What information could be obtained by interviewing a factory manager or community liaison officer?
- If the limits on discharge were made stricter, what would be the implications for the factory?
- Would tighter legislation have economic consequences for the factory?
- What would be the ethical implications of changing the legislation?
- If the factory had to close, how would that affect local people?

Activity 9: That's the question

With a partner, discuss Anne's list of questions:

- Which questions would she need to address before carrying out her fieldwork?
- Which would help her identify 'milestones' for her project planning chart??
- Which might be answered using primary data from her fieldwork?
- Which provide some starting points for her Research review?
- Which could she discuss as part of the Analysis or Evaluation of her project?
- Which questions:
  - fit into the '5W' classification and hence need factual answers?
  - invite speculation?
  - might be addressed using ethical frameworks?
  - invite the expression of a point of view and the use of argument?

Resource reference

The Edexcel Level 3 Extended Project Guide has more advice and information on the 5W questions (chapter 1) ethical frameworks (chapter 2) planning (chapter 3) research review, discussion/evaluation (chapters 4 and 5) points of view, argument and counter-argument (chapters 2 and 5).

The OCR Extended Project Guide gives some specific advice on Investigation projects in the section on Project Formats. There is advice on planning (Managing a Project), researching information (Using Resources) and on evaluation (Reviewing).

Project springboard
Writing down a list of questions can help you break down your project into specific tasks.
Look for opportunities to explore wider questions arising from your Investigation/Field Study project.

Preparing for fieldwork
At an early stage of her project, Anne looked back at her own records of biology and geography fieldwork, and visited the websites of the Field Studies Council (FSC) and the British Ecological Society (BES) to find out more about standard procedures and techniques. This research helped her to develop an appropriate methodology for her own project.

Activity 10: Preparing for fieldwork
Research some standard techniques for collecting and analysing data.
Visit the FSC website and explore the free online resources for students, teachers and environmental professionals.
http://www.field-studies-council.org/publications/for-teachers-and-students/free-online-resources.aspx

Visit the BES website and explore the education pages. Take a virtual tour of some ecological habitats and find out how they are studied.
http://www.britishecologicalsociety.org

Resource reference
The AQA Extended Project Companion has some specific advice collecting primary data in chapter 2.
There is also advice on planning (chapter 1) and presentation (chapter 4).

7. Into the air

Into the air

Gil had a keen interest in physics and hoped to study the subject at university. On reading a magazine article he became intrigued by devices known as rail guns. A rail gun is based on the principle of a linear electric motor, which means that it uses the force produced by an electric current in a magnetic field to launch projectiles at high speed. Rail guns might in future be used for launching small satellites into space without the need for rockets, and they also have the potential for military use. Gil decided to build a rail gun and investigate its performance for his project.

Gil's further reading of books and websites showed him how to arrange the necessary permanent magnets and power supplies, and he carried out a short pilot study to find out whether he could design and build a rail gun with equipment available in his school. He was constrained both by safety considerations and the availability of apparatus, but succeeded in making a working model that fired objects at quite low speeds (less than 1 m s^{-1}). He planned to investigate how the speed and range depended on parameters such as the strength of the magnetic field, and then scale up his results to draw some conclusions about the likely behaviour of larger, more powerful rail guns.

His research question was: How does the range of a rail gun depend on its physical parameters? What are the implications of these results for the future of rail guns?

Project springboard
Magazines such as *New Scientist, Scientific American* and the *Review* magazines (e.g. *Physics Review, Chemistry Review, Biological Sciences Review, Geography Review*) can provide ideas for Investigation projects.

Activity 11: Extended Project
The examiners responsible for the Edexcel Extended Project qualification have identified three ways in which a project might be extended: deepening understanding, broadening skills and widening perspectives.
Read the account of Gil's project and identify ways in which he was able to demonstrate the extended nature of his project.

Theory
Gil carried out extensive background research that helped him to plan and carry out his laboratory work, and to write his project report. He used physics text-books to explore the theory of electromagnetism, which took him well beyond the requirements of his physics course specification. He linked this theory to the design of his model rail gun and used it to make some predictions about the likely speed of the projectiles. The same theory enabled him to understand how magnetic field strength could be measured using a simple probe.

He found out about the speeds necessary to launch a satellite into orbit or to send a vehicle into space, and related these speeds to theoretical calculations involving gravitational fields and energy conservation. He also drew on his knowledge of laws of motion to relate the range of a projectile to its launch speed.

The Introduction to Gil's project report contained a full account of the underlying theory, explaining how it related to his laboratory work.

