Geography

Unit 1: Geographical Methods and Techniques

[ADVANCED HIGHER]

Acknowledgement
Learning and Teaching Scotland gratefully acknowledge the following for permission to reproduce copyright material: Bowen, A and Pallister, J, *Tackling Geography Coursework*, London: Hodder & Stoughton, 1997; Cole, J P, *Situations in Human Geography*, Oxford: Basil Blackwell, 1975; Nagle, G, *Skills and Techniques for Geography A-Level*, Cheltenham: Stanley Thornes (Publishers) Ltd, 1998; Nagle, N and Spencer, K, *Geographical Enquiries*, Cheltenham: Stanley Thornes (Publishers) Ltd, 2000; Taylor, L, *Geographical Techniques*, Cambridge: Pearson Publishing, 1997; Speak, P and Carter, A H C, *Map Reading and Interpretation*, London: Longman, 1970; Worthington, B D R and Gant, Robert, *Ordnance Survey Mapwork*, Basingstoke: Macmillan, 1983; *Statistics Guide*, Milton Keynes: The Open University Press, 1985. Thanks are also due to Douglas McConnach for maps and diagrams.

Introduction

Advanced Higher Geography contains three units. They are:

Unit 1: Geographical Methods and Techniques
Unit 2: The Geographical Study
Unit 3: Geographical Issues.

In Advanced Higher Geography, students must study and apply geographical methods and techniques to analyse information, identify relationships and present information. This part of the pack, **Geographical Methods and Techniques**, gives the student practice, advice and information in these areas.

It cannot be emphasised strongly enough that this unit gives the student practice using statistical, mapping and graphical skills and techniques that they can use in Unit 2 (The Geographical Study) and Unit 3 (Geographical Issues). Therefore, as students work their way through this unit they should put some thought into how they can apply these skills and techniques to subsequent work.

Further, competence in the statistics section of this unit is a key element for the core skills of Numeracy and Problem-solving being automatically certificated at Higher level.

The unit assessment requires candidates to provide evidence of the successful application of these methods and techniques.

Unit evidence (for **internal** assessment) requires the student to have evidence of:

1. Competence in one Physical Geography fieldwork method.
2. Competence in one Human Geography fieldwork method.
3. One piece of evidence to show that the student can produce a map and/or a diagram.
4. One piece of evidence to show that the student can extract and interpret information from a 1:25000 OS map.

Where a candidate is undertaking the unit as part of a course, these skills will also be assessed through the **externally** set and assessed written paper of two hours' duration.

It is assumed that the student has already passed Higher Geography and is competent in OS map reading and interpretation and the examination of geographical methods and techniques at Higher level.

This unit is divided into three parts that correspond to the three skill areas that make up Geographical Methods and Techniques. These are:

- statistical awareness
- the production and interpretation of maps and diagrams
- fieldwork survey/measurement and recording techniques.

Statistical awareness has been further divided into three sub-sections:

Section 1: The graphical presentation of data
Section 2: Descriptive statistics
Section 3: Inferential statistics.

The production and interpretation of maps and diagrams has likewise been divided into three sub-sections:

Section 4: The design and layout of maps
Section 5: The interpretation of OS maps and related data
Section 6: Topographic analysis based on OS maps.

Fieldwork survey/measurement and recording techniques only represents one section (Section 7).

The three units which make up Advanced Higher Geography and the contents of the three skill areas of Geographical Methods and Techniques are shown diagrammatically overleaf.

ADVANCED HIGHER GEOGRAPHY

UNIT 1
GEOGRAPHICAL METHODS AND TECHNIQUES

UNIT 2
THE GEOGRAPHICAL STUDY

UNIT 3
GEOGRAPHICAL ISSUES

SKILL AREA
FIELDWORK METHODS

PHYSICAL TOPICS
Morphological mapping
Vegetation sampling
Slope analysis
Stream analysis
Meteorology in the local setting
Soil profiles and characteristics
Pebble analysis

HUMAN TOPICS
Urban land use mapping
Traffic, pedestrian and environmental quality surveys
Questionnaire design and implementation
Use of secondary sources
Reilly's law of retail gravitation
Huff's probability law

SKILL AREA
STATISTICAL AWARENESS

DESCRIPTIVE STATISTICS
Types of data, nominal, ordinal, etc.
Mean, mode, median
Range, standard deviation, etc.

INFERENTIAL STATISTICS
Sampling
Student's t-test
Chi squared test
Nearest neighbour analysis
Correlation – Spearman's
Pearson's
Linear regression

SKILL AREA
MAPS AND DIAGRAMS

Lettering, linework, shading
Dot maps
Isoline maps
Choropleth maps
Proportional symbols
Divided prop. symbols
Flow maps
Interpretation of OS maps
Rural land use mapping
River profiles
Cross sections, transects

Unit evidence (for **internal** assessment) requires the student to have evidence of:

1. Competence in one Physical Geography fieldwork method.
2. Competence in one Human Geography fieldwork method.
3. One piece of evidence to show that the student can produce a map and/or a diagram.
4. One piece of evidence to show that the student can extract and interpret information from a 1:25000 OS map.

About this learning pack

The number and titles of the study sections

This pack was written to help students complete Unit 1, *Geographical Methods and Techniques,* of Advanced Higher Geography. It contains all three skill areas that make up Unit 1. These are:

- fieldwork survey/measurement and recording techniques
- statistical awareness
- the production and interpretation of maps and diagrams.

The order of the study sections in the pack, however, is different from the order found in the 'Arrangements' document as it was felt that students should familiarise themselves with mapping, graphical and statistical techniques before embarking on fieldwork.

Study time within each section

As Unit 1, *Geographical Methods and Techniques*, is itself divided up into three skill areas and seven sections, it is suggested that the student begins the pack early in the academic session (if applicable) and stays within the recommended time limit, i.e. of 40 hours' duration. Most of this time (perhaps 25 hours) will be spent on the statistical awareness section. The rest of the time should be spent on the production and interpretation of maps and diagrams.

It might be advantageous to students if this pack was separated into the three skill areas so that they could work through these at different times. For example, students might wish to begin with the fieldwork methods at the beginning of the session when the weather is better and leave the mapping section to nearer the time of the external examination when practice at OS mapping would be an advantage.

Notes on Section 5

The exercises in Section 5 are similar to the 'planning question' (Q2) in the external examination. For simplicity, a suitable site for the location of the various buildings has been chosen. Please emphasise to students that in the examination **they** will have to choose the most suitable site. Further, only a selection of examples have been given as exercises; there are many other options, e.g. the location of picnic sites in a rural area, the location of a car par in the CBD or in a National Park – the possibilities are endless. Please consult past CSYS examination papers for additional ideas. The author has tried to include all the advantages of each of the sites in the exercises, although students may be able to find some more.

It is suggested in the T2 assessment that the student should be given a question from a past Advanced Higher or CSYS examination paper which can be used as evidence that he/she can extract and interpret information from 1:25000 OS maps.

Assessment evidence

At the end of each of the sub-sections are tutor-marked assignments which may be used as assessment evidence. It is suggested that the student keeps this evidence together with the candidate record sheet in a separate ring binder.

Explanation of the symbols used

The pack contains three main study sections, each corresponding to the three skill areas. The study sections are made up of the following:

Worked example	This gives a step-by-step explanation through a relatively complex statistical or graphical technique.
A	Indicates an activity to be undertaken with a number, indicating the number of the activity within each section.
A comment	All of the activities are followed by the answer and a comment which gives the student help and advice where appropriate.
T	Indicates a tutor assignment. Successful completion of these tutor assignments could provide enough evidence to show that the student can apply the appropriate techniques to given sets of data or can extract and interpret maps and diagrams. It is therefore suggested that the tutor assignments should be kept separate from the pack to provide evidence for internal assessment.

Other resources required

A bibliography is included at the end of this unit. All of these books are suitable for further work and study.

The HSDU publication *Fieldwork Methods and Techniques, Student Guide* is also an invaluable source of information to the student embarking on fieldwork.

Further, most Geography departments will have a supply of map-reading textbooks and fieldwork books in their departmental library which the tutor will be able to recommend to the student.

Tutor assignments – answers

T_2

1. Ordinal data (Elements of this data type can be placed in order but we cannot do arithmetic with them.)

2. Nominal data (They do not have any inherent order.)

3. Nominal data (Again, there is no inherent order.)

4. Both are ratio data (We can do arithmetic with elements of this data set, we can rank them and a zero value is a meaningful concept.)

T3

Part 1: Median, Mode, Range and Inter-Quartile Range

Calculate mean, median, mode, range, inter-quartile range, standard deviation and coefficient of variation for the annual precipitation at Glasgow. There are **21** values.

The data . . .

93 81 81 130 78 94 100 96 86 100 109 107 89 98 105 111 110 105 89 104 80

. . . ranked in order helps to find the mode . . .

78 80 **81** **81** 86 **89** **89** 93 94 96 98 **100** **100** 104 **105** **105** 107 109 110 111 130

There are in fact **4 Modes**: **81**, **89**, **100** and **105 (cm)**. The data is **multi modal**.

. . . the range . . . **Range = 130 – 78 = 52 (cm)**

. . . the median . . .

78 80 81 81 86 89 89 93 94 96 **98** 100 100 104 105 105 107 109 110 111 130

Median=98 (cm) ^ (middle value)

. . . Q1 and Q3, by splitting the data into two halves, using a 'shared' middle value:

78 80 81 81 86 **89** 89 93 94 96 98 ----- 98 100 100 104 105 **105** 107 109 110 111 130

Q1=89 (cm) ^ (middle value of lower half) **Q3=105 (cm)** ^ (middle value of upper half)

. . . and hence . . . the **Inter-Quartile Range = (105 – 89) = 16 (cm)**

See next page for a summary of the results.

T3

Part 2: Mean, Standard Deviation and Coefficient of Variation

Calculate mean, median, mode, range, inter-quartile range, standard deviation and coefficient of variation for the annual precipitation at Glasgow. There are **21** values.

	x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$
	93	97.43	–4.43	19.61
	81	97.43	–16.43	269.90
	81	97.43	–16.43	269.90
	130	97.43	32.57	1060.90
	78	97.43	–19.43	377.47
	94	97.43	–3.43	11.76
	100	97.43	2.57	6.61
	96	97.43	–1.43	2.04
	86	97.43	–11.43	130.61
	100	97.43	2.57	6.61
	109	97.43	11.57	133.90
	107	97.43	9.57	91.61
	89	97.43	–8.43	71.04
	98	97.43	0.57	0.33
	105	97.43	7.57	57.33
	111	97.43	13.57	184.18
	110	97.43	12.57	158.04
	105	97.43	7.57	57.33
	89	97.43	–8.43	71.04
	104	97.43	6.57	43.18
	80	97.43	–17.43	303.76
$\Sigma x =$	2046		$\Sigma(x-\bar{x})^2 =$	3327.14
$\bar{x} = \frac{\Sigma x}{n} =$	**97.43**		$\frac{\Sigma(x-\bar{x})^2}{n} =$	158.44
			$\sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} =$	**12.59**

Mean = 97.4 (cm) **Standard Deviation = 12.6 (cm)**

Coefficient of Variation $= 12.59 \times \frac{100}{97.43} =$ **12.92** (No units. The Coefficient of Variation is just a number.)

Summary:
Mean = 97.4cm
Median = 98cm
Modes = 81, 89, 100 and 105cm
Range = 52cm
Inter-Quartile Range = 16cm
Standard Deviation = 12.6cm
Coefficient of Variation = 12.9

T4

Student's t-test: litter dropped example

Amount of litter dropped by the public in two areas

	Column 1	Column 2	Column 3		Column 4	Column 5	Column 6
Site A **Beside the Car Park**	**(x)**	**$(x-\bar{x})$**	**$(x-\bar{x})^2$**	***Site B*** **At the Beauty Spot**	**(y)**	**$(y-\bar{y})$**	**$(y-\bar{y})^2$**
1	6	–3	9	1	11	6	36
2	13	4	16	2	4	–1	1
3	7	–2	4	3	1	–4	16
4	8	–1	1	4	6	1	1
5	7	–2	4	5	8	3	9
6	10	1	1	6	4	–1	1
7	12	3	9	7	1	–4	16
$\Sigma_x =$	63	$\Sigma(x-\bar{x})^2 =$	44	$\Sigma_y =$	35	$\Sigma(y-\bar{y})^2 =$	80
$n_x =$	7	$\frac{\Sigma(x-\bar{x})^2}{n_x} =$	6.29	$n_y =$	7	$\frac{\Sigma(y-\bar{y})^2}{n_y} =$	11.43
$\bar{x}$ =	**9**	$\sqrt{\frac{\Sigma(x-\bar{x})^2}{n_x}} =$	**2.51**	$\bar{y}$ =	**5**	$\sqrt{\frac{\Sigma(y-\bar{y})^2}{n_y}} =$	**3.38**
		this is σ_x				**this is σ_y**	

Standard Error (SE) of $x = \frac{\sigma_x}{\sqrt{n_x}} = \frac{2.5}{\sqrt{7}} = 0.948$

Standard Error (SE) of $y = \frac{\sigma_y}{\sqrt{n_y}} = \frac{3.38}{\sqrt{7}} = 1.278$

SE of x = 0.948

SE of y = 1.278

$(\text{SE of } x)^2 = 0.898$

$(\text{SE of } y)^2 = 1.633$

$\bar{x} - \bar{y} = 4$

$(\text{SE of } x)^2 + (\text{SE of } y)^2 = 2.531$

$\sqrt{(\text{SE of } x)^2 + (\text{SE of } y)^2} = 1.591$

$$t = \frac{(\bar{x}-\bar{y})}{\sqrt{(\text{SE of } x)^2 + (\text{SE of } y)^2}}$$

$$= \frac{(9-5)}{\sqrt{0.898 + 1.633}}$$

$$= \frac{4}{\sqrt{2.531}}$$

$$= \frac{4}{1.591}$$

Therefore **t = 2.514**

Note that **t** has a positive value in this example.

Since this is a two-tailed test it does not matter whether **t** is 'plus' or 'minus', we simply drop the sign when using the **t**-tables.

(Different tables are required for one-tailed tests and then the sign of **t** *does matter. This is beyond the scope of this course.)*

To look up the **t**-table we need:

> degrees of freedom = $n_x - 1 + n_y - 1 = 7 - 1 + 7 - 1 =$ **12**
>
> level of significance = **0.05**

giving a **critical value** from the table of **2.18**

The calculated value of **t** (ignoring sign) is 2.514 which is **greater** than the critical value from the table of **2.18**. Therefore, there is evidence to reject the null hypothesis and accept the alternative hypothesis.

A typical wording for the conclusion might be:

> There is evidence to support the hypothesis that there is a difference in the amount of litter dropped by the public in the two different areas of the country park.

Note that, in this example, both samples are the same size, that is $n_x = n_y$, but that is not a necessary condition of the t-test.

When the t-test is used, as here, to examine the difference between two sample means, the two samples can be of different sizes.

T5

Chi-square test of the relationship between shop type and settlement size

NH: **either** There is <u>no relationship</u> between shop type and settlement size.
or There is <u>no difference</u> in the types of shop found in settlements of different sizes.

AH: **either** There <u>is</u> a relationship between shop type and settlement size.
or There <u>is</u> a difference in the types of shop found in settlements of different sizes.

Required **significance level,** either **0.05** or **0.01**, should be stated at this stage.
For something like this, **0.05** would normally be appropriate, but as an exercise, the student has been asked to test at both levels and comment on the findings.

OBSERVED FREQUENCIES

	Shop Type				
Town	**Low Order**	**Middle Order**	**High Order**	**Row Total**	
A 10,000 pop.	16	5	3	**24**	
B 35,000 pop.	28	26	22	**76**	
Column Total	**44**	**31**	**25**	**100**	**(grand total)**

EXPECTED FREQUENCIES

	Shop Type				
Town	**Low Order**	**Middle Order**	**High Order**	**Row Total**	
A 10,000 pop.	11	7	6	**24**	
B 35,000 pop.	33	24	19	**76**	
Column Total	**44**	**31**	**25**	**100**	**(grand total)**

Note that although one **OBSERVED VALUE** is **less than 5**, none of the **EXPECTED VALUES** is **less than 5** and so the analysis is **VALID**.

$$\frac{(O-E)^2}{E}$$

	Low Order	Middle Order	High Order
A 10,000 pop.	2.80	0.80	1.50
B 35,000 pop.	0.88	0.25	0.47

2.80 0.80 … + … 0.25 0.47 = chi-square = **6.71**

degrees of freedom = (No. of rows – 1) x (No. of columns – 1) = 1 x 2 = **2**

chi-square critical value for **2** degrees of freedom and a significance level of **0.05** = **5.99**
chi-square critical value for **2** degrees of freedom and a significance level of **0.01** = **9.21**

The **calculated value = 6.71** is **greater than 5.99,** the 0.05 critical value
but **less than 9.21**, the 0.01 critical value

THUS: There is evidence to support the **AH**:
There is a relationship between shop type and settlement size.
or There is a difference in the types of shop found in settlements of different sizes.

BUT: If the chi-square value had been significant also at the 0.01 level, we could have said:
There is a strong relationship between . . .
or There is a big difference in . . .

T6

Nearest Neighbour Analysis

Village	Nearest Neighbour	Distance (d) in km	
A	F	1.5	
B	C	2.0	
C	B	2.0	
D	E	3.0	
E	B	3.0	
F	G	1.0	
G	F	1.0	
H	I	1.5	
I	H	1.5	
J	K	2.0	
K	J	2.0	
	Sum of d =	20.5	This is called σ_d

Nearest Neighbour Index (NNI) = $2 \times \bar{D} \times \sqrt{\left(\frac{N}{A}\right)}$

Where
N = Number of villages = 11
A = Area (in square kilometres) = 64 sq km

$\bar{D} = \frac{\sigma_d}{N} = \frac{20.5}{11} = 1.863$

Therefore, **NNI**

$= 2 \times 1.863 \times \sqrt{\frac{11}{64}}$

$= 2 \times 1.863 \times 0.415$

= **1.55**

This represents a significant degree of clustering

T7

Spearman's rank correlation

NH: Gross National Product is not related to the percentage agricultural workforce.
AH: Gross National Product is related to the percentage agricultural workforce.

For this example, the student has been instructed to use both the 0.01 and 0.05 levels of significance

	Country	GNP $/cap 1993	Rank (lowest to highest)	agricultural workforce %	Rank (lowest to highest)	d	d^2
1	Australia	17,510	9	6	4	5	25
2	Bangladesh	220	2	59	11	–9	81
3	Bolivia	770	7	47	10	–3	9
4	Brazil	3,020	8	25	8	0	0
5	China	490	5	73	13	–8	64
6	Egypt	660	6	42	9	–3	9
7	Ethiopia	100	1	88	15	–14	196
8	India	290	4	62	12	–8	64
9	Italy	19,620	11	9	7	4	16
10	Japan	31,450	14	7	6	8	64
11	Kenya	270	3	81	14	–11	121
12	Norway	26,340	13	6	4	9	81
13	Switzerland	36,410	15	6	4	11	121
14	UK	17,970	10	2	1	9	81
15	USA	24,750	12	3	2	10	100
							Σd^2 = 1032

Note that the three values of 6% (for agricultural workforce) are in the rank positions 3, 4 and 5 but because they all have the same value they must all be given the average rank, i.e. 4.

Formula: $$R(\text{Spearman}) = \frac{1 - (6 \times \Sigma d^2)}{n \times (n^2 - 1)}$$

Here we have:

$(6 \times \Sigma d^2) = 6192$

$n = 15$

$n^2 = 225$

$n^2 - 1 = 224$

$n \times (n^2 - 1) = 3360$

$$\frac{(6 \times \Sigma d^2)}{n \times (n^2 - 1)} = 1.843$$

$$\frac{1 - (6 \times \Sigma d^2)}{n \times (n^2 - 1)} = -0.843$$

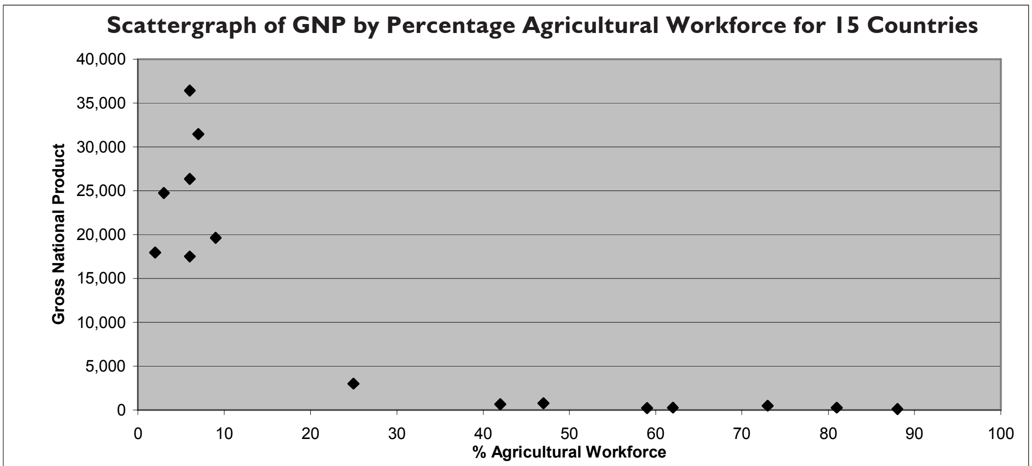
Note the negative sign: as GNP goes up, % agri workforce goes down, but it is still a strong correlation (close to –1).

Spearman's rank correlation coefficient = 0.843
degrees of freedom = no. of paired observations (n) = **15**

Spearman's rank critical value for **15** degrees of freedom and a significance level of **0.05 = 0.525**
Spearman's rank critical value for **15** degrees of freedom and a significance level of **0.01 = 0.689**

For the significance level of 0.05 . . .
The **calculated value** of **0.843** is **greater than** the **critical value** of **0.525** and so we can **reject the null hypothesis** in favour of the **alternative hypothesis**.
THUS: There is evidence to suggest that Gross National Product is related to percentage agricultural workforce.

For the significance level of 0.01 . . .
The **calculated value** of **0.843** is **greater than** the **critical value** of **0.689** and we have even stronger evidence in favour of the **alternative hypothesis**.
THUS: There is strong evidence to suggest that GNP is related to percentage agricultural workforce.



T8

Pearson's correlation coefficient and linear regression

Relationship between distance from source and pebble size

Use the 'dependency' statement: '*Average pebble size depends on distance from source*'

Therefore: y = average pebble size and x = distance from source

	1	2	3	4	5	6	7
	Distance from source (km)	**Average pebble size (mm)**					
Site no.	**x**	**y**	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
1	0.4	18	–1.6	9	2.56	81	–14.4
2	0.8	15	–1.2	6	1.44	36	–7.2
3	1.2	12	–0.8	3	0.64	9	–2.4
4	1.6	10	–0.4	1	0.16	1	–0.4
5	2	8	0	–1	0	1	0
6	2.4	4	0.4	–5	0.16	25	–2
7	2.8	7	0.8	–2	0.64	4	–1.6
8	3.2	5	1.2	–4	1.44	16	–4.8
9	3.6	2	1.6	–7	2.56	49	–11.2
sum x, y	18	81	sum cols 5, 6, 7 ->		9.6	222	**–44**
mean x, y	2	9	(sum col5) x (sum col6)			2131.2	**(numerator)**
			sqr root of fig above			**46.16**	
						(denominator)	

Pearson's correlation coefficient = –44 divided by 46.16 = –0.95

Degrees of freedom = 9 – 2 = **7**

Critical values from the table = **0.666** (0.05 level of significance)
0.798 (0.01 level of significance)

Our correlation coefficient of **0.95** (ignoring the sign) is greater than both critical values. Therefore we can support the AH that there is evidence of a <u>strong</u> relationship between average pebble size and distance from source of river.

Scattergraph and Line of Regression

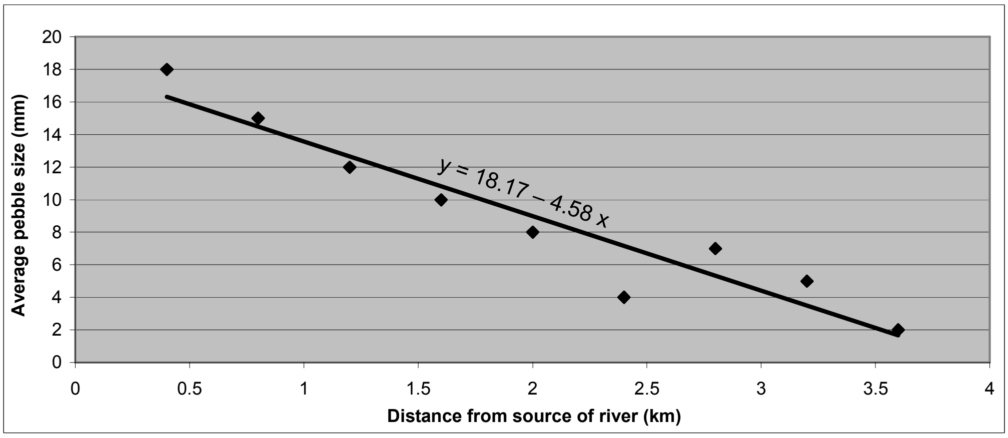
$$\mathbf{b} = \frac{\text{(sum of column 7)}}{\text{(sum of column 5)}}$$

$$= \frac{-44}{9.6}$$

$$= \mathbf{-4.58}$$

$$\mathbf{a} = \bar{y} - (b \times \bar{x})$$
$$= 9 - (-4.58 \times 2)$$
$$= 9 - -9.17$$
$$= \mathbf{18.17} \text{ (remember, minus minus = plus)}$$

Line of Regression is:
Average pebble size = 18.17 –4.58 x Distance from source of river



STUDENT GUIDE

Introduction

Advanced Higher Geography contains three units. They are:

Unit 1: Geographical Methods and Techniques
Unit 2: The Geographical Study
Unit 3: Geographical Issues.

In Advanced Higher Geography, you must study and apply geographical methods and techniques to analyse information, identify relationships and present information. This part of the pack, **Geographical Methods and Techniques**, gives you practice, advice and information in these areas.

It cannot be emphasised strongly enough that this unit gives you practice using statistical, mapping and graphical skills and techniques that you can use in Unit 2 (The Geographical Study) and Unit 3 (Geographical Issues). Therefore, as you work your way through this unit you should put some thought into how you can apply these skills and techniques to subsequent work.

Further, competence in the statistics section of this unit is a key element for the core skills of Numeracy and Problem-solving being automatically certificated at Higher level.

The unit assessment requires you to provide evidence of the successful application of these methods and techniques.

Unit evidence (for **internal** assessment) requires you to have evidence of:

1. Competence in one Physical Geography fieldwork method.
2. Competence in one Human Geography fieldwork method.
3. One piece of evidence to show that you can produce a map and/or a diagram.
4. One piece of evidence to show that you can extract and interpret information from a 1:25000 OS map.

Where you are undertaking the unit as part of a course, these skills will also be assessed through the **externally** set and assessed written paper of two hours' duration.

It is assumed that you have already passed Higher Geography, or equivalent, and are competent in OS map reading and interpretation and the examination of geographical methods and techniques at Higher level.

This unit is divided into three parts that correspond to the three skill areas that make up Geographical Methods and Techniques. These are:

- statistical awareness
- the production and interpretation of maps and diagrams
- fieldwork survey/measurement and recording techniques.

Statistical awareness has been further divided into three sub-sections:

Section 1: The graphical presentation of data
Section 2: Descriptive statistics
Section 3: Inferential statistics.

The production and interpretation of maps and diagrams has likewise been divided into three sub-sections:

Section 4: The design and layout of maps
Section 5: The interpretation of OS maps and related data
Section 6: Topographic analysis based on OS maps.

Fieldwork survey/measurement and recording techniques only represents one section (Section 7).

The three units which make up Advanced Higher Geography and the contents of the three skill areas of Geographical Methods and Techniques are shown diagrammatically overleaf.

ADVANCED HIGHER GEOGRAPHY

UNIT 1
GEOGRAPHICAL METHODS AND TECHNIQUES

UNIT 2
THE GEOGRAPHICAL STUDY

UNIT 3
GEOGRAPHICAL ISSUES

SKILL AREA
FIELDWORK METHODS

PHYSICAL TOPICS
Morphological mapping
Vegetation sampling
Slope analysis
Stream analysis
Meteorology in the local setting
Soil profiles and characteristics
Pebble analysis

HUMAN TOPICS
Urban land use mapping
Traffic, pedestrian and environmental quality surveys
Questionnaire design and implementation
Use of secondary sources
Reilly's law of retail gravitation
Huff's probability law

SKILL AREA
STATISTICAL AWARENESS

DESCRIPTIVE STATISTICS
Types of data, nominal, ordinal, etc.
Mean, mode, median
Range, standard deviation, etc.

INFERENTIAL STATISTICS
Sampling
Student's t-test
Chi squared test
Nearest neighbour analysis
Correlation – Spearman's
Pearson's
Linear regression

SKILL AREA
MAPS AND DIAGRAMS

Lettering, linework, shading
Dot maps
Isoline maps
Choropleth maps
Proportional symbols
Divided prop. symbols
Flow maps
Interpretation of OS maps
Rural land use mapping
River profiles
Cross sections, transects

Unit evidence (for **internal** assessment) requires you to have evidence of:

1. Competence in one Physical Geography fieldwork method.
2. Competence in one Human Geography fieldwork method.
3. One piece of evidence to show that you can produce a map and/or a diagram.
4. One piece of evidence to show that you can extract and interpret information from a 1:25000 OS map.

About this learning pack

In Advanced Higher Geography, you must study and apply complex geographical methods and techniques to analyse information, identify relationships and present information. This pack shows you the main geographical methods and techniques that you need to know in order to be able to do this.

The number and titles of the study sections
The pack contains the three skill areas which make up Unit 1. These are:

- statistical awareness
- the production and interpretation of maps and diagrams
- fieldwork survey/measurement and recording techniques.

The order of the study sections in the pack, however, is different from the order found in the 'Arrangements' document as it was felt that you should familiarise yourself with mapping, graphical and statistical techniques before embarking on fieldwork.

Study time within each section
As Unit 1, *Geographical Methods and Techniques*, is itself divided up into three skill areas and seven study sections, it is suggested that you begin the pack early in the academic session (if applicable) and stay within the recommended time limit, i.e. of 40 hours' duration. Most of this time (perhaps 25 hours) will be spent on the statistical awareness section. The rest of the time should be spent on the production and interpretation of maps and diagrams.

It might be advantageous to you if you separated this pack into the three skill areas and you could work through these at different times. For example, you might wish to begin with the fieldwork methods at the beginning of the session when the weather is better and leave the mapping section to nearer the external examination when practice at OS mapping would be an advantage.

Assessment evidence
This pack contains tutor-marked assignments which may be used as assessment evidence. It is suggested that you keep this evidence together with your record sheet in a separate ring binder.

Explanation of the symbols used

The study sections are made up of the following:

Worked example	This gives a step-by-step explanation through a relatively complex statistical or graphical technique.
A	Indicates an activity to be undertaken with a number, indicating the number of the activity within each section.
A comment	All of the activities are followed by the answer and a comment which give you help and advice where appropriate.
T	Indicates a tutor assignment. Successful completion of these tutor assignments could provide enough evidence to show that you can apply the appropriate techniques to given sets of data or can extract and interpret maps and diagrams. It is therefore suggested that the tutor assignments should be kept separate from the pack to provide evidence for internal assessment.

Other resources required

A bibliography is included at the end of this unit. All of these books are suitable for further work and study.

The HSDU publication *Fieldwork Methods and Techniques, Student Guide* is also an invaluable source of information to the student embarking on fieldwork.

Further, most Geography departments will have a supply of map-reading textbooks and fieldwork books in their departmental library that the tutor will be able to recommend to you.

SECTION 1

Introduction

We will start off with the easy part of the course! This section looks at advanced graphical techniques that you may be able to use in your geographical study. Obviously you should already be familiar with the use of pie charts, line graphs and bar graphs, but in the examination you may be asked to interpret and explain any patterns you can see in any of the following graphical techniques.

There is a tutor assignment at the end of the section (pp. 52–5).

This section is straightforward and should only take about 5 hours to complete.

The techniques in this section (with additional exercises) can be found in many standard fieldwork textbooks. Ask your tutor or consult the bibliography.

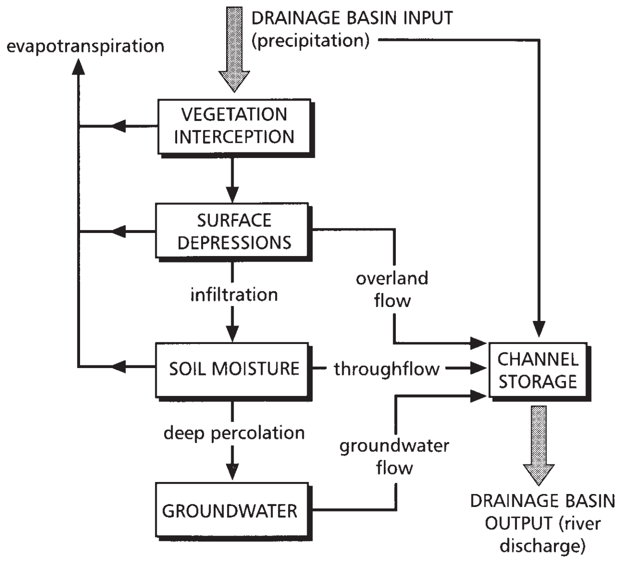
Systems diagrams

Systems diagrams are simplifying devices that can be used to communicate quite complex relationships. They usually comprise boxed annotations linked by lines or arrows. They are used to put across concepts and ideas, e.g. the hydrological cycle, the ecosystem or the spiral of unemployment.

Worked example

Describe a **drainage basin** as a system using the diagram below.

The drainage basin as a system



Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 48

Starting at the top of the diagram and working down the page, we can see that the **input** into the drainage basin is **precipitation**. Most of the precipitation falls onto the **vegetation** (you can tell this by the width of the arrow) and from there some of it is lost through **evapotranspiration**. Of the remainder, the water flows from the vegetation down into **surface depressions** where again some is lost through evapotranspiration and some becomes **overland flow** into **channel storage.**

The rest infiltrates into the soil to become **soil moisture** and again some is lost through evapotranspiration and some becomes **throughflow** into channel storage. Some of the soil moisture then percolates deep into the rocks to become **groundwater** which flows into channel storage.

Moving back to the top of the diagram again, it can be seen that some of the original **precipitation** directly becomes **channel storage**. The precipitation joins the **overland flow, throughflow** and **groundwater flow** to become the rather large volume of **drainage basin output** from the drainage system. (You can tell this by the width of the arrow.)

AI

In the space below, construct a systems diagram and describe the hydrological cycle. Use a Physical Geography textbook to help you with the diagram if your memory of the 'hydrosphere' from Higher Geography is poor! Explain your diagram.

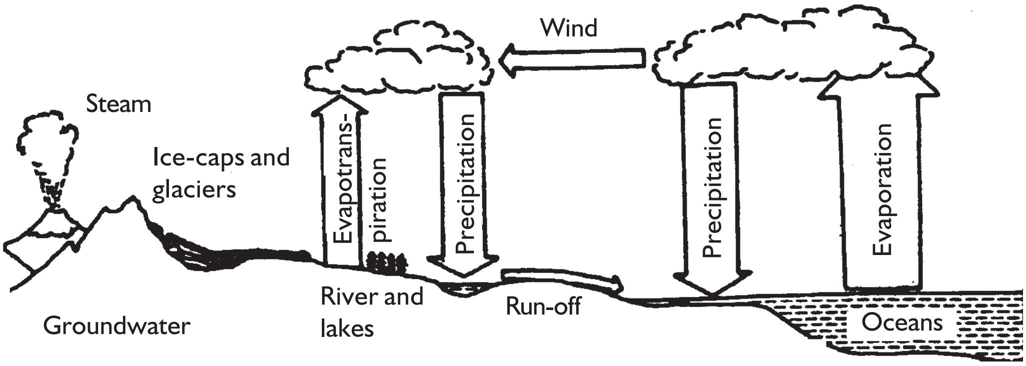
The Hydrological Cycle

Explanation

AI comment

Your diagram should look like this.

The Hydrological Cycle

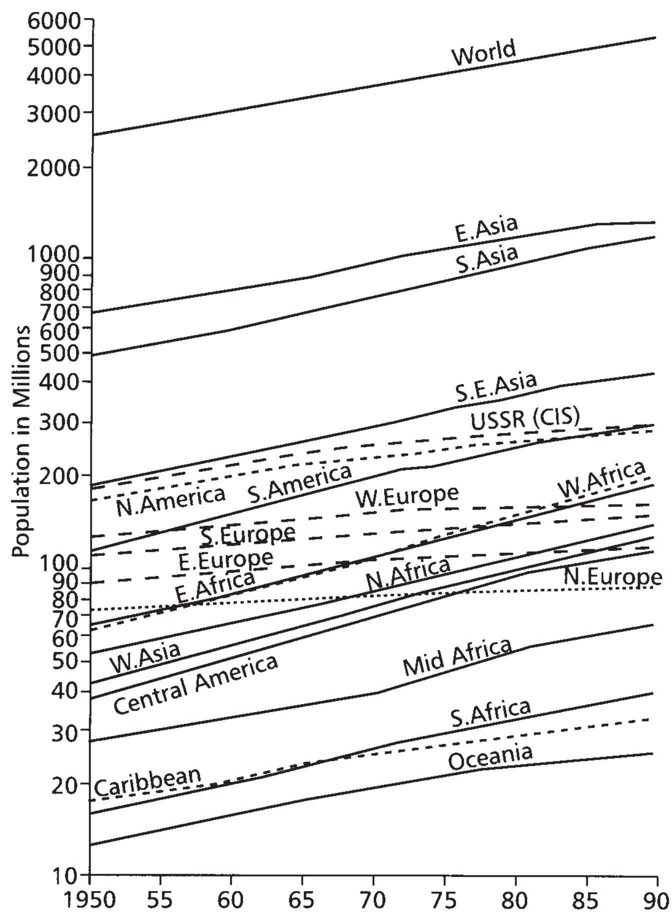


Explanation
There is constant recycling of water between the atmosphere, land and oceans. Precipitation from the atmosphere falls to earth where some of it is stored as snow and ice, or in lakes or as groundwater. The rest either evaporates into the atmosphere again or flows over the land as runoff (or through the soil and rocks as throughflow or groundwater flow). This water eventually ends up in the oceans where it evaporates back into the atmosphere to become water vapour. The water vapour falls back into the ocean as precipitation or is blown inland by winds where it condenses to form precipitation and the whole cycle begins again.

Logarithmic graphs

Logarithmic graphs are used to compare things that differ enormously in size, for example the population growth of the continents (see below). Logarithmic graphs are drawn on logarithmic graph paper where the logarithmic scale compresses the range of values. It gives more space to the smaller values and reduces the amount of space for the larger values. Thus it can show relative growth quite clearly. On the scale there are cycles of values. Each cycle increases by a larger amount, usually to the power of 10. Thus the first cycle increases by 10 each time up to 100, the second cycle increases by 100 to 1000 and so on.

World population growth



Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 51

A2

The rank-size rule

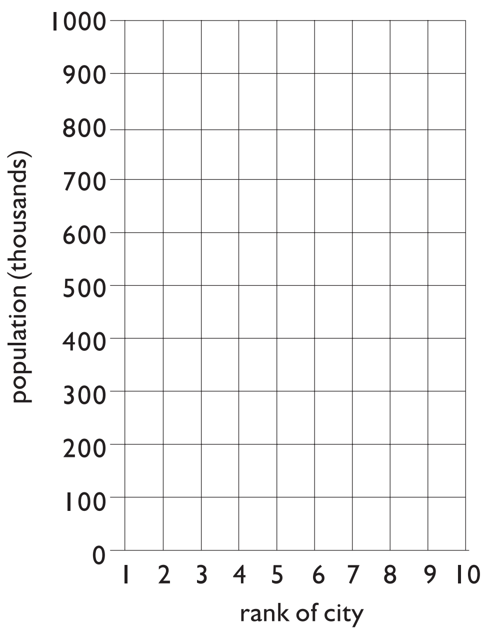
This rule states that within a country the settlements are ranked in descending order with the largest city placed first, the second largest should have half the population of the first, the third largest should have one third the population of the first, etc.

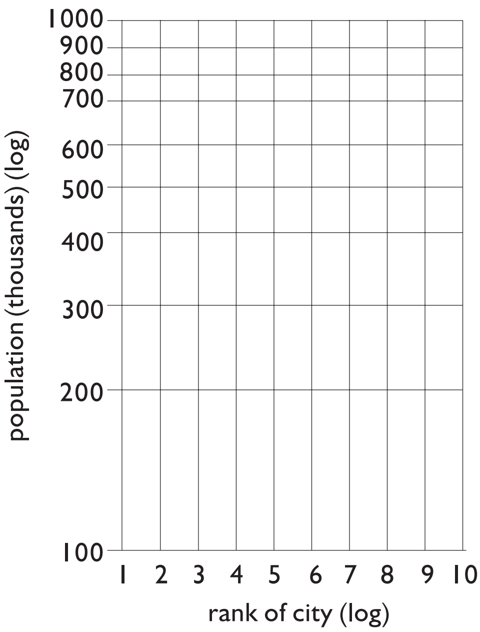
Look at the figures for the population of cities in a country. Plot these figures on graph paper (left diagram) and log graph paper (right diagram) and draw both of these lines. Compare the graphs.

Comparison

Population of cities (in thousands)

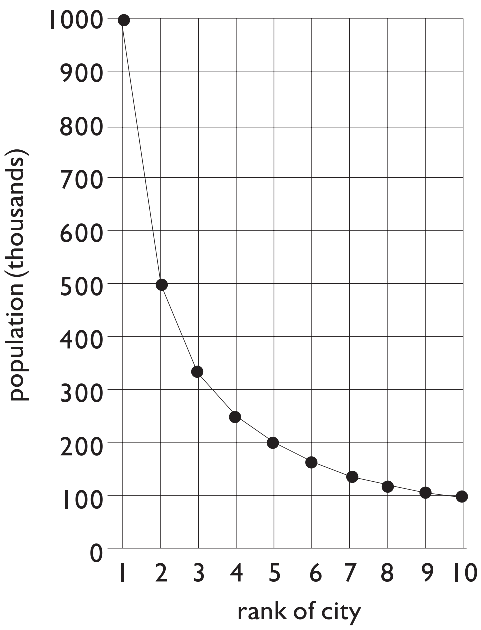
Rank	City	Population
1	A	1000
2	B	500
3	C	300
4	D	250
5	E	200
6	F	166
7	G	142
8	H	128
9	I	111
10	J	100

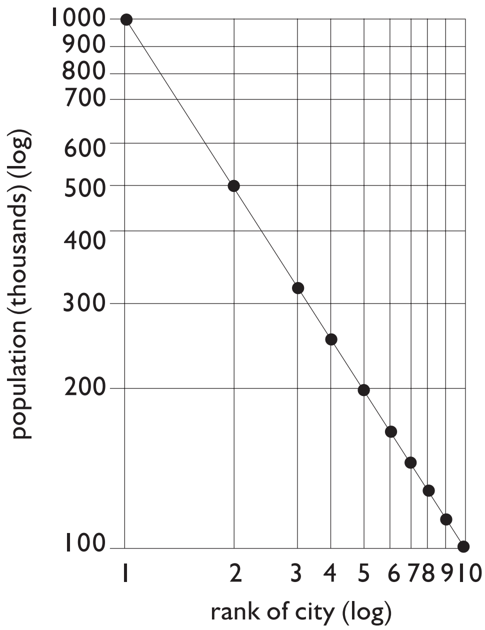




A2 comment

Your diagrams should look like this.





Comparison

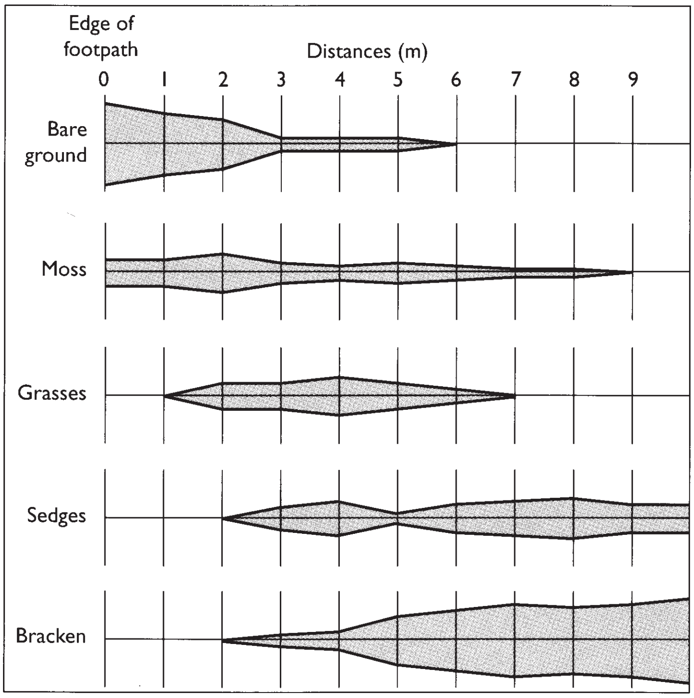
If the rank size rule was applied when the figures were plotted on graph paper, a steeply downward sloping curve would be produced (above left), but if they were plotted on graph paper with a logarithmic scale the relationship would appear as a more or less straight line (above right).

Kite diagrams

Kite diagrams are most useful for making observations of plants at regular points along a transect line. For example, the kite diagram below was drawn in the following way:

- Make a scale line for the distance covered in the survey (this is 10 metres in the diagram).
- One row is required for each vegetation type.
- Each row needs to have the same scale and be wide enough (and far enough apart from the others) to allow 100% for each plant.
- Draw a line through the middle of each row.
- At each survey point the value is plotted **on both sides of the line** down the middle of the row so that symmetry is achieved.
- The points are joined up for each row (which gives the diagram its kite appearance). Notice the gradual manner in which the kite tails off in both directions.
- Shade in between the kite lines.

Kite diagram showing vegetation cover from the edge of a footpath



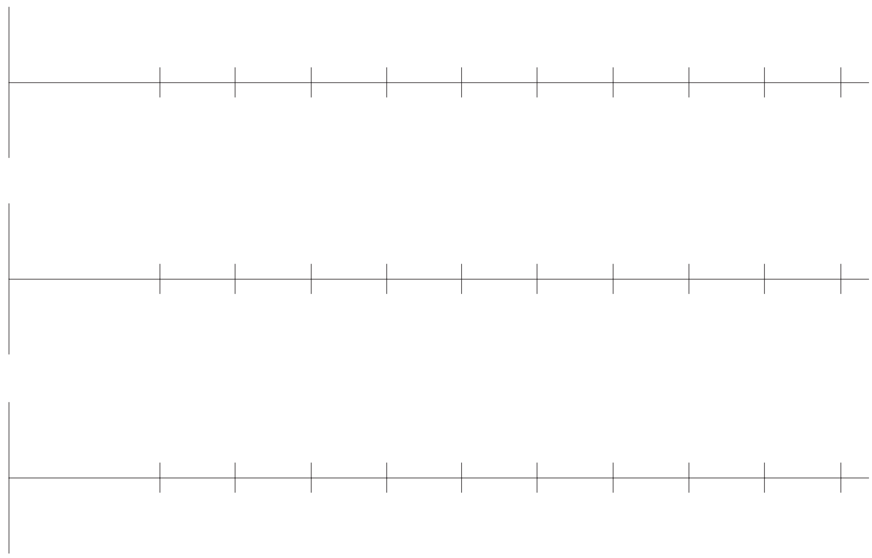
Source: *Tackling Geography Coursework*, Ann Bowen and John Pallister, p. 71

A3

sea ◄——————————► inland

Site	1	2	3	4	5	6	7	8	9	10
Species										
Couch grass	60	50	20	10						
Dandelion	10	10	20	30	50	30	20	10	10	5
Meadow grass	0	0	0	0	10	10	20	30	40	60

Draw kite diagrams to represent the frequency of each vegetation species found along a sand dune transect.