Other studies
Gil was able to find details of some other work on rail guns. These studies mostly reported on the different designs used for the projectile system and power supply and stated the power demands. In the Evaluation section of his report, he compared some of these designs to his own.

He found that rail guns had first been properly researched in the 1980s as part of US President Reagan's 'star wars' defence initiative. This made him think about the ethics of using scientific research to develop weapons. He developed quite strong opinions on the matter, which he explored in the Discussion section of his project report.

Apparatus
From his pilot study, Gil identified aspects of his original model that could be improved. Making use of his school's workshop facilities, he was able to design and make a safe, reliable model rail gun that allowed him to vary its physical parameters. In order to achieve this, he needed to become skilled in some workshop techniques that were new to him. He described this aspect of his project in the Method section of his report, which was illustrated with diagrams and photographs and included a review of safety considerations.

In the Method section, he also described how he measured the field of a large permanent magnet. To do this, he first calibrated a magnetic-field probe using a current-carrying coil to produce a field that could be calculated using electromagnetic theory. In this section he also described how he measured the speed and range of the projectiles.

Data and results

Using his model rail gun, Gil varied the electric current, the distance between the rails and the strength of the permanent magnet while keeping other factors constant. In each case, he collected data on the force applied to the projectile during launch, and the projectile's speed and range. He plotted graphs of his results and, by referring to the errors and uncertainties in his data, evaluated the extent to which his graphs agreed with his theoretical predictions.

A mathematical analysis of his results led Gil to produce some equations linking the performance of the gun to its physical parameters. He used these equations to calculate the launch speed of a much larger rail gun, which enabled him to evaluate whether such devices could realistically be used for launching satellites.

Activity 12: Laboratory work

A teacher who read Gil's report commented that it was an 'exemplary laboratory investigation, which could be used as a template for such work'. Explain what you think the teacher meant, and identify features of Gil's laboratory work that make it 'exemplary'. Then think how you could incorporate those features into your own laboratory work.

8. In the town

In the town

Ellie had enjoyed her work experience in a hotel then took a weekend job in a town-centre cafe. Her ambition was one day to run her own catering business.

The cafe was open from 10 am to 6 pm, and was busy all day serving tea, coffee, lunch and snacks to shoppers and to visitors who came to explore the town's historic centre. For her project, Ellie was interested in collecting data on the cafe's customers and it's sales of food and drink, but her tutor was concerned that she might gather a lot of data without a having a very clear focus. Ellie talked to the cafe owner, who was wondering whether to open in the evenings. He suggested to Ellie the idea of researching data for a business plan, which gave her a clear objective for an Investigation. The research question that she agreed with her tutor was: Is there a good business case for extending the opening hours of the cafe to serve evening meals?

Ellie wrote down several ideas about data she could collect and other things to do. Taking her tutor's advice about having a clear focus, she then had to decide which were worth following up.

Activity 13: What to do?

In a group, discuss Ellie's suggestions (below) and say which you think would be worth pursuing, which could be adapted to be more useful (and how) – and which are probably not worth spending time on. Add any suggestions of your own. Remember that the project work must relate to the main objective, and should also demonstrate 'extension' in terms of deepening understanding, broadening skills and widening perspectives.

Ellie's suggestion:

- Survey the cafe's weekend customers to find out whether they are local people or visitors and whether they would come to the cafe for an evening meal.
- Survey people visiting the town centre in the evenings to find out if they would visit the cafe, ask them what type of meal they would choose (snack, fast food, top-quality restaurant meal) and how much they would spend.

- Study the cafe's records to find out how much the average customer spends, and what types of food and drink lead to the highest profits.
- Interview the cafe owner.
- Cook some sample menus for evening meals, offer them to passers-by and ask for their opinions.
- Use Yellow Pages to find the locations of other cafes in the town
- Learn how to use a spreadsheet to model the cash flow of a business.
- Use textbooks to find out about theories of small-business management.
- Interview someone from the local tourist office to find out about patterns of visitors to the town.
- Find out about bank loans for business expansion.
- Research legislation relating to cafes and restaurants e.g. working hours, licensing, health and hygiene, wages ...
- Make a market map of the area.
- Collect information about the menus and opening hours of other cafes and restaurants nearby.
- Visit websites that give advice about running a small business.

Ellie's work enabled her to produce a business plan for the cafe, and helped the owner to decide whether to serve evening meals.

Project springboard

Work experience can provide a good starting point for project work. Having a real-life situation to research can provide a good focus for your Investigation.