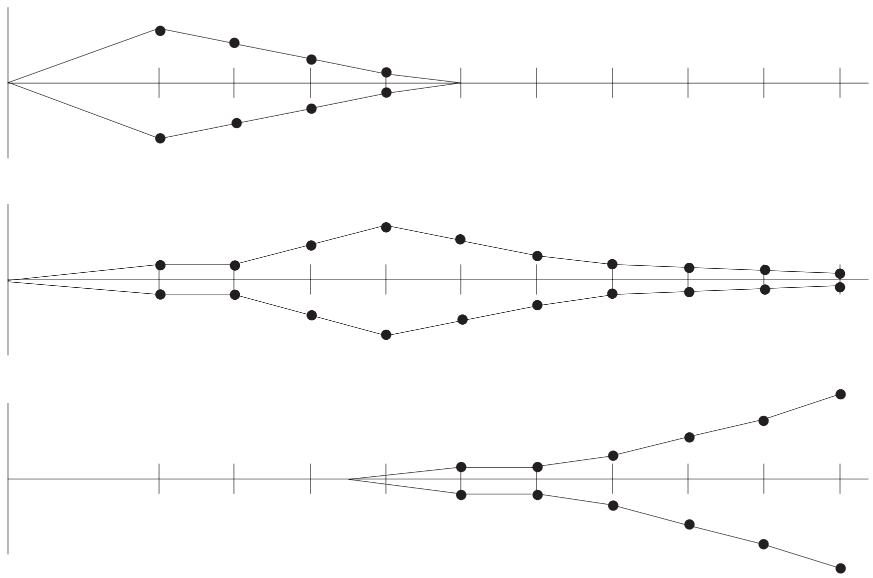


Explain why the kite diagram is a useful technique to apply in this context.

What relationships can you see between distance from the sea and vegetation types? (Hint : use your Higher Geography 'Biosphere' notes to help you.)

A3 comment

Your diagrams should look like this.



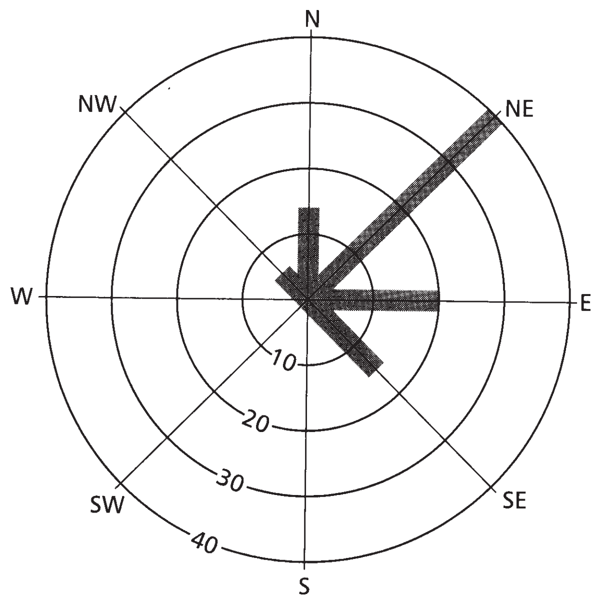
The kite diagram is a useful technique because it illustrates diagrammatically how the vegetation type changes along a sand dune transect with increasing distance from the sea.

Your answer should include the fact that the more hardy plants, which can withstand the arid, salty, sandy conditions, are located in more exposed areas near to the sea. On moving inland the vegetation type changes to less hardy species as the water and humus content in the soil increases, the soil becomes more acid and the area is more protected from the sea breezes.

Polar graphs

A polar graph or rose diagram is used to show direction as well as magnitude. For example, the diagram below shows that most corries in the Lake District face northwards and eastwards, and by reading off the scale (located between the south and south-west points) it can be seen that 40 corries face in a north-easterly direction, 20 in an easterly direction and 15 in a northerly direction.

The orientation of corries in the Lake District



Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 54

Polar graphs are easy to construct.

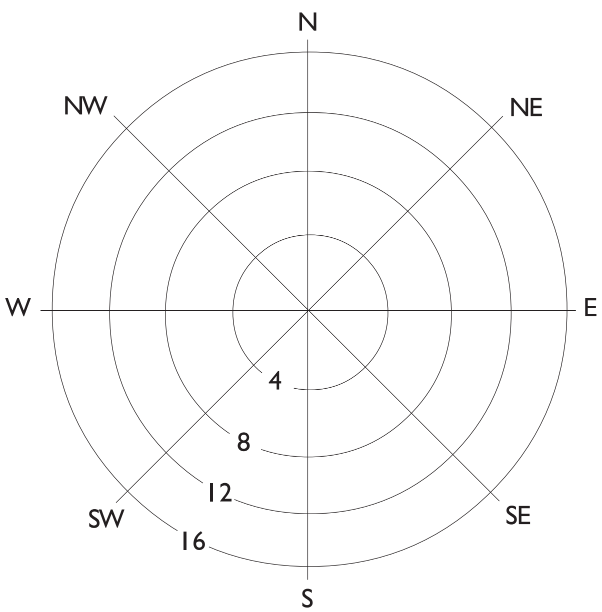
- Using a compass and protractor, draw in lines that correspond to north, north-east, east, south-east, south, south-west, west and north-west.
- Draw a scale. The scale relates to the radius of the diagram, e.g. in the above diagram 1 cm represents 10 corries. Circles are drawn at 1 cm intervals.
- Mark the scale on the vertical axis (north axis). In the diagram above, the scale has been marked between south and south-west.
- Plot the data for each directional sector.
- To make the diagram more detailed the direction bar can be divided. In the case of a wind rose, it may show the percentage frequency of different wind speeds.

A4

The orientation of the long axis of pebbles in glacial deposits gives an idea of the direction in which the glacier moved.

From the following information, complete the polar (rose) diagram and describe the direction in which the glacier moved.

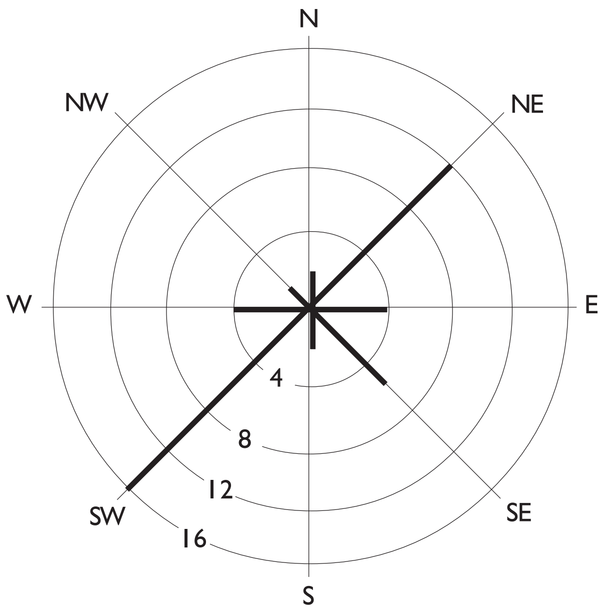
Direction	**number of stones**
North	2
North-east	12
East	4
South-east	6
South	2
South-west	16
West	4
North-west	2



Answer

A4 comment

Your diagram should look like this.



Answer

This should state that the predominant movement of the glacier is north-east to south-west.

Triangular graphs

Triangular graphs are used to show data that can be divided into three parts. This might include data such as soil (sand, silt and clay), employment (primary, secondary and tertiary) and housing (terraced, detached, semi-detached); see below.

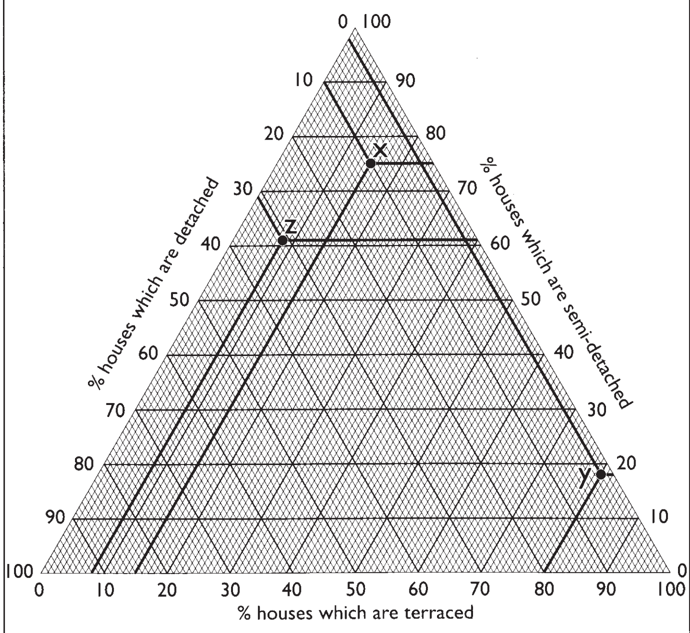
The triangular graph obviously has three sides! In the diagram below the x-axis has the percentage of terraced houses and the percentage **increases** from left to right. The left-hand side has the percentage of detached houses and the percentage **decreases** from the bottom of the triangle to the top. The right-hand side has the percentage of semi-detached houses and the percentage **increases** from the bottom of the triangle to the top.

Triangular graphs are tricky to construct – it is easy to get confused! However, the main advantage of the triangular graph is that it allows a large amount of data to be shown on one diagram and, in many cases, once plotted value groupings become evident.

Survey of the housing type in an area

Survey of 100 houses in each of three areas

Area X		*Area Y*		*Area Z*	
Terraced	15	Terraced	80	Terraced	8
Semi-detached	75	Semi-detached	18	Semi-detached	61
Detached	10	Detached	2	Detached	31



Source: *Tackling Geography Coursework*, Ann Bowen and John Pallister, p. 55

A5

Look at the triangular graph on the next page, and attempt the following:

a. Read off the values for soils A, B, C and D.

A ____________________

B ____________________

C ____________________

D ____________________

b. Plot and match the correct names to the following samples:

Sample	% clay	% silt	% sand	soil type
E	80	10	10	
F	45	5	50	
G	5	90	5	
H	5	5	90	
I	34	36	30	

Soil type

E ____________________

F ____________________

G ____________________

H ____________________

c. Answer the following:

i) Which type of parent rock is likely to produce soil H?

____

ii) How was soil G likely to have been formed?

____

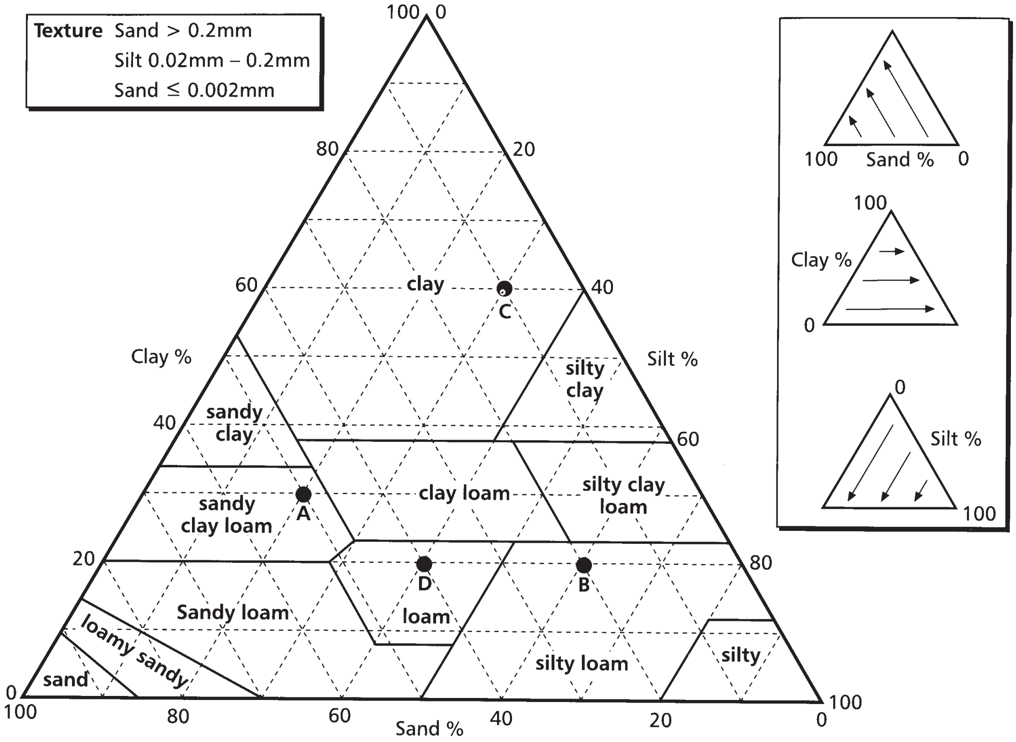
iii) Which of the five soil samples would be: (a) free draining, (b) difficult to drain?
In both cases give a reason for your answer.

____

iv) Which of the soil samples is the most fertile? Give a reason for your answer.

____

Diagram showing soil type



Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 56

A5 comment

a.

Sample	% clay	% silt	% sand
A	30	20	50
B	20	60	20
C	60	30	10
D	20	40	40

b.

Sample	% clay	% silt	% sand	***soil type***
E	80	10	10	*clay*
F	45	5	55	*sandy clay*
G	5	90	5	*silty soil*
H	5	5	90	*sandy soil*
I	34	36	30	*clay loam*

c. i) The type of rock is sandstone, due to the high % of sand in the soil.

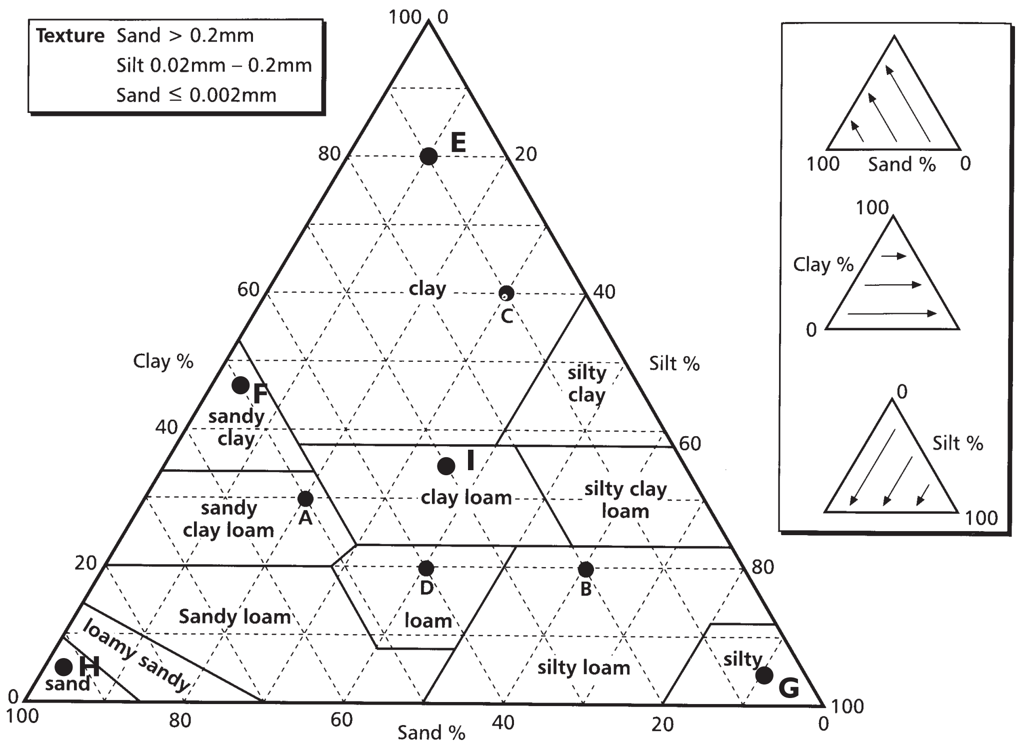
ii) Because of the high percentage of silt, this soil is likely to have been formed by rivers flooding and depositing silt on the floodplain.

iii) a) Soil H with the highest percentage of sand, would be the most free draining, because the large spaces between the sand grains let water percolate through easily.

b) The most difficult to drain would be soil E with its high percentage of clay. Clay particles bind together and make it difficult for water to percolate through.

iv) Sample I is the most fertile soil. It is a clay loam and has all the best attributes of the three soil types.

Diagram showing soil types



Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 56

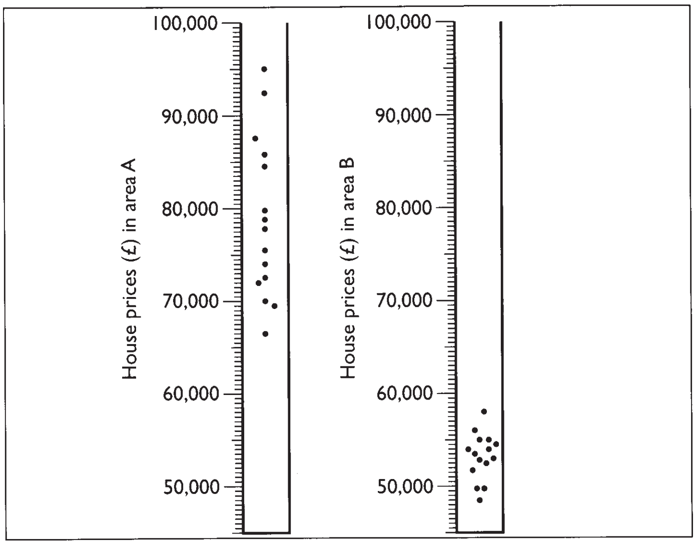
Dispersion diagrams

A dispersion diagram shows the range of a set of data and their tendency to group or disperse. It may also be used for comparing two groups of data. For example, the box below contains raw data listing prices for the same type of house in two areas in the same town. From this a dispersion diagram can be drawn.

- Make two columns – one for area A and one for Area B.
- Choose a scale that covers the range of values (with this method you need not start with zero if it is not required).
- Plot each house column in the range with a dot.

(Note from this diagram that the median, upper and lower quartiles and inter-quartile range can be determined but do not worry about this here – you will find out about quartiles in the next section.)

Dispersion diagram showing house prices in two areas



Source: *Tackling Geography Coursework*, Ann Bowen and John Pallister, p. 66

A6

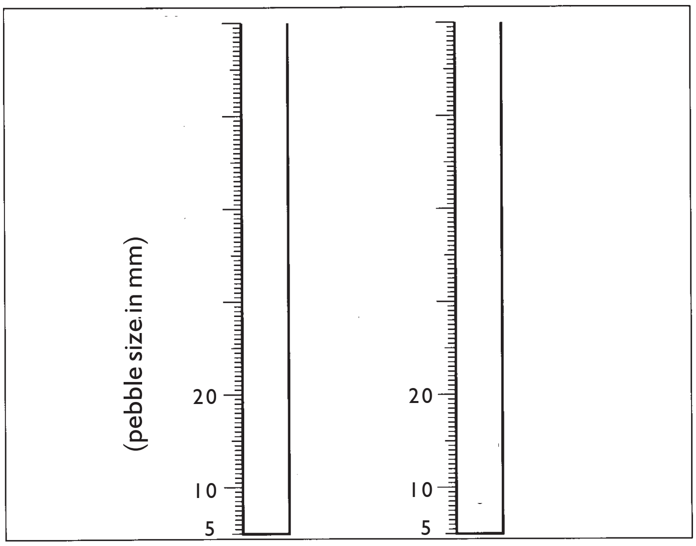
Two samples of stones were taken from a river in Arran. Sample A was taken from the upper course and sample B was taken from the lower course. The stone sizes (in mm) were as follows:

Sample A: 43, 30, 39, 56, 32, 46, 40, 19, 45, 26, 36, 38, 43, 20, 51, 45, 32, 45, 21, 34, 48, 23, 24, 37, 43, 27, 56, 37, 32.

Sample B: 23, 18, 27, 16, 12, 20, 31, 23, 18, 26, 31, 12, 9, 25, 13, 22, 23, 12, 17, 26, 10, 9, 25, 31, 25, 30, 16, 19, 12, 19.

a. Draw two dispersion diagrams on the frames provided below.

Sample A **Sample B**

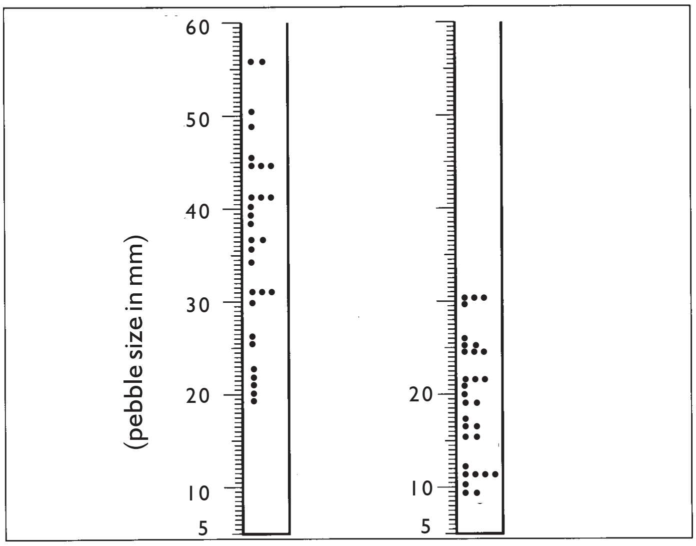


b. Describe and account for the differences in stone size.

A6 comment

a. The two dispersion diagrams can be seen on the frames below.

Sample A **Sample B**



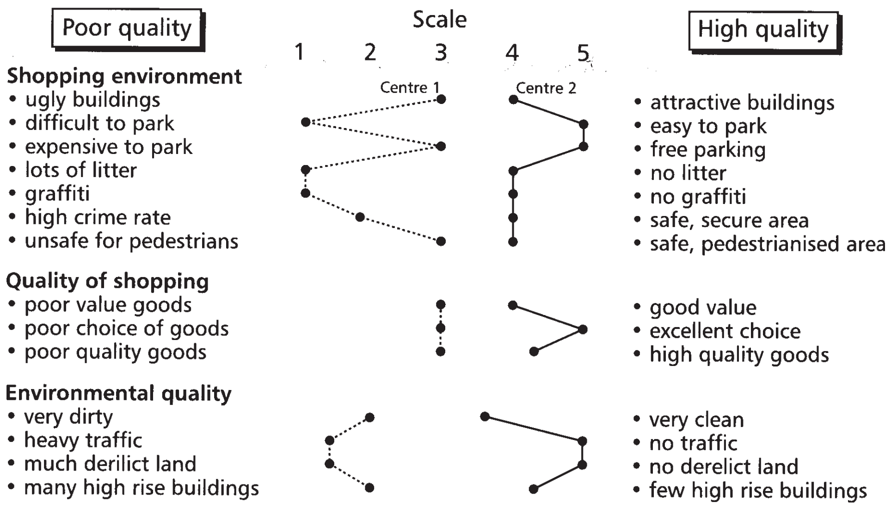
b. The stones in sample B are generally smaller in size than the stones in sample A. In sample A most of the stones are distributed around the 30–49 mm intervals, whereas in sample B the stones are found mainly in the 10–29 mm intervals.

The reason for these results is due to the fact that as the river enters its lower course its speed begins to fall away and it no longer has the energy to carry its load. Therefore the larger stones are deposited first then the smaller ones next.

Bipolar analysis

Bipolar analysis is a technique for comparing things. The example below shows how it can be used visually, using questionnaire results.

- In a questionnaire survey fifty shoppers were asked to rate two different shopping centres according to a range of categories and on a scale of 1– 5, with 5 being the highest score.
- Once the results have been collated the **average value** for each category is calculated. This is done by adding up the scores given by the shoppers for each of the categories and then dividing by the total number of shoppers.
- A dot showing this average value is then placed on the diagram (see below).
- Once this has been done for every category the dots are joined together.
- A visual comparison is then made by looking at the lines.



Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 62

The diagram above clearly shows that the new shopping centre (Centre 2) is perceived to be more attractive than the established centre (Centre 1). The main aspects of difference are the provision of parking spaces, cleanliness, safety, pedestrianisation, and variety of shops. Centre 1 contrasts unfavourably for such things as graffiti, litter, poor parking, high crime rate and concern about personal safety.

A7

An environmental quality survey was undertaken in two areas of a town. Area A was an inner-city area with mixed housing and industry and area B was a suburban housing area. The results for each area are shown below; a score of 1 in a row means the area is bad, 5 means it is good.

Look at the survey scores for the two areas (A and B) and in the third frame on the next page, draw a bipolar diagram of both sets of results.

Compare the two areas by looking at their respective results and explain the differences.

Area A: Inner city

Poor	***1***	***2***	***3***	***4***	***5***	***Good***
Poor air quality		x				Clean air
Lots of litter	x					Clean
Poor transport links		x				Good transport links
Poor-quality buildings	x					Good-quality buildings
Crime, vandalism		x				No evidence of crime
Unpleasant to look at	x					Pleasing to look at
Inadequate amenities	x					Many amenities
Not enough open space	x					Open space

Area B: The suburbs

Poor	***1***	***2***	***3***	***4***	***5***	***Good***
Poor air quality					x	Clean air
Lots of litter					x	Clean
Poor transport links			x			Good transport links
Poor-quality buildings				x		Good-quality buildings
Crime, vandalism				x		No evidence of crime
Unpleasant to look at					x	Pleasing to look at
Inadequate amenities				x		Many amenities
Not enough open space					x	Open space

Bipolar analysis

Poor	***1***	***2***	***3***	***4***	***5***	***Good***
	(A)				(B)	
Poor air quality						Clean air
Lots of litter						Clean
Poor transport links						Good transport links
Poor-quality buildings						Good-quality buildings
Crime, vandalism						No evidence of crime
Unpleasant to look at						Pleasing to look at
Inadequate amenities						Many amenities
Not enough open space						Open space

Comparison of the two areas:

A7 comment

Your diagram should look like this.

Poor	***1***	***2***	***3***	***4***	***5***	***Good***
		(A)			(B)	
Poor air quality		x			x	Clean air
Lots of litter	x				x	Clean
Poor transport links		x	x			Good transport links
Poor-quality buildings	x			x		Good-quality buildings
Crime, vandalism		x		x		No evidence of crime
Unpleasant to look at	x				x	Pleasing to look at
Inadequate amenities	x			x		Good amenities
Not enough open space	x				x	Open space

Comparison

From the bipolar diagram it can be seen that this inner-city area does not boast a particularly pleasant environment. The air quality is poor and there is lots of litter around. There are poor transport links, the buildings are of poor quality and are unpleasant to look at. The whole area suffers from crime and vandalism. There is not enough open space and the amenities for the people who live there are inadequate.

In comparison the suburbs have clean air and a clean environment. The transport links, however, are only reasonable and the buildings are of reasonable quality with less evidence of criminal activity. The whole area is pleasing to look at with open space and an appropriate level of amenities.

Tutor assignment

T1

You should submit this work to your tutor as it will be formally assessed. Success in this assignment means you have evidence that you can understand the graphical presentation of data.

1. **Systems diagrams**

   Draw a diagram to show the farming system. Your diagram should include the farm's physical and human inputs and its outputs. Describe and explain the diagram.

2. **Logarithmic graphs**

   a. On semi-logarithmic graph paper, plot the following points and join together the plotted points.

Population growth of selected areas

Area	**1950**	**1960**	**1970**	**1980**	**1990**
World	2500	2700	3000	3500	3900
E. Asia	690	750	800	1000	1200
S. Asia	500	550	700	800	1000
N. America	170	180	185	200	210
S. America	110	140	160	180	200
W. Europe	120	150	155	160	160
E. Africa	65	75	90	120	150
Mid Africa	29	30	35	42	55
Oceania	6	10	13	20	22

   b. Describe and explain the pattern shown.

   c. What are the advantages and disadvantages of using logarithmic graph scales?

3. **Kite diagrams**

a. Draw kite diagrams on graph paper to display the following information collected from a transect along sand dunes in S. England.

sea ◄────────────────────► inland

Site (metres inland)	**0**	**20**	**40**	**60**	**80**	**100**	**120**	**140**	**160**	**180**	**200**	**220**	**240**	**260**
Species														
Couch grass		2												
Marram grass			4	5	1	1		2	2					
Bell heather			1	5	5	6	2	1						
Ling						5	7	4	7	8				
Small trees												2	3	2

b. Explain why the kite diagram is a useful technique to apply in this context.

a. What relationships can you see between distance from the sea and vegetation type?

4. **Polar graphs**

a. Draw diagrams to display the information from each of the three weather stations below:

Weather station:	**A**	**B**	**C**
Location:	Shetland Isles (N. Scotland)	Cornwall (S.W. England)	E. Anglia (E. England)
Wind direction:	no. of days	no. of days	no. of days
N	3	8	5
NE	2	4	3
E	0	3	2
SE	1	0	1
S	7	3	4
SW	9	2	8
W	6	9	4
NW	2	2	4

b. Describe and account for any similarities and differences in the wind roses.

5. **Triangular graphs**

a. Plot the following points on triangular graph paper:

Country	% primary workforce	% secondary workforce	% tertiary workforce
Bangladesh	59	13	28
Brazil	25	25	50
Cambodia	74	7	19
Cameroon	79	4	14
Ethiopia	88	2	10
Italy	9	32	59
UK	21	39	40
USA	3	25	72

b. Account for any pattern of clustering you can see on the graph.

6. **Dispersion diagrams**

The figures below in Groups A and B show average life expectancies of the population for two groups of countries.

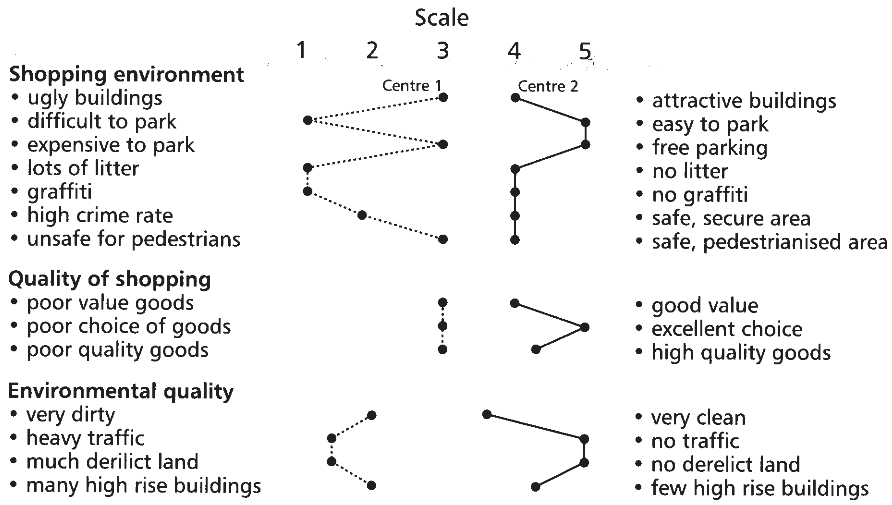
a. Choose a scale which covers the range of values and draw two dispersion diagrams.

Group A 80, 79, 75, 70, 74, 74, 75, 78, 77, 81, 80, 78, 76, 74.
Group B 56, 55, 60, 62, 40, 45, 46, 50, 52, 55, 59, 53, 59, 60.

b. Describe and account for the differences shown in the two groups.

7. **Bipolar analysis**

Customers were asked to compare two shopping centres. One was a traditional 'High Street' shopping centre and the other was a covered-in shopping mall. The results of the questionnaire are shown below.



a. Which is the High Street shopping centre and which is the shopping mall? Give reasons for your answer.

b. Describe fully what each centre looks like from the responses given.

SECTION 2

Introduction

Over recent years the understanding of statistics has been extremely useful in a great many areas of study and Geography is no exception. This section will introduce you to the use of statistics in Geography.

In the 1960s Geography adopted a more scientific approach. This meant that statistical techniques were used to test hypotheses (hunches) and theories **quantitatively** (i.e. using numbers). At first it seemed that mathematics had taken over the subject, but it is now accepted that by using statistical techniques we can identify patterns, discover relationships and examine cause–effect links. However, it is essential to select the most appropriate statistical techniques for use in the interpretation of geographical data. Therefore, it is important to have some understanding of statistical methods.

There are many types of statistics. At the most basic level there are **descriptive statistics** that **summarise data**, while the more advanced **inferential statistics** look at **relationships** between sets of data.

This section of the unit is divided into two parts. The sections on **descriptive statistics** and **inferential statistics** include all of the statistical methods you need to know for Advanced Higher Geography. Below is a list of the topics in each section.

Descriptive statistics

1. **Types of data:** nominal, ordinal, interval, ratio.
2. **Measures of central tendency:** the mean, median, mode.
3. **Measures of dispersion:** range, inter-quartile range, standard deviation, variance, coefficient of variation, standard error of the mean of a sample.

Inferential statistics

1. **Basic concepts:** sampling, hypothesis testing, levels of significance.
2. **Student's t-test:** a statistical test used to compare the mean of a sample with another mean.
3. **The chi-square test:** the test we use to measure the difference between what is observed and what is expected according to an assumed hypothesis.
4. **Nearest neighbour analysis:** a method by which spatial patterns can be measured objectively, e.g. it can be used to identify a tendency towards clustering or dispersion.

5. **Spearman's rank correlation coefficient:** used to test the degree of relationship between two variables by ranking the variables.
6. **Pearson's correlation coefficient:** uses figures instead of ranking in its calculations to test the degree of relationship between two variables.
7. **Linear regression:** using this technique we can calculate, mathematically, how to draw the best straight line through points on a scatter graph and examine relationships between variables in more detail.

Sections 2 and 3 are more time consuming than Section 1. They will probably take you 20–25 hours to complete. You will need a scientific calculator or even better a computer (if available) to help you with your calculations!

Tutor assignments are located at intervals through Sections 2 and 3.

Many of the techniques in Sections 2 and 3, with additional exercises, can be found in fieldwork textbooks or geographical techniques textbooks. Ask your tutor or consult the bibliography.

You may be able to use some of these techniques in your Geographical Study. Some of the examples used here have been taken from fieldwork exercises to give you an idea of where they can be used. In addition, remember that question 4 of the Advanced Higher is based on statistical techniques. Ask your tutor if you can see a past paper to get an idea of the layout of the question. It is more straightforward than you think!

Data sets

In any project or enquiry it is important to use descriptive statistics as they provide invaluable summaries of data and are often the only statistics that are appropriate. Before studying descriptive statistics, we need to know about different **types of data**.

There are four different types of data:

- **Nominal data** refers to data that have names. For example, in Geography we might categorise land use as commercial, residential, agricultural and recreational; or rock types as sedimentary, igneous and metamorphic.

- **Ordinal data** refers to data that can be placed in ascending or descending order. For example, settlement type: city, town, village and hamlet.

- **Interval data** refers to real numbers. With interval data there is no true zero. For example, if the temperature for Oxford is 13 degrees centigrade and the temperature for Miami is 26 degrees centigrade, it is not strictly correct to say that Miami is twice as warm as Oxford for if we had measured the two temperatures in Fahrenheit they would be 59 and 79 degrees respectively. [*Interval data is very uncommon. Temperature is the only example you are likely to come across*.]

- **Ratio data** also refers to real numbers but this time they do possess a true zero, e.g. it is possible to have 0 mm of rainfall and in this case Miami with 2100 mm of rainfall is three times as wet as Oxford with 700 mm of rainfall. [*Most of the numerical data you will use will be ratio data*.]

A1

Answer the following questions in the spaces provided, giving reasons for your answers in all cases.

What types of data are the following?

1. low order services, middle order services, high order services.
2. 12°C, 3°C, 16°C, –3°C.
3. earth, air, fire, water.
4. 2 hectares, 7 hectares, 1 hectare, 10 hectares.
5. Primary, secondary, tertiary, quaternary.
6. 30 mm, 75 mm, 10 mm, 3 mm.
7. arable, pastoral, plantation, subsistence.
8. 32°F, 60°F, 12°F, 100°F.

1. ______________________________
2. ______________________________
3. ______________________________
4. ______________________________
5. ______________________________
6. ______________________________
7. ______________________________
8. ______________________________

A1 comment

- Q3: earth, air, fire, water, and Q7: arable, pastoral, plantation, subsistence are both examples of **nominal** data as both data sets have **names** but cannot be placed in order. This is the most basic type of data.

- Q1: low order services, middle order services, high order services, and Q5: primary secondary, tertiary, quaternary are examples of **ordinal** data because the data can be ranked in order.

- Q2: 12°C, 3°C, 16°C, –3°C and Q8: 32°F, 60°F, 12°F, 100°F are examples of **interval** data. They both refer to temperatures, one in degrees Celsius and the other in degrees Fahrenheit. The difference between these and ratio data is that there is **no true or meaningful zero**. (£0 means you have no money, but 0°C doesn't mean there is no temperature!)

- Q4: 2 hectares, 7 hectares, 1 hectare, 10 hectares and Q6: 30 mm, 75 mm, 10 mm, 3 mm are both examples of **ratio** data. These are **real numbers** which **possess a proper zero**.

Central tendency

Now we can look at ways of **summarising data** using statistics. One method of summarising data is to calculate some sort of **average**. This is sometimes referred to as a measure of **central tendency**, giving one figure to describe a complete data set. The measurements used for central tendency include the **mean**, the **median** and the **mode**.

- **Mean** – this is the same thing as the average.
- **Median –** this is the middle value when all the values are placed in ascending (or descending) order.
- **Mode** – this is the most frequently occurring value.

The mean
One of the most commonly used statistics is the mean. This is found by totalling the values for all observations (Σx) and dividing by the total number of observations (n).

The formula for finding the mean is: $\text{Mean} = \frac{\Sigma x}{n}$

The median
Another method to use is the median. This is the middle value when all the data are placed in ascending or descending order. Where there are two middle values we take the mean of these.

The median is the middle value when all values are placed in ascending or descending order.

The mode
Another measure of central tendency is the mode.

The mode is the value that occurs most often.

Sometimes there are two (or more) modes. When there are two modes the data are said to be bi-modal. (If each of the values in the data occurs only once, the data has *no mode*.)

Selecting the correct method of central tendency

In many cases it is better to use one method of central tendency rather than the others. For example:

- If a student sits four Geography exams the best measurement of central tendency to report on would be the **mean** of the four exam marks as this would give a good picture of the student's overall performance.

- If data was published on the selling price of homes in a particular town, the best measurement of central tendency to use here would be the **median** price. This is because it does not take into account the relatively few homes with very high or very low selling prices which could make the mean too high or too low to be a good measurement of central tendency.

- If data provides a frequency distribution the only suitable measurement of central tendency would be the **mode**. For example look at the data below:

  Number of primary industries in Area X = 2
  Number of secondary industries in Area X = 12
  Number of tertiary industries in Area X = 34

  The mode here would be tertiary industry. There is no way you can compute the mean or median of such data. Similarly, the mode is the only measure of central tendency that can be used for nominal (named) or ordinal (ranked) data.

A2

Answer the following questions in the spaces provided on the following page.

1. The weekly pocket money for nine first-year secondary school students was found to be: £3, £12, £4, £6, £1, £ 4, £2, £5, £8. What is their mean weekly pocket money?

2. Find the average daily rainfall from the following measurements that were taken each day over a one-week period: 6, 2, 7, 6, 9, 0, 5 (mm).

3. Find the median of the data in Q1 and Q2.

4. Find the mode of the pocket money from the list of figures given in Q1 and the mode of the rainfall from the list of figures given in Q2.

5. Which measure of central tendency is the best one to use in the following three examples? In each case give a reason for your answer.

   i) A distance runner entered seven marathons. His times were 3 hours 45 minutes, 4 hours 05 minutes, 3 hours 55 minutes, 4 hours 25 minutes, 4 hours 20 minutes, 3 hours 48 minutes, 3 hours 55 minutes. He wanted to find his overall performance.

   ii) In the 1992 Boston marathon there were two categories of official finishers: male and female. The following table provides a frequency distribution for that data:

male	**female**
6562	1562

   Which measurement of central tendency should be used here?

   iii) The average wage of the employees in an office from the managing director to the office junior.

1.

2.

3.

4.

5.

i)

ii)

iii)

A2 comment

1. The mean weekly pocket money for the nine pupils is $\frac{45}{9} = £5$.

2. The average rainfall is $\frac{35}{7} = 5$ mm.

3. The median for Q1 is
   1 2 3 4 **4** 5 6 8 12.
   Place the values in ascending order and select the middle value.

   The median for Q2 is
   0 2 5 **6** 6 7 9.

4. The mode of Q1 is 4 and the mode of Q2 is 6. Both of these numbers appear most often.

5. i) The **mean** time of the seven marathons, as this would give a good picture of his overall performance.

   ii) The **mode** would be used here. There is no way you can compute the mean and median of such data.

   iii) The **median** wage as this does not take into account the very high wages of the few directors in the company or the very low wages of the few office juniors.

Groups of data

Sometimes the data we collect are in group form (see below). Finding the **mean** is slightly more difficult. We use the midpoint of the group and multiply this by the frequency. The table below shows the length of stay of a number of holidaymakers in a seaside resort.

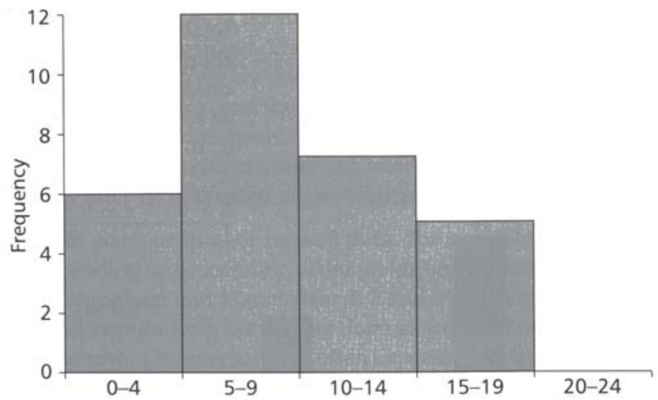
Length of stay in days	Midpoint (x)	Number of holidaymakers (f)	Midpoint × Frequency (fx)
0–4	2	6	12
5–9	7	12	84
10–14	12	7	84
15–19	17	5	85
20–24	22	0	0
Total		n = 30	$\sum(fx) = 265$

The **mean** is $\frac{\sum(fx)}{n} = \frac{265}{30} =$ 8.8 which is in the 5–9 group.

The **median** will be the average of the 15th and 16th values when all the values are placed in order. This will also fall in the 5–9 group. (The 5–9 group starts with the 7th value and ends with the 18th value, so the median is nearer the 'end' of the group – that is, it is nearer 9 than 5. In fact it is 8, but the method of calculating it is not required for this course.)

We cannot find the **mode** for grouped data but we can find the **modal group**. The modal group is the group that occurs most frequently (i.e. the 5–9 group). Note that the modal group is sometimes known as the **modal class**.

The data above can be plotted in the form of a bar graph, which makes it easier to find the modal group.



Number of holidaymakers

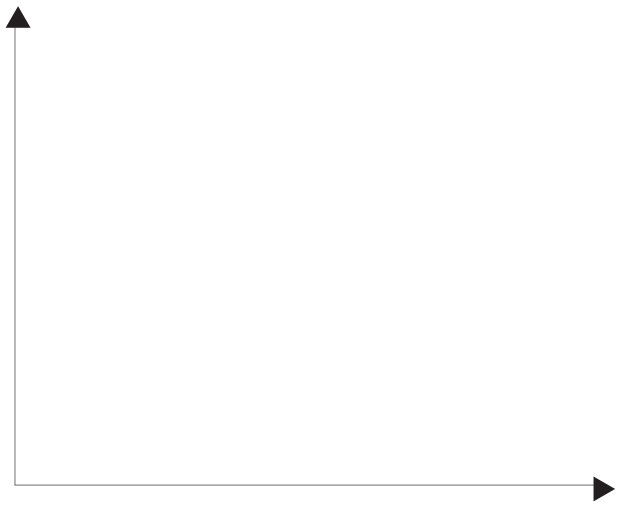
Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 10

A3

A survey of distances travelled by shoppers to a retail park is shown below:

Distance (km)	**Number of shoppers**
0–4	30
5–9	40
10–14	16
15–19	14
20–24	12
25–29	8

1. Plot this data in the form of a bar graph in the space provided below.



2. Complete the table below and work out the mean, modal group and median distance travelled by shoppers to the retail park.

Distance (km)	Midpoint (x)	No. of shoppers(f)	Midpoint × Frequency (fx)
0–4		30	
5–9		40	
10–14		16	
15–19		14	
20–24		12	
25–29		8	
Total		n =	$\sum(fx) =$

Mean ______________________________

Modal group ______________________________

Median ______________________________

3. Which of these measurements is:
   a) the most satisfactory
   b) the least satisfactory?

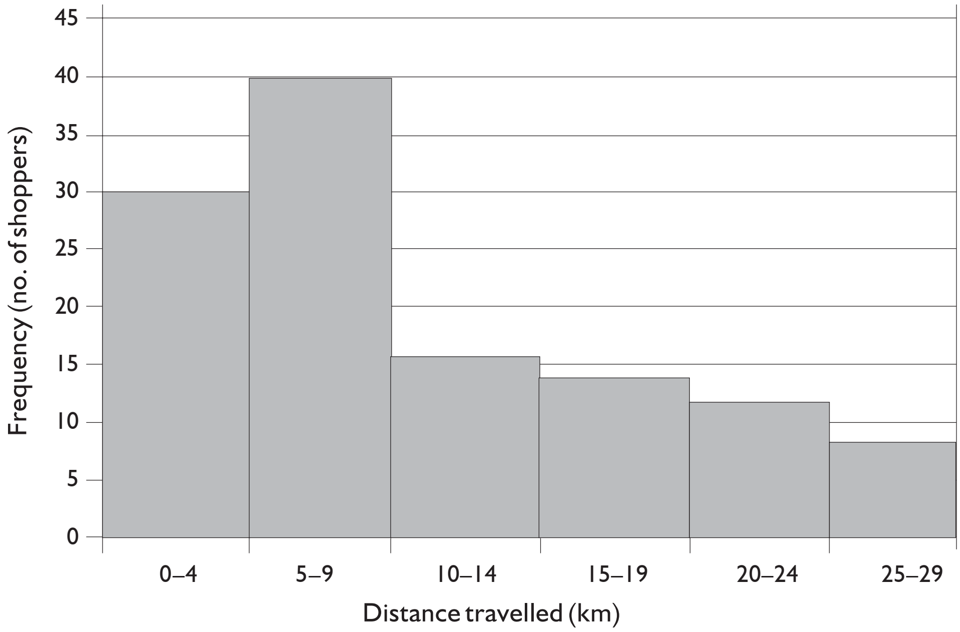
   In both cases justify your answer.

   a) ______________________________

   b) ______________________________

A3 comment

1. The bar graph should look like this.



2. The completed table should look like this.

Distance (km)	Midpoint (x)	No. of shoppers (f)	Midpoint × Frequency (fx)
0–4	2	30	60
5–9	7	40	280
10–14	12	16	192
15–19	17	14	238
20–24	22	12	264
25–29	27	8	216
Total		n = 120	$\sum(fx) = 1250$

Mean. The mean is $\frac{\Sigma(fx)}{n}$ which is $\frac{1250}{120}$ = 10.42 km. This is found in the 10–14 km group.

Modal group. The modal group is the group that occurs most often. In this case it is the 5–9 km group with 40 shoppers.

Median. The median is the average of the 60th and 61st numbers when all the values are placed in order. In this case the median group is also the 5–9 km group, because group 1 (0–4 km) with 30 customers added to group 2 (5–9 km) with 40 customers gives an answer of 70 which contains both the 60th and 61st numbers.

3. a) The 5–9 km group contains both the mode and the median, and in this case either would be a good measure to use.
   b) The average is in the 10–14 km group which shows that the small number of customers travelling the farthest distance (24–29 km) is affecting the mean. This suggests that the mean might not be the best measure of central tendency in this case.

T2

You should submit this work to your tutor as it will be formally assessed. Success in this assignment means you have evidence that you can understand data sets and measurements of central tendency.

Attempt the following.

1. Which type of data (nominal, ordinal, interval, ratio) is the following: primary industry, secondary industry, tertiary industry?

   ____

2. A geologist wants to classify his rock samples into the following categories: metamorphic, sedimentary, igneous. Which type of data is this?

   ____

3. Look at the following table:

Country	land area (000 km^2)	population (millions)
Belgium	30.5	10.4
France	552	57
Spain	505	39
UK	243	57.5
Norway	324	4.2

   Which type of data is shown in the first column of the table above?

   ____

4. Which type of data is shown in the second and third columns of the table above?

   ____

5. Find the **mean, median** and **mode** for the following data:

   The ages of the teachers in a primary school are as follows: 37, 28, 40, 33, 48, 41, 28, 24, 44, 27, 24.

   Mean ______________________________
   Median ______________________________
   Mode ______________________________

6. How long on average does it take a group of cyclists to travel to work along a cycle path:

22, 19, 24, 31, 16, 48, 29, 29, 21, 15, 22, 28, 27, 23, 37, 31, 23, 30, 26, 16, 26, 29 minutes.

Mean ______________________________

7. A student passes four Geography exams. His grades are 88, 78, 95 and 100. If asked for his average, which measurement of central tendency is the student likely to report? Why?

____

8. A national association of second-hand car dealers publishes data on the resale prices of cars in Britain. Which measurement of central tendency is the most appropriate for such resale prices? Explain your answer.

____

9. The SQA publishes data on the number of Higher examinations passed by male candidates and female candidates. Which measure of central tendency should be used here? Why?

____

10. Complete the table below and work out the mean daily rainfall in millimetres. Also find the modal group and median group.

Daily rainfall (mm)	Midpoint (x)	No. of days(f)	Midpoint × Frequency (fx)
0–2		20	
3–5		15	
6–10		7	
11–15		4	
16–20		2	
Total		n =	$\sum(fx) =$

Mean ______________________________
Modal group ______________________________
Median ______________________________

11. The marks for the students in a Geography examination were as follows:

88	82	89	70	85	63	100	86	67
39	90	96	76	34	81	84	89	96

a) Calculate the mean mark.

b) Group these exam marks using the classes 30–39, 40–49, 50–59, 60–69, 70–79, 80–89, 90–100.

c) What is the width of the classes?

d) Which class of marks is the modal group?

Measures of dispersion

So far we have looked at ways of summarising data by showing some sort of average or central tendency, but often it is useful to show how much figures differ from the average. This measure is known as **dispersion** and there are a number of ways of showing it.

> **The range**
>
> *The range is the difference between the maximum and the minimum values.*

The range has its limitations, however. For example, in a Geography exam it was found that the top mark was 90% and the bottom mark 10% giving a range of 80%, but most of the class got between 45% and 70%, i.e. they were all grouped near the middle, and the range was not able to pick this up.

The marks were:
10 25 45 47 49 51 51 52 52 54 56 57 58 60 62 66 68 70 75 90

> **The inter-quartile range**
>
> *The inter-quartile range is the range of the middle half of the values.*

The inter-quartile range is a better measurement to use than the range because it only refers to the middle half of the results. In other words the extremes are omitted and cannot affect the answer.

To calculate the inter-quartile range we must first find the **quartiles**. There are three quartiles, called Q1, Q2 and Q3 and they are the values that split the data into four equal parts (remember to arrange the values in order, lowest to highest, before you start). We do not need to bother about Q2, it is just the median and we have already seen how to calculate it. We only need to find Q1 and Q3. Q1 is simply the middle value of the *bottom* half of the data and Q3 is the middle value of the *top* half of the data. So we just split the data in half and find Q1 and Q3 using the same method as we used to find the median. Once we have found Q1 and Q3, we can calculate the **inter-quartile range which is Q3 minus Q1.**

For example, the figures below are the same twenty Geography exam marks arranged in order lowest to highest:

10 25 45 47 **49** **51** 51 52 52 54 56 57 58 60 **62** **66** 68 70 75 90

Because there is an even number of values (twenty), we can easily split them into two groups of ten. Q1 is the value in the middle of the lower group. If there is no single middle value, we have to take the average of the two middle values, just as we did with the median. We have to do that with this example. The two middle values are 49 and 51 and the average of these two is (49+51) ÷ 2 = 50 which is therefore Q1. Similarly, Q3 lies in the middle of the top half of the data and again we have to take the average of two middle values, 62 and 66. Therefore Q3 = (62+66) ÷ 2 = 64. The inter-quartile range then is Q3 – Q1 = 64 – 50 = 14.

There is just one 'trick' to learn. If there is an odd number of values, we cannot split the data into two equal groups. When this happens, there will always be a single middle value and we simply include this value in **both** halves of the data.

For example, we will add one more student's mark, 59, to the list of exam marks so that we now have an odd number of values, i.e. twenty-one. The extra mark is shown in brackets.

10 25 45 47 49 51 51 52 52 54 **56** 57 58 (59) 60 62 66 68 70 75 90

There are now twenty-one marks and the middle value is 56. We can now split the data into two equal halves by including 56 in both halves, thus:

10 25 45 47 49 **51** 51 52 52 54 56 – – – –56 57 58 59 60 **62** 66 68 70 75 90

Q1 and Q3 are now easy to find. Q1 is the middle value of the bottom half, i.e. 51 and Q3 is the middle value of the top half, i.e. 62. The inter-quartile range is
Q3 – Q1 = 62 – 51 = 11.

A4

1. Try adding another mark somewhere in the middle of the range of figures above to get a total of twenty-two marks and then calculate the inter-quartile range.

2. Try this again with another mark added in the middle of the range, i.e. giving a total of twenty-three marks. (Remember to use the 'trick'.)

3. The data below shows how long on average it takes a group of cyclists to travel to work along a cycle path (in minutes):

   22, 19, 24, 31, 16, 48, 29, 29, 21, 15, 22, 28, 27, 23, 37, 31, 23, 30, 26, 16, 26, 29

   Calculate the median and inter-quartile range.

4. The data below shows the minimum temperatures of ten weather stations in Britain on a winter's day. The temperatures are:

   5°, 9°, 3°, 2°, 7°, 9°, 8°, 2°, 2°, 3° (°Centigrade)

   Calculate the median and inter-quartile range.

A4 comment

1. ....11..

2. ....10..

3. Median = 26. Inter-quartile range = 7.

4. Median = 4. Inter-quartile range = 6.

Standard deviation

The **standard deviation** is one of the most important measures of dispersion. It gives us a much more accurate figure than the range or inter-quartile range as it takes into account all values and is not unduly affected by extreme values. It measures the **dispersion (or spread) of figures around the mean**. A **large number** for the standard deviation means that there is a **wide spread** of values around the mean, whereas a **small number** for the standard deviation implies that the values are grouped **close together** around the mean.

The formula for standard deviation is: $\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{n}}$

Worked example

Find the standard deviation of the minimum temperatures of ten weather stations in Britain on a winter's day. The temperatures are:

5°, 9°, 3°, 2°, 7°, 9°, 8°, 2°, 2°, 3° (°Centigrade)

- To calculate the standard deviation we construct a table like the one below:

x	$\bar{x}$	$(x - \bar{x})$	$(x - \bar{x})^2$
5	5	0	0
9	5	4	16
3	5	–2	4
2	5	–3	9
7	5	2	4
9	5	4	16
8	5	3	9
2	5	–3	9
2	5	–3	9
3	5	–2	4
$\Sigma x = 50$			$\Sigma(x - \bar{x})^2 = 80$
$\bar{x} = \frac{\Sigma x}{n} = \frac{50}{10} = 5$			$\frac{\Sigma(x - \bar{x})^2}{n} = 8$
			$\sqrt{\left[\frac{\Sigma(x - \bar{x})^2}{n}\right]} = 2.825$

- Column 1. List all the values (temperatures) (x). Add them up (Σx) and calculate the mean ($\bar{x}$).
- Column 2. Write the mean temperature ($\bar{x}$) in every row in the second column.
- Column 3. Subtract each value (temperature) from the mean. It does not matter if you obtain a negative number.
- Column 4. Square all the figures you obtained in column 3 to get rid of the negative numbers.
- Add up all the figures you calculated in column 4 to get $\Sigma(x - \bar{x})^2$
- Divide $\Sigma(x - \bar{x})^2$ by the total number of values, i.e. in this case 10 (the number of weather stations).
- Take the square root of this figure to obtain the standard deviation. (Round your answer to one decimal place.)

The standard deviation of the minimum temperatures is:
2.8 (rounded to one decimal place)

A5

A sample of coniferous trees had the following heights 3, 2, 1, 2, 3, 4, 3, 7, 6, 5 (metres).

Find the standard deviation.

x	$\bar{x}$	$(x - \bar{x})$	$(x - \bar{x})^2$
$\Sigma x =$			$\Sigma(x - \bar{x})^2 =$
$\bar{x} = \frac{\Sigma x}{n} =$			$\frac{\Sigma(x - \bar{x})^2}{n} =$
			$\sqrt{\left[\frac{\Sigma(x - \bar{x})^2}{n}\right]} =$

The standard deviation of the height of the trees is

A5 comment

Your table should look like this.

x	$\bar{x}$	$(x - \bar{x})$	$(x - \bar{x})^2$
3	3.6	–0.6	0.36
2	3.6	–1.6	2.56
1	3.6	–2.6	6.76
2	3.6	–1.6	2.56
3	3.6	–0.6	0.36
4	3.6	0.4	0.16
3	3.6	–0.6	0.36
7	3.6	3.4	11.56
6	3.6	2.4	5.76
5	3.6	1.4	1.96
$\Sigma x = 36$			$\Sigma(x - \bar{x})^2 = 32.4$
$\bar{x} = \frac{\Sigma x}{n} = 3.6$			$\frac{\Sigma(x - \bar{x})^2}{n} = 3.24$
			$\sqrt{\left[\frac{\Sigma(x - \bar{x})^2}{n}\right]} = 1.8$

The standard deviation of the height of the trees is 1.8 metres (rounded to one decimal place)

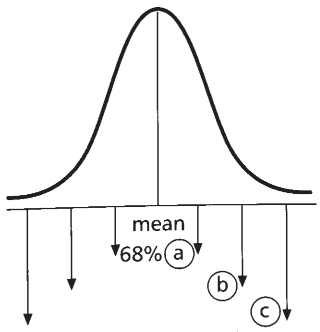
Normal distribution

You might wonder why we go to all this trouble to calculate the standard deviation when the range and the inter-quartile range are so much easier. The reason is that the standard deviation is much more useful. For example, the answer of 2.8°C, that we got in the worked example on minimum temperatures, means that there is a 68% chance* of the temperature falling within ±2.8°C of the mean temperature of 5°C, that is one standard deviation away from the mean. Normally, values are said to lie between one, two or three standard deviations from the mean.

* WHERE DID THE 68% CHANCE COME FROM?

The graph below shows a normal distribution curve, and many data sets when graphed form this type of curve. It is a bell-shaped curve where most of the data clusters around the middle or mean value and gradually declines the farther you get from the mean until very few data appear at the extremes. The distribution of people's heights is a good example. Most people are near average height, some are short, some are tall and a few are very short or very tall.

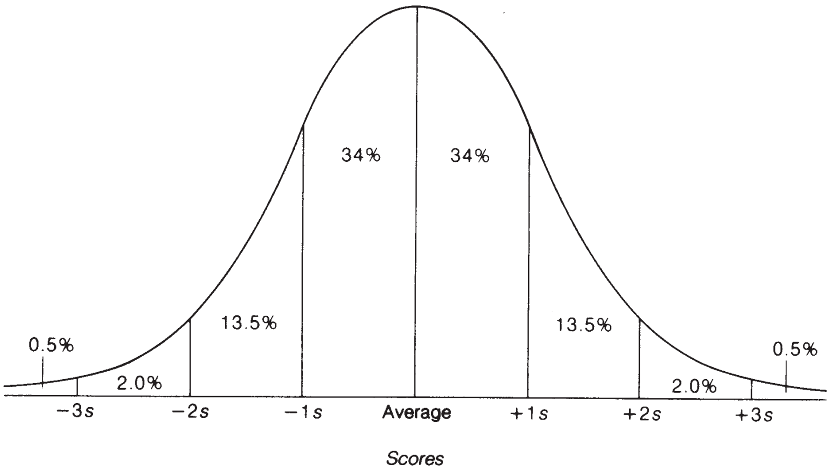
If you look at the graph you can see that most of the data (68%) is located within one standard deviation on either side of the mean, even more (95%) is located within two standard deviations on either side of the mean, and almost all (99%) of the data is located within three standard deviations on either side of the mean.



(a) 68% of the values lie within ±1 standard deviation of the mean
(b) 95% of the values lie within ±2 standard deviations of the mean
(c) 99% of the values lie within ±3 standard deviations of the mean

Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 15

A6



Source: *Statistics guide*, Open University, p. 9

The Wechsler Intelligence Scale for Children is designed to produce a distribution of IQ scores with an average of 100 and a standard deviation of 15.

Look at the normal curve and answer the following questions:

1. What is the average IQ of the group of children?
2. What percentage of scores fall between IQ 85 and IQ 115?
3. If a child has an IQ of 130, how many standard deviations is that away from the mean?
4. What is the range of IQs for 95% of the children?
5. If the standard deviation was 10 instead of 15 would the bell-shaped curve be broader or narrower? Comment on your answer.

1. ______________________________

2. ______________________________

3. ______________________________

4. ______________________________

5. ______________________________

A6 comment

1. The average IQ of the group of children is 100. Most scores cluster around the middle or average value and the number of scores gradually declines the farther you get from the average until very few scores appear at the extremes.

2. As the standard deviation is 15, 68% of the children will be found one standard deviation to either side of the average. In this case one standard deviation on either side is 85 (100 – 15 = 85) and 115 (100 + 15 = 115).

3. The answer is two standard deviations away from the mean. The mean is 100, the standard deviation is 15, one standard deviation is 115 (100 + 15); two standard deviations is 130 (100 + 2 x 15 = 130). Note that similarly an IQ of 70 is two standard deviations away from the mean, but in the other direction.

4. 95% of the children lie within two standard deviations on either side of the mean. This means they have an IQ of between 70 and 130. (100 – 2 x 15 = 70) ; (100 + 2 x 15 = 130).

5. The bell-shaped curve would be narrower. A small standard deviation ‘pulls’ the curve in towards the mean; a large standard deviation spreads it out.

The coefficient of variation and the standard error of the mean

The coefficient of variation

The coefficient of variation indicates the spread of values around the mean by a percentage.

The formula for the coefficient of variation is: $\frac{\text{standard deviation} \times 100}{\text{mean}}$

The higher the coefficient of variation the more widely spread the values are around the mean. The purpose of the coefficient of variation is to let us compare the spread of values between two different data sets that have different means.

The standard error of the mean

If we take a sample of 40 students we might find that the average number of Highers they passed was 4. However if we interview another 40 students from the same school we might find that that the average number of Highers they passed was different from this. How can we be certain that the average (mean) obtained from our sample represents the average number of Highers passed by **all** students? (i.e. the average of whole population – which is called the **true** mean)? We do this by finding the standard error of the mean of the sample.

The standard error of the mean is used to make an estimate of the limits within which the true mean lies.

The standard error of the mean is calculated by finding the standard deviation of our sample data (in this case the 40 students) and dividing the standard deviation by the square root of the number in our sample. This is the standard error of the mean and we know that the true mean for all the students who sat Highers has a 68% chance of lying within one standard error from our sample mean. Therefore, for any estimate of the mean that we arrive at (from a sample), it is possible to estimate within certain limits where the **true** population mean lies.

The formula for the standard error of the mean of a sample is:

$$SE = \frac{\text{standard deviation of the sample}}{\text{square root of number in sample}}$$

Worked example

Instead of noting down and calculating the mean atmospheric pressure for all the days in a year, we might select at random just 10% of the possible readings. Thus every 10th reading produced the following statistics:

1010	1019	1009	1018	1028	998
1020	1022	1018	1024	1008	1021
1024	1006	1032	1011	1006	1020
1016	1030	1000	1009	1029	1031
1008	989	1016	1010	1001	1008
1027	1004	1016	1012	1004	985

Data in millibars (mb)

Calculations on this data produce these formulae:
Sample mean = 1013.6 mb
Sample standard deviation = 11.2 mb

Obviously we have to accept that another thirty-six random readings will produce a different mean and standard deviation. Our sample is therefore subject to error.

We can estimate the value of the true mean using our sample mean and sample standard deviation, by fitting them into the formulae above.

$$\text{e.g. true mean} = \frac{1013.6 \pm 11.2}{\sqrt{36}} = \frac{1013.6 \pm 11.2}{6} = 1013.6 \pm 1.9 \text{ mb}$$

i.e. between 1011.7 and 1015.6

We have therefore found that the true mean for the whole year is likely to lie between 1011.7 mb and 1015.6 mb.

(This is correct. The calculation of the actual mean atmospheric pressure for the **whole** year produces a mean of 1014.6mb.)

A7

Using a sample of 40 students, estimate within one standard deviation (68%) where the true mean for the number of Highers lies. Their number of Highers is as follows:

1,4,2,1,1,5,3,3,2,4,4,2,3,5,0,5,4,3,1,3,3,2,2,5,4,4,1,4,2,3,5,0,1,4,3,5,4,4,1,2.

Calculations on this data produce:
Sample mean = 2.87
Sample standard deviation = 1.47

Calculate the standard error of the mean.

$$SE = \frac{\text{standard deviation of the sample}}{\text{square root of number in sample}}$$

=

=

=

A7 comment

Sample mean = 2.87
Sample standard deviation = 1.47

Number in sample = 40

$$SE = \frac{\text{standard deviation of the sample}}{\text{square root of number in sample}}$$
$$= \frac{1.47}{\sqrt{40}}$$
$$= \frac{1.47}{6.32}$$
$$= 0.23$$

Therefore our true mean number of Highers obtained by all the students has a 68% chance of lying between 2.87 ± 0.23. That is between 3.10 and 2.64 Highers.

T3

You should submit this work to your tutor because it will be formally assessed. Success in this assignment means you have evidence you can understand measurements of central tendency.

Calculate the mean, median, mode, range, inter-quartile range, standard deviation and coefficient of variation for the annual precipitation at Glasgow from the climatic data below.

Annual precipitation (in cm) at Glasgow over a 21-year period:

93 81 81 130 78 94 100 96 86 100 109 107 89 98 105 111 110 105

89 104 80

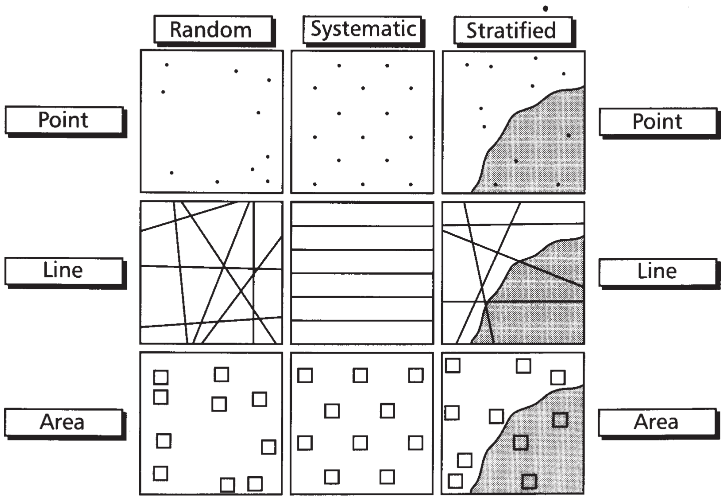
Basic concepts

Sampling

We have already mentioned the word **sample**. Everyone in everyday life encounters samples. In the morning before stepping into the bath we put our hand into the water to make sure the temperature is OK; in the evening we taste the pasta to make sure it is *al dente*. We are sampling the water and sampling the pasta!

In most investigations it is not possible to survey everything in the study area. For example you cannot ask every shopper on a busy Saturday what he or she is buying; similarly in a study of sand dune vegetation it is not possible to survey every plant in the whole area. As a result some form of selection is needed and this is called **sampling**.

There are three standard sampling techniques – random, systematic and stratified.

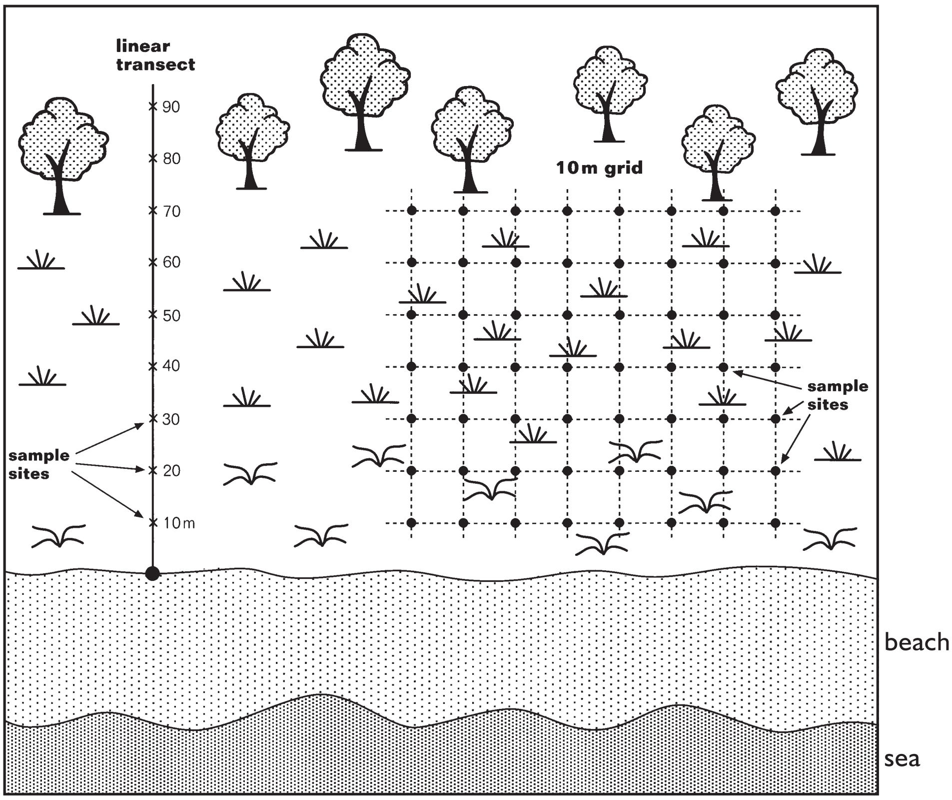


Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 65

1. **Random sampling** occurs if each item from a population has an equal chance of being selected. To select from this population each member is assigned a number using computer-generated random number tables. (See Appendix 2, Table 5.) Once a number has been chosen it can be related to a grid reference, an angular direction, a distance, or whatever else it is we are sampling.

2. **Systematic sampling** occurs if each sample is selected in a regular manner. For example, it could mean taking a vegetation sample every 10 metres (linear sample), taking a sample at a series of points located at the intersection of a 10-metre grid (point sample), or selecting every tenth customer at the supermarket, etc.

3. **Stratified sampling** is a technique used to ensure that a valid sample is selected for each sub-group we are interested in. For example, in a sand dune investigation seven-eighths of the dune might be managed and the rest unmanaged. With either of the other two methods we could, if we are unlucky, end up with too small a sample in the smaller sub-group, i.e. unmanaged dune. With stratified sampling, we simply decide before we start what sample size we need, in order to get valid results, in each sub-group (i.e. both managed and unmanaged dune). We then select the sample in each sub-group using one of the first two methods.



Source: *Geographical Enquiries*, Garrett Nagle and Kris Spencer, p. 64

Sample size and bias

Many investigations fail because the size of the sample is **too small** and this leads to **unreliable results**. When deciding how to select a sample, the main concern is to **remove bias**, i.e. to obtain a **representative sample**. Bias can arise for the following reasons:

- the population from which the sample is taken is already biased, e.g. all one rock type;
- insufficient care is taken to select a sample which represents the full population;
- the time when a sample is taken, e.g. during work hours.

AI

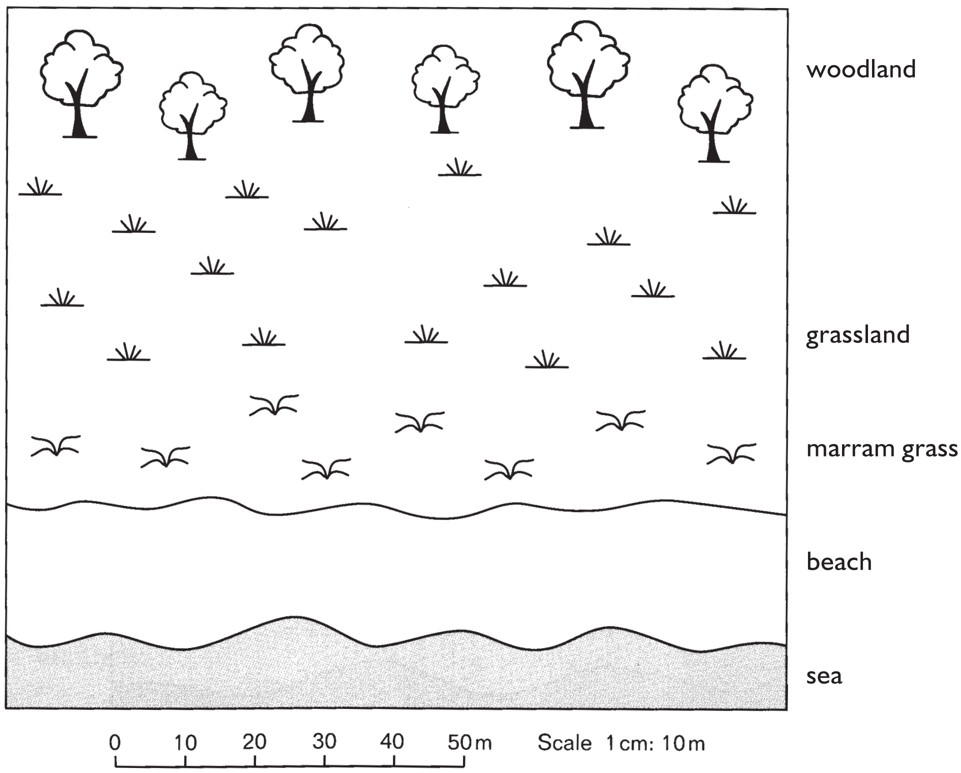
Answer the following questions in the spaces provided on the next page.

1. Explain why the following are not representative samples:
   a. A sample of 20 Glasgow dentists is taken to estimate the average income of all Glasgow residents.
   b. The political opinion of 150 voters in the retirement community of Sun City, Arizona is used as a sample of the political opinions of all Arizona voters.

2. The owner of a business that employs 100 people wants to select 25 of them at random to respond to a questionnaire. Construct a list of 25 random numbers between 1 and 100 that can be used to select the sample using the random number tables on page 208.

3. Suggest how you might obtain a sample of:
   a. pebbles on a river bed
   b. pebbles on a beach
   c. farms on an island of 10 km^2.

4. Construct a grid with 1 cm squares over the diagram below. Using the point sampling method calculate the percentage of each land use from the map of the following:
   a. woodland; b. grassland; c. marram grass; d. beach; e. sea.



Source: *Geographical Enquiries*, Garrett Nagle and Kris Spencer, p. 63

1. a. ______________________________

____

b. ______________________________

____

2. ______________________________

____

3. a. ______________________________

____

b. ______________________________

____

c. ______________________________

____

4. a. ______________________________

____

b. ______________________________

____

c. ______________________________

____

d. ______________________________

____

e. ______________________________

____

AI comment

1. a. Glasgow dentists earn considerably more than most Glasgow residents so the sample is not representative. Furthermore, a sample of 20 is too small to make any inference about the whole population of Glasgow, which is more than 600,000.

   b. The 150 voters in the retirement community are likely to be over 65 and relatively wealthy. They will probably have political views that will be determined by their age, where they live, and their wealth. These may be very different from the views of younger or poorer people.

2. The owner of the business should assign a number to each employee (either in alphabetical order or chronological order of when they joined the company). Then using the random numbers table on page 208, he should either work along the rows or down the columns and select the numbers in the order in which they appear in the tables. These numbers will relate to the numbers assigned to his employees.

3. a. The pebbles on a river bed could be chosen by linear sampling. A line could be stretched across the river and pebbles could be taken at 25-cm intervals along the line.

   b. A linear sample could again be collected here with lines either running parallel to the sea or at right angles to it.

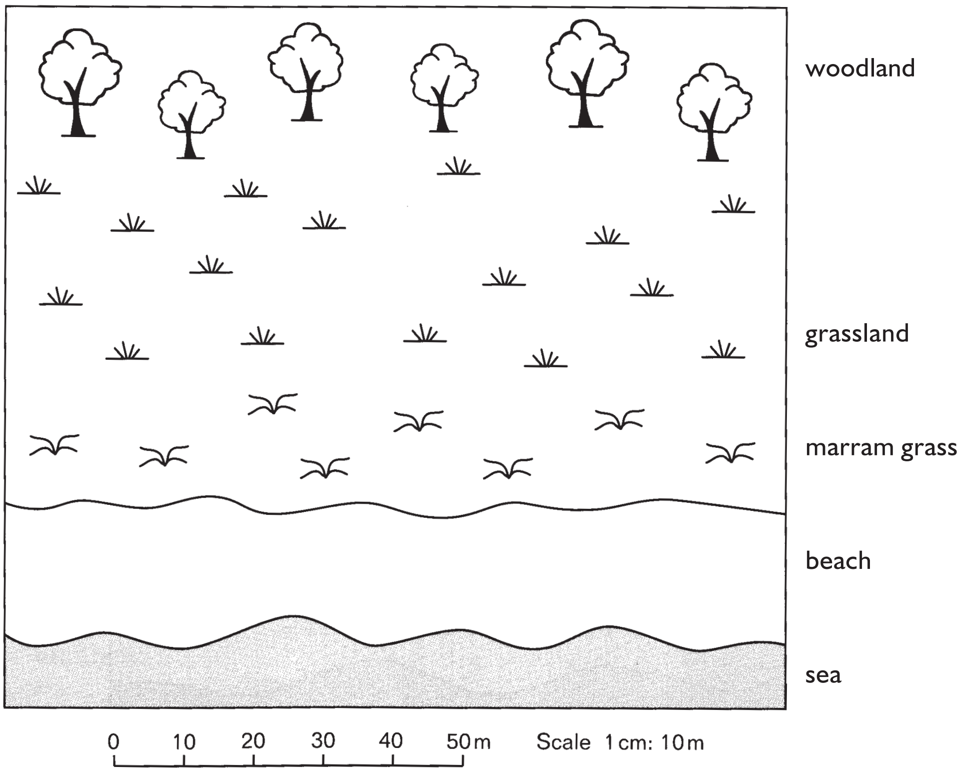
   c. Point sampling could be undertaken here. Place a grid over a map of the island and choose the farms that are located at the intersection of the points on the grid.

4. We calculate the area of land given over to each land use in the following way:
   - Calculate the area of the map in km using the linear scale provided.
   - Draw a 1 cm^2 grid on an acetate sheet and place the grid over the map.
   - Count the number of ‘hits’ for woodland (i.e. the number of times woodland occurs at the intersection of the grid lines).
   - Do the same for the other land uses.
   - Count up the total number of hits for all land uses.

- Work out the percentage of hits for each of the land uses by using the following formula:

$$\frac{\text{woodland hits}}{\text{total number of hits}} \times \text{area} = \text{.... km}^2 \text{ of woodland}$$

- Do the same for each of the other land uses.



Source: *Geographical Enquiries*, Garrett Nagle and Kris Spencer, p. 63

Hypothesis testing

As we have seen, a **sample** is defined as part of a larger **population** from which it is drawn. The sample may be drawn not only for the information it contains about the sample itself but also about what **inferences** (assumptions) can be made about the whole population that the sample was drawn from.

For example, suppose we are interested in finding out if there is any difference between the average number of Highers passed by students at Hilltop High School compared to the average for Smartypants Academy. We could send out a questionnaire to every student in both schools (i.e. **the population**) but this would be expensive and time-consuming. Instead, we could draw a **sample** of students from each school and give them the questionnaire and this would save money and time. Having obtained these samples we could then calculate the average number of Highers passed at each school. There will almost certainly be a difference, but how big does this difference have to be before we can say it is a 'real' or 'significant' difference – that is, to be sure that the difference we have found is not just a chance result from an unlucky sample (or samples)?

To answer this, we calculate a figure known as the **test statistic**, which is based on the data from our samples. Different types of problem require different test statistics, but the good thing is that statisticians have calculated all possible values of these test statistics and put them in **statistical tables**. All we need to do is to calculate our value and compare it with the value in the table to get our answer.

The whole process of asking a question like the one above and carrying out a statistical analysis to get the answer is known as **hypothesis testing**.

A **hypothesis** is a statement or a hunch. To test a hypothesis, the first thing we do is write down a statement. This statement is called the null hypothesis (written as **NH**) – that is the '**no change position**' and the researcher is trying to find evidence to prove that there **is a change**! Setting up a null hypothesis might seem a silly way to go about things, but in putting forward any case, it is fairer to the opposition to start with the assumption that you are wrong and then try to prove that you are right! It is a bit like 'innocent until proved guilty'!

The null hypothesis will usually have words such as:

'There is **no** difference between…'
'There has been **no** change since…'
'There is **no** relationship between…'

In the above case our null hypothesis (or NH) would be:

> 'There is **no difference** between the average number of Highers passed by students at Hilltop High School and the average number passed by students at Smartypants Academy.'

If you are **proved correct** then you **reject the null hypothesis** and **accept the alternative hypothesis (AH).** In this case the alternative hypothesis would be:

> 'There **is a difference** between the average number of Highers passed by students at Hilltop High and the average number passed by students at Smartypants Academy.'

[*Note that you will sometimes see the null hypothesis written as* H_0 *and the alternative hypothesis written as* H_1.]

A2

1. What is the meaning of the terms
   a. hypothesis
   b. null hypothesis
   c. alternative hypothesis?

2. For each of the investigations below write down
   a. the null hypothesis
   b. the alternative hypothesis.

   i) The Ministry of Agriculture and Fisheries wants to find out if the yield of wheat in British farms has changed after using brand X fertiliser instead of using brand Y fertiliser.

   ii) Public transport officials found that the mean travel time to work for Glasgow residents in 1995 was 20 minutes. A transport official wants to use a sample of this year's travel times to see if they have changed since 1995.

   iii) An economist wants to establish whether a relationship exists between the GNP of ten economically less developed countries and the provision of doctors per 1,000 people in those countries.

1. a. ____________________

   b. ____________________

   c. ____________________

2. i) a. ____________________

   b. ____________________

   ii) a. ____________________

   b. ____________________

   iii) a. ____________________

   b. ____________________

A2 comment

1. a. A **hypothesis** is a statement or a hunch.
   b. The null hypothesis (written as **NH**) is the 'no change position' and is <u>not</u> what the researcher is trying to prove.
   c. The alternative hypothesis (**AH**) <u>is</u> what the researcher is trying to find evidence to prove.

2. i) *Null hypothesis*. 'There is **no change** in wheat yields using brand X compared to brand Y.'
   Alternative hypothesis. 'There **is** a change in wheat yields using brand X compared to brand Y.'

   ii) *Null hypothesis*. 'The mean travel time of 20 minutes to work for Glasgow residents **has not changed** since 1995.'
   Alternative hypothesis. 'The mean travel time of 20 minutes to work for Glasgow residents **has changed** since 1995.'

   iii) *Null hypothesis*. 'There **is no relationship** between the GNP of ten economically less developed countries and the provision of doctors per 1,000 people.'
   Alternative hypothesis. 'There **is a relationship** between GNP of ten economically less developed countries and the provision of doctors per 1,000 people.'

Remember

Setting up a null hypothesis might seem a silly way to go about things, but in putting forward any case, it is fairer to the opposition to start with the assumption that you are wrong and then try to prove that you are right!

Significance

Before carrying out the test, we have to decide on a **significance level.** This is something that lets us determine at what point to **reject** the null hypothesis and accept the alternative hypothesis.

Significance is based on probability and chance. For example, in making **inferences** from a sample (in our case deciding that there is a difference in the average number of Highers obtained per student between the two schools, based on our samples of students), we run the risk that bad luck or **chance** might affect the results.

Statisticians have calculated the probability of 'chance' events occurring that might affect our results. They have come to the conclusion that, if the probability of an event occurring by chance is very small (for example less than 1 in 20), the result is **significant**, i.e. the result is **not** just a chance event. The practice is to refer to results as **significant** or **highly significant**. The significance levels at the end of this unit (pp. 204–7) are 0.05 and 0.01. This means that there is only a 1 in 20 (0.05) or a 1 in 100 (0.01) probability of the event occurring by chance.

You will be told which significance level to use in your worked exercises. The significance levels are shown at the top of each column of the tables. The values in the tables are called **critical values**.

> From your calculations and the significance tables on pages 204–7 you will find that if the value of the test statistic you have calculated is **greater** than the value in the table (the critical value), you can reject the null hypothesis and accept the alternative hypothesis.

Therefore, using suitable statistical techniques on our samples of students to obtain a value for the appropriate test statistic, and looking up the significance tables we might be able to reject our null hypothesis:

> **NH:** 'There is no difference between the average number of Highers passed by students at Hilltop High School and the average number passed by students at Smartypants Academy.'

and accept the alternative hypothesis:

> **AH:** 'There **is** a difference between the average number of Highers passed by students at Hilltop High School and the average number passed by students at Smartypants Academy.'

Degrees of freedom

In carrying out the various hypothesis tests, you will come across the term 'degrees of freedom'. This is something that affects the sample size in your test. Usually it makes the sample size slightly smaller. The reasons for this are beyond the scope of this course – all you have to do is be careful to follow the rules for each test. These will be stated clearly in each case.

You need to know the number of degrees of freedom in order to use the statistical tables.

A3

Using the appropriate tables on pages 204–7, find out if the following null hypotheses should be accepted or rejected. If the null hypothesis is rejected, state your findings in terms of the alternative hypothesis.

1. A sample of 20 fields was examined. Of these, 10 were given brand X fertiliser and 10 were given brand Y fertiliser to see if the wheat yields from the two groups of fields were different. A Student's t-test was undertaken to see if these differences were significant.
   The test statistic obtained from the Student's t-test was 2.00.

   - Write down the null hypothesis and the alternative hypothesis.
   - Look up the tables for critical values of Student's t-test at significance levels 0.05 (95%) and 0.01(99%) with 18 degrees of freedom.
   - Write down whether the figure of 2.00 is significant at these significance levels.
   - Write down whether you accept or reject the null hypothesis.

2. Some Geography students want to see if land use changes up a hillside. They divide the hillside into land below 200 metres, land between 201 and 400 metres and land above 401 metres. They undertake a land-use survey and then back in class calculate the number of 'hits' of five different land uses at different heights that they get from a land-use map they have drawn of the area. They wish to use the chi-square test to find out if the differences in land use with height are significant.
   The test statistic obtained from the chi-square test was 23.5.

   - Write down the null hypothesis and the alternative hypothesis.
   - Look up the tables for critical values of chi-square at significance levels 0.05 (95%) and 0.01(99%) with 8 degrees of freedom.
   - Write down whether the figure of 23.5 is significant at both significance levels.
   - Write down whether you accept or reject the null hypothesis.

A3 comment

1. • The null hypothesis is 'There is **no change** in wheat yields using brand X compared to brand Y.'
The alternative hypothesis is 'There **is** a change in wheat yields using brand X compared to brand Y.'

• The tables for critical values of Student's t-test at significance levels 0.05 (95%) and 0.01(99%) with 18 degrees of freedom are 2.10 and 2.88 respectively.

• The figure of 2.00 is **not** significant at significance levels 0.05 (95%) and 0.01(99%) with 18 degrees of freedom because it is **smaller** than the critical values obtained from the tables of 2.10 and 2.88 (respectively).

• Therefore the null hypothesis 'There is no change in wheat yields using brand X compared to brand Y' should be accepted.

2. • The null hypothesis is 'There is **no difference** between land uses at different heights' and the alternative hypothesis is 'There **is a difference** between land uses at different heights.'

• The tables for critical values of chi-square at significance levels 0.05 (95%) and 0.01(99%) with 8 degrees of freedom are 15.5 and 20.1 respectively.

• The figure of 23.5 **is** significant at both significance levels of 0.05 (95%) and 0.01(99%) with 8 degrees of freedom because it is **larger** than the critical values of 15.5 and 20.1 obtained from the tables.

• Reject the null hypothesis and accept the alternative hypothesis 'There **is** a difference between land uses at different heights.'

Student's t-test

A **Student's t-test** is a test of the difference between **two samples**. It is applied only to data measured on an **interval** or **ratio** scale. The null hypothesis of a Student's t-test is always that two sample means are the same. The alternative hypothesis is either that the two means are different or that one mean is greater than the other. (We have already looked at an example of the t-test with the wheat yields and brand X and brand Y fertiliser!)

The formula for a Student's t-test is: $t = \frac{\bar{x} - \bar{y}}{\sqrt{[(\text{SE of x})^2 + (\text{SE of y})^2]}}$

Worked example

(Calculating Student's t-test is quite a laborious task. It is easier to use a calculator or even better to use a computer.)
A researcher wants to find out if there is any difference between the birth rates of more developed and less developed countries.

- Set out the null hypothesis:
  NH: There is **no difference** between the mean birth rates of more developed and less developed countries.
  (In other words if it were possible to obtain the birth rates for all 'more developed' countries and for all 'less developed' countries, the mean birth rate for each type of country would be the same or very similar. The null hypothesis assumes that any difference we find between the two sample means is due solely to chance. However, if the two sample means are **very** different it will be hard to accept that assumption – but how different do they have to be before we would actually reject the null hypothesis? The Student's t-test gives us the answer.)
- State the alternative hypothesis:
  AH: There **is a difference** between the mean birth rates of more developed and less developed countries.
- Decide on a significance level. In this case we will make it 0.05 (or a 1 in 20 probability of chance affecting the figures).
- Set out the table as shown on the next page:
  1. Column 1 is the birth rates for the more developed countries (x). We need to total this ($\sum x$) and find the mean $\left(\bar{x} = \frac{\Sigma x}{n_x}\right)$.
  2. Column 2. Subtract the value of x from the mean of x to get $(x - \bar{x})$.
  3. Column 3 is column 2 squared. Total this to give $\sum(x - \bar{x})^2$ and divide this by the number in the sample (n_x). Take the square root of this to obtain σ_x the standard deviation of x.

     $$\sigma_x = \sqrt{\left[\frac{\Sigma(x-\bar{x})^2}{n_x}\right]}$$

  4. Calculate the standard error of x, i.e. divide the standard deviation (σ_x) by the square root of the total number of values ($\sqrt{n_x}$). (The total number of values is the number of countries in column 1.) SE of x = $\frac{\sigma_x}{\sqrt{n_x}}$
  5. Column 4 is the birth rates for the less developed countries (y). Total this ($\sum y$) and find the mean $\left(\bar{y} = \frac{\Sigma y}{n_y}\right)$.
  6. Column 5: subtract the value of y from the mean of y to get $(y - \bar{y})$.
  7. Column 6 is column 5 squared. Total this to give $\sum(y - \bar{y})^2$ and divide this by the number in the sample (n_y) . Take the square root of this to obtain σ_y, the standard deviation of y.

     $$\sigma_y = \sqrt{\left[\frac{\Sigma(y-\bar{y})^2}{n_y}\right]}$$

8. Calculate the standard error of y, i.e. divide the standard deviation (σ_y) by the square root of the total number of values ($\sqrt{n_y}$). (The total number of values is the number of countries in column 4.) SE of y = $\frac{\sigma_y}{\sqrt{n_y}}$

9. Calculate the value of t using the following equation:

$$t = \frac{(\bar{x} - \bar{y})}{\left[\sqrt{(\text{SE of x})^2 + (\text{SE of y})^2}\right]}$$

Birth rates in more and less developed countries

Column 1 More developed countries (x)	**Column 2** $(x - \bar{x})$	**Column 3** $(x - \bar{x})^2$	**Column 4** Less developed countries (y)	**Column 5** $(y - \bar{y})$	**Column 6** $(y - \bar{y})^2$
Switzerland 12	–2	4	Colombia 30	–10	100
Germany 10	–4	16	Zambia 51	11	121
Norway 13	–1	1	Egypt 35	-5	25
France 14	0	0	Kenya 48	8	64
Japan 16	2	4	India 35	–5	25
New Zealand 18	4	16	Brazil 29	–11	121
USA 15	1	1	Bangladesh 33	–7	49
Spain 16	2	4	Ethiopia 50	10	100
Italy 12	–2	4	Mali 49	9	81
$\Sigma x = 126$		$\Sigma(x-\bar{x})^2 = 50$	$\Sigma y = 360$		$\Sigma(y-\bar{y})^2 = 686$
$n_x = 9$		$\frac{\Sigma(x-\bar{x})^2}{n_x} = 5.56$	$n_y = 9$		$\frac{\Sigma(y-\bar{y})^2}{n_y} = 76.22$
$\bar{x} = 14$		$\sqrt{\frac{\Sigma(x-\bar{x})^2}{n_x}} = 2.36$ (this is σ_x)	$\bar{y} = 40$		$\sqrt{\frac{\Sigma(y-\bar{y})^2}{n_y}} = 8.73$ (this is σ_y)

SE of x = $\frac{\sigma_x}{\sqrt{n_x}} = \frac{2.36}{\sqrt{9}} = 0.786$

SE of y = $\frac{\sigma_y}{\sqrt{n_y}} = \frac{8.73}{\sqrt{9}} = 2.910$

$(\text{SE of x})^2 = 0.617$

$(\text{SE of y})^2 = 8.469$

Therefore, your value of $t = \dfrac{(\bar{x} - \bar{y})}{\sqrt{(\text{SE of x})^2 + (\text{SE of y})^2}}$

$$= \frac{14 - 40}{\sqrt{0.617 + 8.469}}$$

$$= \frac{-26}{\sqrt{9.086}}$$

$$= \frac{-26}{3.014}$$

$$t = -8.625$$

Note that the value of **t** you calculate can be either **positive** or **negative**. However, for this course, we can **ignore the sign** and just use the **figure** when comparing it with the critical value from the significance table. This **figure** is called the **calculated value of t**.

- Look up the significance table for **t** using 16 degrees of freedom $[(n_x - 1) + (n_y - 1) = 16]$ and the 0.05 significance level to get the critical value:

The critical value from the table = 2.12

- Since the calculated value of **t** is **greater** (t = – 8.625 but remember to ignore the minus sign here) than the critical value from the table, we can **reject** the null hypothesis:

  NH: There is **no** difference between the mean birth rates of more developed and less developed countries.

  and **accept** the alternative hypothesis:

  AH: There **is** a difference between the mean birth rates of more developed and less developed countries.

- Now look for geographical factors to explain your findings.

NOTE that if the calculated value of t is **less** than the critical value, we **cannot reject the null hypothesis** and we must therefore say that ‘There is no evidence to support the alternative hypothesis’.

One-tailed and two-tailed tests

The Student's t-test is a test of the difference between **two mean values** and when you are calculating the Student's t you will come across the terms 'one-tailed tests' and 'two-tailed tests'.

With a **one-tailed test** the hypothesis you wish to test (the alternative hypothesis) is framed so that the **direction is clear**, e.g.

AH: Men are <u>taller</u> than women.

On the other hand you might be interested simply in whether **differences exist in any direction**. This is a **two-tailed test** and your alternative hypothesis would be:

AH: Men are a <u>different</u> height from women.

Note that the null hypothesis is the same in both cases:

NH: There is <u>no difference</u> between the height of men and women.

The alternative hypothesis used in the worked example on the previous page,

AH: There <u>is a difference</u> between the mean birth rates of more developed and less developed countries.

is a **two-tailed test** as we have no expectations about the direction of the difference.

PLEASE NOTE: Different statistical tables are required for one-tailed and two-tailed tests. In this course the examples are all for two-tailed tests and the tables in the appendices are for two-tailed tests.

If a one-tailed test *is* required for any reason, this will be clearly stated and a separate table and instructions for one-tailed tests will be provided.

A5

The table below shows the percentage of arable land on farms in two areas. Test the hypothesis: There is a difference in the mean percentage of arable land on farms in the two areas.

Calculate the degrees of freedom to use, df = $(n_x - 1) + (n_y - 1)$ and test to 0.01 significance level. Clearly state the null hypothesis and the alternative hypothesis.

Null hypothesis

Alternative hypothesis

Percentage of arable land on farms in two areas

Area 1	Area 2
16	32
21	31
18	35
25	29
14	33
28	29
19	34
26	31
22	36
16	32
21	29
19	33

You might wish to use the table on the next page to assist with the calculations.

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6
Area 1 (x)	$(x - \bar{x})$	$(x - \bar{x})^2$	Area 2 (y)	$(y - \bar{y})$	$(y - \bar{y})^2$
16			32		
21			31		
18			35		
25			29		
14			33		
28			29		
19			34		
26			31		
22			36		
16			32		
21			29		
19			33		

Value for t = ____________________

Degrees of freedom = ____________________

Critical value of t from the tables at 0.05 significance
and __________ degrees of freedom = ____________________

Hypothesis

____

____

____

____

A5 comment

Your table should look like this.

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6
Area 1 (x)	$(x-\bar{x})$	$(x-\bar{x})^2$	Area 2 (y)	$(y-\bar{y})$	$(y-\bar{y})^2$
16	–4.42	19.51	32	0	0
21	0.58	0.34	31	–1	1
18	–2.42	5.84	35	3	9
25	4.58	21.01	29	–3	9
14	–6.42	41.17	33	1	1
28	7.58	57.51	29	–3	9
19	–1.42	2.01	34	2	4
26	5.58	31.17	31	–1	1
22	1.58	2.51	36	4	16
16	–4.42	19.51	32	0	0
21	0.58	0.34	29	–3	9
19	–1.42	2.01	33	1	1
$\Sigma x = 245$		$\Sigma(x-\bar{x})^2 = 202.92$	$\Sigma y = 384$		$\Sigma(y-\bar{y})^2 = 60$
$n_x = 12$		$\frac{\Sigma(x-\bar{x})^2}{n_x} = 16.91$	$n_y = 12$		$\frac{\Sigma(y-\bar{y})^2}{n_y} = 5.0$
$\bar{x} = 20.42$		$\sqrt{\frac{\Sigma(x-\bar{x})^2}{n_x}} = 4.11$	$\bar{y} = 32$		$\sqrt{\frac{\Sigma(y-\bar{y})^2}{n_y}} = 2.24$
		(this is σ_x)			(this is σ_y)

$$\text{SE of } x = \frac{\sigma_x}{\sqrt{n_x}} = \frac{4.11}{\sqrt{12}} = 1.187 \qquad \text{SE of } y = \frac{\sigma_y}{\sqrt{n_y}} = \frac{2.24}{\sqrt{12}} = 0.645$$

$(\text{SE of } x)^2 = 1.409$ $(\text{SE of } y)^2 = 0.417$

Therefore, your value of $t = \frac{(\bar{x}-\bar{y})}{\sqrt{(\text{SE of } x)^2 + (\text{SE of } y)^2}}$

$$= \frac{20.42 - 32}{\sqrt{1.409 + 0.417}}$$

$$= \frac{-11.58}{\sqrt{1.826}}$$

$$= \frac{-11.58}{1.351}$$

$$t = -8.572$$

Value for t = **8.572** (remember to ignore the minus sign!)
Degrees of freedom = $(n_x - 1) + (n_y - 1) = 11 + 11 = 22$
Critical value of t from the tables at 0.05 significance and 22 degrees of freedom = **2.07**

Hypothesis test results:
The value of t (8.572) is much bigger than the critical value from the table (2.07).
Therefore we can reject the null hypothesis in favour of our alternative hypothesis that:
There is a difference in the mean percentage of arable land on farms in the two areas.

T_4

You should submit this work to your tutor as it will be formally assessed. Success in this assignment means you have evidence that you can understand the Student's t-test.

Students are surveying the amount of litter dropped by the public in two areas of a country park. One area is beside a car park and picnic area, and the other is at an isolated beauty spot. The figures for the amount of litter dropped were taken by counting the amount of litter found in randomly thrown quadrats. The results were as follows:

Site A beside the car park		**Site B at the isolated beauty spot**	
Sample	No. of pieces of litter	Sample	No. of pieces of litter
1	6	1	11
2	13	2	4
3	7	3	1
4	8	4	6
5	7	5	8
6	10	6	4
7	12	7	1

Set up the null hypothesis and calculate the value of t. Use the 0.05 level of significance with 12 degrees of freedom.

Suggest (geographical) reasons for the result obtained.

The chi-square test

We often have occasion to make comparisons between two characteristics of something to see if they are linked or related to each other. One way to do this is to work out what we would **expect** to find if there was **no relationship** between them (the usual null hypothesis) and compare that with what we actually **observe**. The test we use to measure the difference between what is observed and what is expected according to an assumed hypothesis is called the chi-square test.

For example, some null hypotheses might be:
- There is no relationship between the height of the land and the vegetation cover.
- There is no difference in the location of superstores and of small grocers' shops.
- There is no connection between the size of farm and the type of farm.

The chi-square test can only be used on data that has the following characteristics:

- The data must be in the form of frequencies
- The frequency data must have a precise numerical value and it must be organised into categories or groups
- The total number of observations must be greater than 20
- The expected frequency in any one cell of the table must be greater than 5. However, if there only a few cells with 5 or less, it may be possible to combine 2 adjacent rows or columns so that all expected frequencies are greater than 5.

The formula for the chi-square test is: $\chi^2 = \frac{\Sigma(O-E)^2}{E}$

Where: χ^2 = the value of chi-square
O = the observed value
E = the expected value
$\Sigma(O-E)^2$ = all the values of (O – E) squared then added together.

Worked example

- **Write down the null hypothesis and alternative hypothesis and set the level of significance**

  NH: There is no difference in the distribution of old industry and modern industrial estates in the postal districts of Preston.

  AH: There is a difference in the distribution of old industry and modern industrial estates in the postal districts of Preston.

  We will set the level of significance at 0.05.

- **Construct a table with the information you have observed or obtained.**

  OBSERVED FREQUENCIES (O)

Post codes	**PN1**	**PN2**	**PN3**	**PN4**	**PN5**	**Row total**	
Old industry	9	13	10	10	8	**50**	
Modern industry	4	3	5	9	21	**42**	
Column total	**13**	**16**	**15**	**19**	**29**	**92**	**(grand total)**

(Note that although there are 3 cells in the table that are not greater than 5, these are observed frequencies. It is only the expected frequencies that have to be greater than 5, and as they are all greater than 5, we do not have a problem.)

- **Work out the expected frequency:**

$$\text{expected frequency} = \frac{\text{row total} \times \text{column total}}{\text{grand total}}$$

e.g. expected frequency for old industry in PN1 $= \frac{(50 \times 13)}{92} = 7.07$

EXPECTED FREQUENCIES (E)

Post codes	**PN1**	**PN2**	**PN3**	**PN4**	**PN5**	**Row total**	
Old industry	7.07	8.70	8.15	10.33	15.76	**50**	
Modern industry	5.93	7.30	6.85	8.67	13.24	**42**	
Column total	**13**	**16**	**15**	**19**	**29**	**92**	**(grand total)**

- **For each of the cells calculate** $\frac{(O-E)^2}{E}$

  e.g. old industry in PN1 is $\frac{(9-7.07)^2}{7.07} = 0.53$

  $\frac{(O-E)^2}{E}$

Post codes	PN1	PN2	PN3	PN4	PN5
Old industry	0.53	2.13	0.42	0.01	3.82
Modern industry	0.63	2.54	0.50	0.01	4.55

- **Add up all the above numbers to obtain the value for chi-square:** $\chi^2 = 15.14$

- **Look up the significance tables at the end of this unit. These will tell you whether to accept the null hypothesis or reject it.**

  The number of degrees of freedom to use is: the number of rows in the table minus 1, multiplied by the number of columns minus 1. This is $(2-1)(5-1) = 1 \times 4 = 4$ degrees of freedom. We find that our answer of 15.14 is **greater** than the critical value of 9.49 (for 4 degrees of freedom and a significance level of 0.05) and so **we reject the null hypothesis**.

  In other words:

  > The distribution of old industry and modern industrial estates in Preston is significantly different.

  Note that, even if we had picked the 'stronger' significance level of 0.01 for this test, our value of χ^2 of 15.14 would still have been bigger than the critical value of 13.28 which is the value in the table for a significance level of 0.01 (and for 4 degrees of freedom). This just makes us more convinced that the distribution of old established industry and modern industrial estates in Preston is significantly different.

- **Now look for geographical factors to explain your findings.**

A6

A land-use survey was undertaken to establish the location of grassland and woodland on a hill slope. A land-use map was then drawn. Point sampling on this map then established the number of 'hits' of grassland and woodland obtained at different locations.

	Under 200m	**201–400m**	**401m+**
grassland	36 'hits'	17 'hits'	7 'hits'
woodland	4 'hits'	23 'hits'	3 'hits'

Set up the null hypothesis and calculate the value of χ^2. Use 0.02 significance at 2 degrees of freedom.

Suggest (geographical) reasons for the result obtained.

Null hypothesis ____________________

Alternative hypothesis ____________________

EXPECTED FREQUENCIES

	Under 200m	**201–400m**	**401m+**
grassland			
woodland			

$$\frac{(O - E)^2}{E}$$

	Under 200m	**201–400m**	**401m+**
grassland			
woodland			

Value of χ^2 ____________________

Critical value for χ^2 (use 0.05 significance at 2 degrees of freedom) ____________________

Hypothesis ____________________

____

Suggest (geographical) reasons for the result obtained ____________________

____

____

____

____

____

____

____

A6 comment

Null hypothesis: There is no relationship between land use and altitude.
Alternative hypothesis: There is a relationship between land use and altitude.

	Under 200m	**201–400m**	**401m+**
grassland	36 'hits'	17 'hits'	7 'hits'
woodland	4 'hits'	23 'hits'	3 'hits'

EXPECTED FREQUENCIES

	Under 200m	**201–400m**	**401m+**
grassland	20	20	20
woodland	20	20	20

$$\frac{(O - E)^2}{E}$$

	Under 200m	**201–400m**	**401m+**
grassland	12.8	0.45	8.45
woodland	12.8	0.45	8.45

Value of $\chi^2 = 43.40$

Critical value for χ^2 (using 0.05 significance at 2 degrees of freedom) = 9.21

The calculated value (43.40) is much greater than the critical value (9.21) and we can therefore reject the null hypothesis.

Thus, there is a relationship between land use and altitude.

T5

You should submit this work to your tutor as it will be formally assessed. Success in this assignment means you have evidence that you can understand the chi-square test.

The types of shops in two settlements were compared to find out if there was a difference in the number of low, middle and high order shops in the towns. *Town A* with a population of 10,000 had 16 low order shops, 5 middle order shops and 3 high order shops, while *Town B* with a population of 35,000 had 28 low order shops, 26 middle order shops and 22 high order shops. Use the chi-square test to find out if there was a difference in the number of low, middle and high order shops in the towns.

Remember to set up the null and alternative hypothesis. Test your chi-square value at both the 0.05 and 0.01 levels of significance and comment on your findings. Use 2 degrees of freedom in each case.

Explain your answer geographically.

[Note that although two of the *observed* frequencies are not greater than 5, all of the *expected* frequencies are greater than 5 and therefore the analysis is valid.]

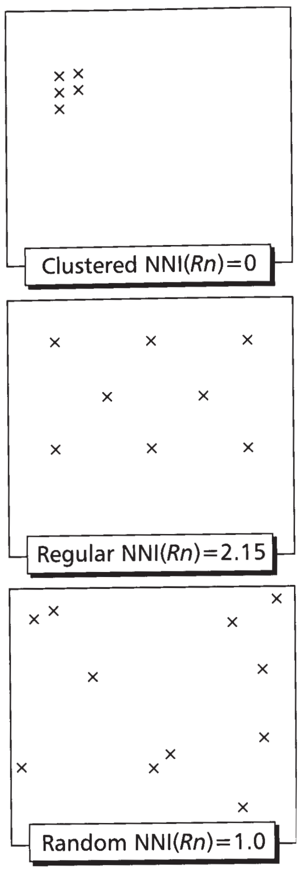
Nearest neighbour analysis

Settlements often appear on the map as dots. Dot distributions are commonly used in Geography, yet their patterns are difficult to describe. One way in which a pattern can be measured objectively is by **nearest neighbour analysis**. It can be used to identify a tendency towards clustering or dispersion for shops, industry, settlements, etc. Nearest neighbour analysis gives an index that enables one region to be compared to another.

The formula used in nearest neighbour analysis is $R_n = 2D\sqrt{\left(\frac{n}{a}\right)}$

The formula used in nearest neighbour analysis produces a figure expressed as R_n (the nearest neighbour index) which measures the extent to which a pattern is clustered, random or regular.

- **Clustered:** $R_n = 0$ All the dots occur close to the same point.
- **Random:** $R_n = 1.0$ There is no pattern (i.e. the distribution of the settlements is random).
- **Regular:** $R_n = 2.15$ There is a perfectly uniform pattern where each dot is equidistant from all its neighbours.



Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 28

How to undertake a nearest neighbour analysis

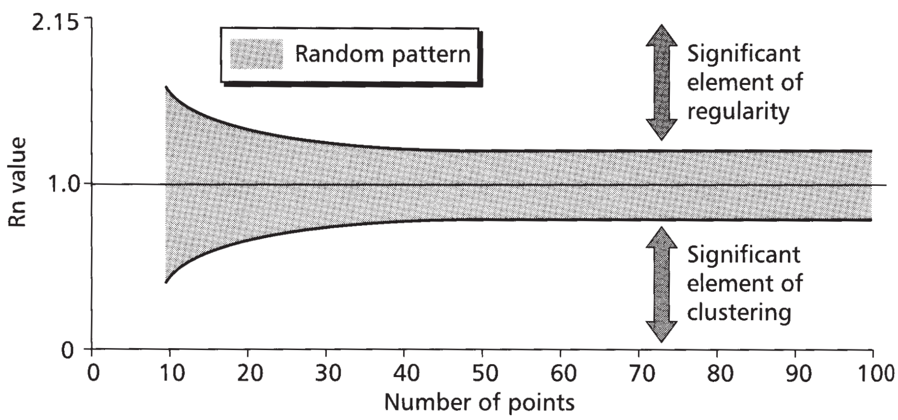
- The formula for nearest neighbour is $R_n = 2D\sqrt{\left(\frac{n}{a}\right)}$

where R_n is the nearest neighbour index
D = the average distance between each point and its nearest neighbour
n = the number of points under study
a = the size of the area under study

Note that $D = \frac{\Sigma d}{n}$

where d is the distance between each point and its nearest neighbour.

- Aim to measure as many points as possible (settlements, shops, plant species, etc.) – the more points the more significant the answer.
- Measure the straight line distance between each point and its nearest neighbour.
- Total all of the distances measured above (Σd) and divide by the number of points to find the average distance ($\Sigma d/n$). This is D in the formula.
- Calculate the total area of your study area (a).
- Fit your calculations into the formula to calculate R_n.
- Using the R_n number, refer to the diagram below to determine how regular or clustered the pattern is.
- Now look for geographical factors to explain your findings.



Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 28

Worked example

Calculate the nearest neighbour for the following settlements located in an area of 56 square kilometres and account for the pattern found.

Settlement	Nearest neighbour	Distance (km)
A	B	1.0
B	A	1.0
C	D	0.6
D	C	0.6
E	F	1.6
F	E	1.6
G	H	1.8
H	I	1.3
I	H	1.3

$$\mathbf{\Sigma d} = 10.8$$

$$\mathbf{n} = 9$$

$$\left(\frac{\mathbf{\Sigma d}}{\mathbf{n}}\right) = \frac{10.5}{9} = 1.2 \text{ (this is also called } \mathbf{D} \text{ in the formula)}$$

$$\mathbf{area} = 56 \text{ km}^2$$

The formula used in nearest neighbour analysis is $\mathbf{R_n = 2D\sqrt{\left(\frac{n}{a}\right)}}$

$$= 2 \times 1.2 \times \sqrt{\frac{9}{56}}$$

$$= 2.4 \times 0.4$$

$$= 0.96$$

The answer of 0.96 is close to 1.0, which means that there is a random settlement pattern in the area.

A7

The distribution of coniferous trees in a Forest Enterprise woodland of 100 × 200 metres was measured. Undertake a nearest neighbour analysis and account for the pattern found.

Tree no.	Nearest neighbour	Distance (m)
1	2	30
2	1	30
3	2	35
4	3	60
5	6	40
6	5	40
7	8	60
8	7	60
9	7	100
10	5	50

$\sum d =$

$n =$

$\left(\frac{\sum d}{n}\right) = \qquad =$ (this is also called D in the formula)

area $=$ m^2 (sq metres)

The formula used in nearest neighbour analysis is $\mathbf{R_n = 2D\sqrt{\left(\frac{n}{a}\right)}}$

$\mathbf{R_n} =$

$=$

$=$

Account for the pattern.

A7 comment

Tree no.	Nearest neighbour	Distance (m)
1	2	30
2	1	30
3	2	35
4	3	60
5	6	40
6	5	40
7	8	60
8	7	60
9	7	100
10	5	50

$$\sum d = 505$$

$$n = 10$$

$$\left(\frac{\Sigma d}{n}\right) = \frac{505}{10} = 50.5$$ (this is also called D in the formula)

$$\text{area} = 100 \times 200 = 20{,}000 \text{ m}^2$$ (sq metres)

The formula used in nearest neighbour analysis is $\mathbf{R_n = 2D\sqrt{\left(\frac{n}{a}\right)}}$

$$\mathbf{R_n} = 2 \times 50.5 \times \sqrt{\frac{10}{20{,}000}}$$

$$= 101 \times 0.02236$$

$$= 2.258$$

As the answer of 2.258 is close to 2.15, the analysis tells us the pattern is almost uniform, i.e. each tree is almost equidistant from its neighbours. This is likely to be the case when the trees have been planted for commercial reasons.

T6

You should submit this work to your tutor as it will be formally assessed. Success in this assignment means you have evidence that you can understand nearest neighbour analysis.

Calculate a nearest neighbour analysis of a number of villages in an area of 64 km^2.

Village	Nearest neighbour	Distance (km)
A	F	1.5
B	C	2.0
C	B	2.0
D	E	3.0
E	B	3.0
F	G	1.0
G	F	1.0
H	I	1.5
I	H	1.5
J	K	2.0
K	J	2.0

Scattergraphs and correlation

In this section we are going to introduce the idea of **correlation** and in particular to look at a technique known as Spearman's rank correlation coefficient. In the next section we will look at another technique called Pearson's correlation coefficient.

Correlation is one of the most widely used statistical techniques in Geography. It is used to examine **the degree of relationship between two numerical variables.** When two variables (for example, town size and number of shops) increase or decrease together, this is known as a **positive correlation**. Where one variable increases as the other decreases (for example, increasing altitude with decreasing temperature), this is known as a **negative correlation**. The table below shows some possible relationships.

Variable 1	Variable 2	Relationship
Size of settlement	Number of shops	*Positive*. As the population of a settlement increases so too does the number of shops.
Temperature	Altitude	*Negative*. Increasing height means decreasing temperature.
Age of a car	Price of car	*Negative*. The older the car the lower the price.
Water pressure	Depth of water	*Positive*. The deeper the water the more pressure there is.
Length of hair	Physical strength	*No relationship*. Length of hair does not affect an individual's strength.

Scattergraphs

Scattergraphs can show the relationship between the variables. The values of the variables are plotted on a graph as a series of dots and then it is often possible to draw a 'best fit' line through the plotted points leaving roughly the same number of points on either side of the line.

In most cases, one variable will be dependent on the other. For example, if we look at the table below, temperature is dependent on altitude – not the other way around. If we want to draw this as a graph, the **dependent** variable (in this case **temperature**) is always put on the **Y-axis** and the other variable, **altitude**, along the **X-axis**. The other variable is called the **independent** variable.

In order to set up the scattergraph the right way round we write a statement in the form:

'one variable'	is dependent on	'the other variable'
e.g. **temperature**	is dependent on	**altitude**
vertical axis **(the Y-axis)**		horizontal axis **(the X-axis)**

The variable on the **left** of the statement always goes on the **vertical** axis (the Y-axis) of the graph – and the variable on the **right** always goes on the **horizontal** axis (the X-axis) of the graph. So the axes of the scattergraph will look like this:

Y-axis

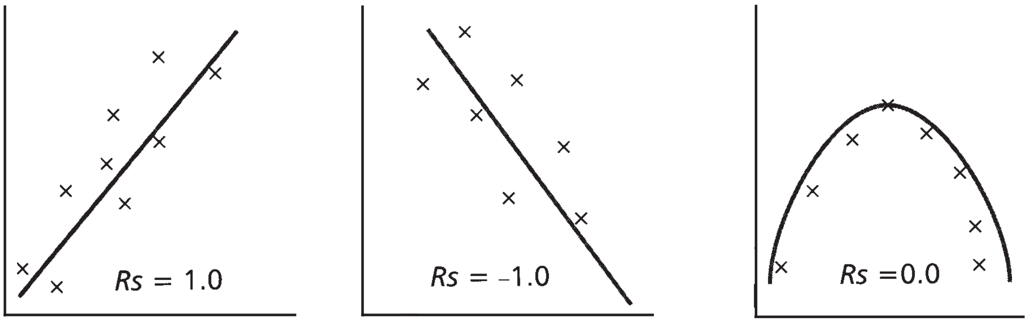
'dependent variable'

(e.g. temperature)

X-axis 'independent variable' (e.g. altitude)

All we need to do now is to plot the points on the graph, using these two axes, to get our scattergraph.

The graphs below show some best-fit lines for positive, negative and no correlation.



Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 25

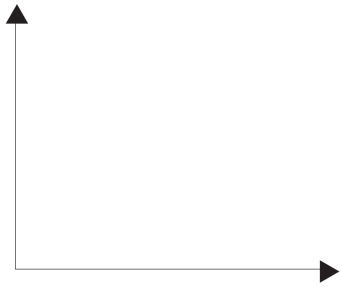
A8

Look again at the table and attempt the exercise underneath.

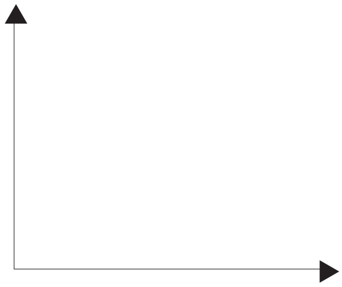
	Variable 1	Variable 2	Relationship
1.	Size of settlement	Number of shops	*Positive*. As the population of a settlement increases so too does the number of shops.
2.	Temperature	Altitude	*Negative*. Increasing height means decreasing temperature.
3.	Age of a car	Price of car	*Negative*. The older the car the lower the price.
4.	Water pressure	Depth of water	*Positive*. The deeper the water the more pressure there is.
5.	Length of hair	Physical strength	*No relationship*. Length of hair does not affect an individual's strength.

In the graph frames on the next two pages write in the independent and dependent variable along the X- and Y-axes, draw in an expected best-fit line and in the space below the graph write out which variable is dependent on the other. Do this for each of the five examples in the table above.

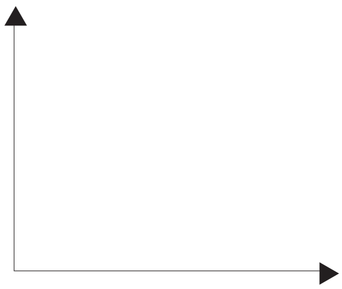
1.



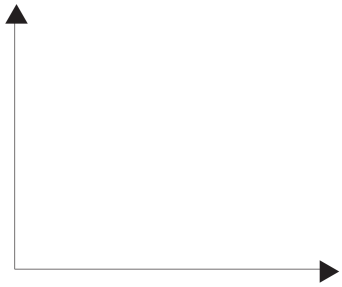
2.



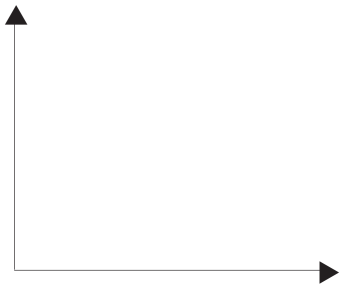
3.



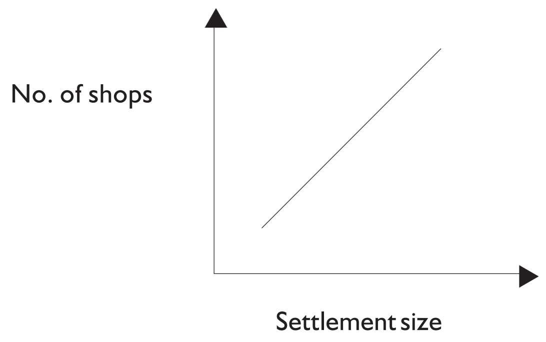
4.



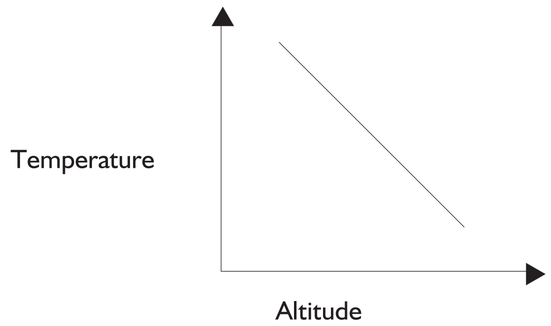
5.



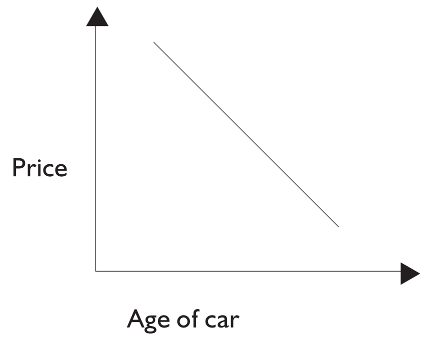
A8 comment



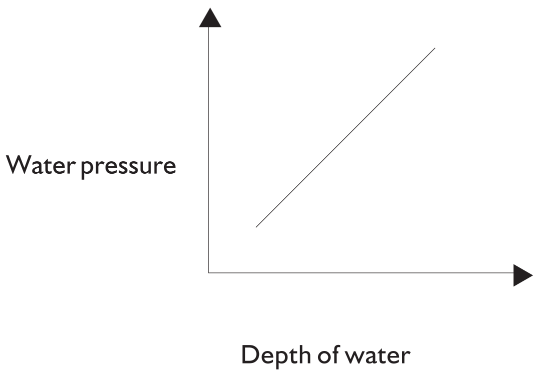
1. ***Size of settlement and number of shops***
   The number of shops depends on the size of the settlement, since usually the bigger the settlement the greater the number of shops. Therefore the number of shops is the dependent variable and should be drawn up the Y-axis. The best-fit line increases from the left-hand side of the page to the right, as there is a positive correlation (the **bigger** the settlement the **greater** the number of shops).



2. ***Temperature and altitude***
   In normal conditions, temperature decreases with increasing height above sea level (altitude), therefore temperature is the dependent variable and should be drawn up the Y-axis. The best-fit line decreases from the left-hand side of the page to the right, as there is a negative correlation (the **higher** the altitude the **lower** the temperature).



3. ***Age and price of a car***
The price of a car usually (unless you own a classic or unusual car!) depends on the age of the car – the older the car the less money you get when you try to sell it. Therefore the price is the dependent variable and should be drawn up the Y-axis. The best-fit line decreases from the left-hand side of the page to the right, as there is a negative correlation (an **increase** in the age of a car means a **decrease** in the money obtained for it, in other words the older the car the less money you will get for it).



4. ***Water pressure and depth of water***
It is a well known fact that for every 10 metres depth of water the pressure increases by one atmosphere, so that on the water's surface the pressure will be one atmosphere, at 10 metres depth there will be two atmospheres pressure, at 20 metres three atmospheres pressure, etc. (This has led to all sorts of problems with North Sea divers when they dive at great depth.) Because the water pressure depends on the depth, the water pressure is the dependent variable and should be drawn up the Y-axis. The best-fit line increases from the left-hand side of the page to the right, as there is a positive correlation. (An **increase** in depth means an **increase** in pressure.)

5. ***Hair length and physical strength***
Despite the fable of Samson and Delilah there is **absolutely no relationship** between the two variables. Therefore it would be impossible to draw a best-fit line!

Spearman's rank correlation coefficient

This correlation is quick and easy to calculate. It only requires ordinal (ranked) data or other data that can be ranked, e.g. Gross National Product, infant mortality, etc. The correlation coefficient obtained is an index measuring between +1 (perfect positive correlation) and –1 (perfect negative correlation) with the value of 0 showing that there is no relationship between the two variables. The nearer the coefficient is to +1 or –1 the stronger the relationship is between the two variables. Here, as elsewhere, we use the 5% (0.05) and 1% (0.01) levels of significance.

How to calculate Spearman's rank correlation:

- State the null hypothesis: The number of shops in a town is not affected by the population size of a town (i.e. no correlation).
- State the alternative hypothesis: There is a relationship (or correlation) between the number of shops and the size of the town.
- List your two sets of data side by side.
- Find the rank values of both sets of data from highest to lowest. When two cases share the same rank, use the average rank. For example, if two cases share ranks 2 and 3, both take on rank 2.5 and the next case in order is ranked 4th. Use a similar method if three or more cases share the same rank. (Do *not* put the cases in order, just find the rank values.)
- Work out the difference in rank between columns 1 and 2 (d).
- Square each of the differences (d^2).
- Total the differences ($\sum d^2$).
- Put them into the formula for Spearman's rank . . .

The formula for Spearman's rank is $r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$

- . . . and calculate the value of $\mathbf{r_s}$. Note that $\mathbf{r_s}$ can be positive or negative.
- Now compare your result with the figure in the table on page 206 to see if your Spearman's rank value is large enough to indicate a significant relationship between the two sets of data. To do this, just use the figure you obtained for $\mathbf{r_s}$ ignoring the sign (+ or –). For this test, the number of degrees of freedom to use **equals** the number of ranked pairs (n). Look down the degrees of freedom column and check your result against the 0.05 and 0.01 significance levels. **If your value of $\mathbf{r_s}$ is greater than the figure in the table, then there is evidence of a correlation between the variables**.
- Now look for geographical factors to explain your findings.

Worked example

State the null hypothesis: The number of shops in a town is not affected by the population size of a town.

Town	size (in 000's)	rank (highest to lowest)	no of shops	rank (highest to lowest)	d	d^2
A	25	1	40	3	–2	4
B	32	2	32	2	0	0
C	40	3	30	1	2	4
D	52	4	95	4	0	0
E	60	5	150	7	–2	4
F	64	6	142	6	0	0
G	76	7.5	132	5	2.5	6.25
H	76	7.5	157	8	–0.5	0.25
I	92	9	230	9	0	0
J	95	10	240	10	0	0
						$\sum d^2 = 18.5$

The formula for Spearman's rank is

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$
$$= 1 - \frac{6 \times 18.5}{10 \times (10^2 - 1)}$$
$$= 1 - \frac{111}{990}$$
$$= +0.888$$

Now compare your result with the figure in the table on page 206 to see if your Spearman's rank value is large enough to indicate a significant relationship between the two sets of data. To do this, just use the figure you obtained for $\mathbf{r_s}$ ignoring the sign (+ or –). For this test, the number of degrees of freedom to use **equals** the number of ranked pairs (n). In this case it is 10.

Look down the degrees of freedom column and check your result against the 0.05 and 0.01 significance levels. **If your value of $\mathbf{r_s}$ is greater than the figure in the table, then there is evidence of a correlation between the variables**.

The nearer to +1 or –1 the better the correlation between the variables. In this case the figure is +0.888 which is quite close to +1 and therefore there is a strong correlation between the variables, which means that there is a relationship between the number of shops and the size of the population of the town and we can **reject** the null hypothesis. Note that because the sign in this case is +, it means that the correlation is a positive correlation; that is, the larger the population the more shops there are.

A scattergraph and best-fit line can be drawn to show this.

Now look for geographical factors to explain your findings.

A9

Attempt the following exercise:

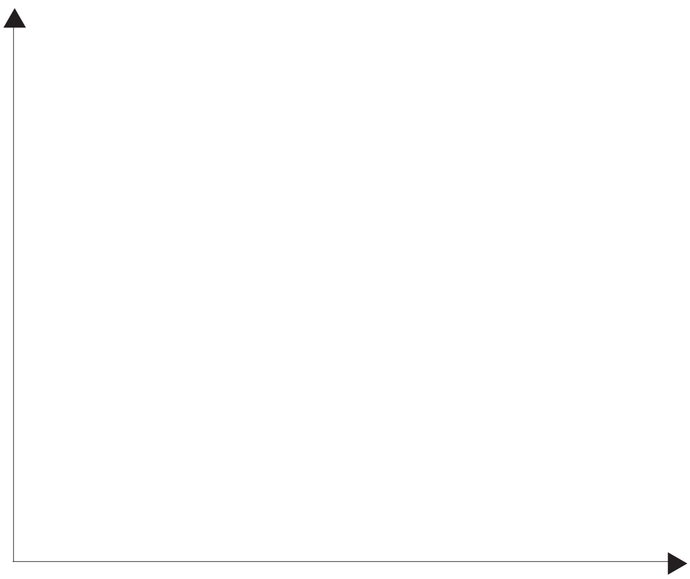
Use Spearman's rank correlation coefficient to calculate the correlation coefficient of the following data.

Write down the null hypothesis and the alternative hypothesis and complete the table below.

Draw a scattergraph to illustrate the relationship.

The table shows annual income and weekly food expenditure of eight families.

Family	**annual income** (in £ooo's)	**rank** (lowest to highest)	**weekly food expenditure** (in £'s)	**rank** (lowest to highest)	**d**	**d^2**
1	55		180			
2	60		210			
3	35		150			
4	20		130			
5	25		120			
6	30		140			
7	18		100			
8	19		120			



A9 comment

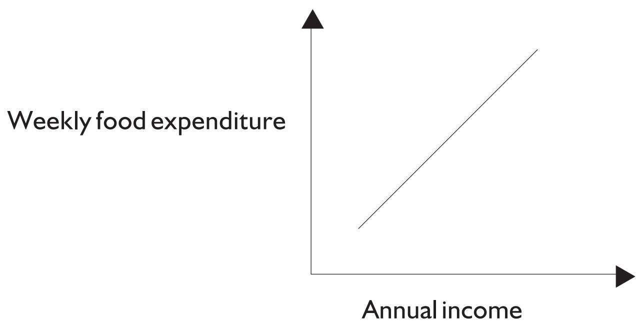
Family	annual income (in £ooo's)	rank (lowest to highest)	weekly food expenditure (in £'s)	rank (lowest to highest)	d	d^2
1	55	7	180	7	0	0
2	60	8	210	8	0	0
3	35	6	150	6	0	0
4	20	3	130	4	–1	1
5	25	4	120	2.5	1.5	2.25
6	30	5	140	5	0	0
7	18	1	110	1	0	0
8	19	2	120	2.5	–0.5	0.25

Now compare your result with the figure in the table on page 206 to see if your Spearman's rank value is large enough to indicate a significant relationship between the two sets of data. To do this, just use the figure you obtained for $\mathbf{r_s}$ ignoring the sign (+ or –). For this test, the number of degrees of freedom to use **equals** the number of ranked pairs (n). In this case it is 8.

Look down the degrees of freedom column and check your result against the 0.05 and 0.01 significance levels. **As the value of r_s of 0.985 is greater than the figure in the table, then there is evidence of a correlation between the variables**.

The nearer to +1 or –1 the better the correlation between the variables. In this case the value is very close to +1 and therefore we **reject** the null hypothesis, There is no relationship between annual income and weekly food expenditure, and accept the alternative hypothesis There is a relationship between annual income and weekly food expenditure. Since the value of $\mathbf{r_s}$ is positive, we can say that there is a positive correlation between the variables and therefore 'the higher the annual income the greater the food expenditure'.

A scattergraph and best-fit line can be drawn to show this.



T7

You should submit this work to your tutor as it will be formally assessed. Success in this assignment means you have evidence that you can understand Spearman's rank correlation coefficient.

Use Spearman's rank correlation coefficient to calculate the correlation coefficient of the following data.

Write down the null hypothesis and the alternative hypothesis, then copy and complete the table below (some rank values have been given to help you). Calculate Spearman's rank correlation coefficient, compare it with the critical value for a significance level of 0.01 from the tables and decide whether to accept or reject the null hypothesis.

Draw a scattergraph to illustrate the relationship.

Explain your results geographically.

Country	**Gross National Product ($ per capita 1993)**	**Rank**	**% agricultural workforce**	**Rank**
Australia	17,510	9	6	4
Bangladesh	220	2	59	11
Bolivia	770	7	47	10
Brazil	3,020	8	25	8
China	490	5	73	13
Egypt	660	6	42	9
Ethiopia	100	1	88	15
India	290		62	
Italy	19,620		9	
Japan	31,450		7	
Kenya	270		81	
Norway	26,340		6	
Switzerland	36,410		6	
UK	17,970		2	
USA	24,750		3	

Pearson's product moment correlation coefficent

This is the most widely used correlation method. It is a more powerful method than Spearman's rank because it **uses the actual values instead of ranks in its calculations to examine the relationship between two variables.** Its full name is the 'Pearson product moment correlation coefficient' and you may see it written like this, especially in statistical tables. It is called **r**.

The formula for Pearson's correlation coefficient is $r = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{[\Sigma(x-\bar{x})^2 \times \Sigma(y-\bar{y})^2]}}$

How to calculate Pearson's correlation coefficient:

- List your two sets of data side by side (x in column 1, y in column 2).
- Total the x figures for column 1, this is Σx. Find the mean for column 1, this is $\bar{x}$.
- Total the y figures for column 2, this is Σy. Find the mean for column 2, this is $\bar{y}$.
- In column 3 subtract each of the values of column 1 from the mean of x. This is called $(x-\bar{x})$. Remember to include negative numbers.
- In column 4 subtract each of the values of column 2 from the mean of y. This is called $(y-\bar{y})$. Remember to include negative numbers.
- In column 5 square the values you obtained in column 3. Add these values to obtain $\Sigma(x-\bar{x})^2$.
- In column 6 square the values you obtained in column 4. Add these values to obtain $\Sigma(y-\bar{y})^2$.
- Now multiply the sum of column 5 by the sum of column 6. Then take the square root, to obtain $\sqrt{[\Sigma(x-\bar{x})^2 \times \Sigma(y-\bar{y})^2]}$. This is the **denominator (bottom line)** of your formula.
- In column 7 multiply column 3 by column 4. Add these values to obtain $\Sigma(x-\bar{x})(y-\bar{y})$. This is the **numerator (top line)** of your formula.
- Now divide the **numerator** by the **denominator** to find **r**. Your answer should lie between +1.0 and –1.0.
- Look up the significance tables to find if the relationship is significant. The number of degrees of freedom is two less than the sample size, in this case, 9 – 2 = 7 degrees of freedom.
- Explain your results geographically.

Here is an example of the layout to use:

Relationship between age and height x = age y = height

	1	2	3	4	5	6	7
Person No.	x	y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
1	*The*	*You*					
2	*x*	*must*					
3	*and*	*calculate*					
4	*y*	*all*					
5	*values*	*the*					
6	*will*	*other*					
7	*be*	*figures.*					
8	*given.*						
9							
	$\sum x =$	$\sum y =$			$\sum(x-\bar{x})^2 =$	$\sum(y-\bar{y})^2 =$	$\sum(x-\bar{x})(y-\bar{y}) =$
	$\bar{x} =$	$\bar{y} =$			$\sum(x-\bar{x})^2 \times \sum(y-\bar{y})^2 =$		**(numerator)**
					$\sqrt{\{\sum(x-\bar{x})^2 \times \sum(y-\bar{y})^2\}} =$ **(denominator)**		

Worked example

Examine the relationship between unemployment and infant mortality in the regions of Italy.

Test the hypothesis that there is a relationship between unemployment and infant mortality. (Use the 0.05 and 0.01 levels of significance as evidence of a relationship.)

Regional contrasts in Italy

x = unemployment (%) y = infant mortality (per 1,000 births)

	1	2	3	4	5	6	7
Region	x	y	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
N.W.	9	9	–2	–1	4	1	2
Lombardy	6	6	–5	–4	25	16	20
N.E.	6	5	–5	–5	25	25	25
Emilia	7	8	–4	–2	16	4	8
Centre	8	9	–3	–1	9	1	3
Lazio	12	14	1	4	1	16	4
Campania	10	11	–1	1	1	1	–1
Abruzzo	24	15	13	5	169	25	65
South	17	13	6	3	36	9	18
	$\Sigma x = 99$	$\Sigma y = 90$			$\Sigma(x-\bar{x})^2 = 286$	$\Sigma(y-\bar{y})^2 = 98$	**$\Sigma(x-\bar{x})(y-\bar{y}) = 144$**
	$\bar{x} = 11$	$\bar{y} = 10$			$\Sigma(x-\bar{x})^2 \times \Sigma(y-\bar{y})^2 = 28028$		**(numerator)**
					$\sqrt{\{\Sigma(x-\bar{x})^2 \times \Sigma(y-\bar{y})^2\}}$ = 167.42 (denominator)		

The formula for Pearson's correlation coefficient is

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{[\Sigma(x-\bar{x})^2 \times \Sigma(y-\bar{y})^2]}}$$

$$= \frac{144}{167.50}$$

$$= 0.86$$

Hypothesis test

NH: (There is no relationship between unemployment and infant mortality in the regions of Italy.)

AH: (There is a relationship between unemployment and infant mortality in the regions of Italy.)

Degrees of freedom = no. of paired observations minus 2 = 9 – 2 = 7

We are asked to test at significance levels 0.05 and 0.01.

The Pearson's coefficient critical value for 7 degrees of freedom and a significance level of 0.05 = 0.666
The Pearson's coefficient critical value for 7 degrees of freedom and a significance level of 0.01 = 0.798.

As the Pearson's correlation coefficient is greater than both of the above figures then we can reject the null hypothesis and accept the alternative hypothesis: There is evidence that there is a strong relationship between unemployment and infant mortality.

A10

Calculate the Pearson correlation coefficient to show the relationship between the percentage of people who work in industry or services and life expectancy. Write this in terms of the null and alternative hypotheses. Use the table below to help you with your calculations. (Use the 0.05 and 0.01 levels of significance with 6 degrees of freedom as evidence of a relationship.)

Country	**% work in industry or services**	**life expectancy**
Ethiopia	25	43
Congo	40	48
Kenya	21	55
Indonesia	50	58
Brazil	76	67
Portugal	85	73
France	90	75
Canada	93	77

NH (There is no relationship between ______________________________)

AH (There is a relationship between ______________________________)

	1	**2**	**3**	**4**	**5**	**6**	**7**
Country	x	y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
Ethiopia	25	43					
Congo	40	48					
Kenya	21	55					
Indonesia	50	58					
Brazil	76	67					
Portugal	85	73					
France	90	75					
Canada	93	77					
	$\Sigma x =$	$\Sigma y =$			$\Sigma(x-\bar{x})^2 =$	$\Sigma(y-\bar{y})^2 =$	**$\Sigma(x-\bar{x})(y-\bar{y}) =$**
	$\bar{x} =$	$\bar{y} =$			$\Sigma(x-\bar{x})^2 \times \Sigma(y-\bar{y})^2 =$		**(numerator)**
					$\sqrt{\{\Sigma(x-\bar{x})^2 \times \Sigma(y-\bar{y})^2\}} =$ (denominator)		

The formula for Pearson's correlation coefficient is:

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{[\Sigma(x-\bar{x})^2 \times \Sigma(y-\bar{y})^2]}}$$

=

=

A10 comment

NH (There is no relationship between % of people who work in industries and services and life expectancy.)
AH (There is a relationship between % of people who work in industries and services and life expectancy.)

	1	2	3	4	5	6	7
Country	x	y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
Ethiopia	25	43	–35	–19	1225	361	665
Congo	40	48	–20	–14	400	196	280
Kenya	21	55	–39	–7	1521	49	273
Indonesia	50	58	–10	–4	100	16	40
Brazil	76	67	16	5	256	25	80
Portugal	85	73	25	11	625	121	475
France	90	75	30	13	900	169	390
Canada	93	77	33	15	1089	225	495
	$\Sigma x=480$	$\Sigma y=496$			$\Sigma(x-\bar{x})^2=6116$	$\Sigma(y-\bar{y})^2=1662$	**$\Sigma(x-\bar{x})(y-\bar{y})=2498$**
	$\bar{x}=60$	$\bar{y}=62$			$\Sigma(x-\bar{x})^2\times\Sigma(y-\bar{y})^2=7106792$		**(numerator)**
					$\sqrt{\{\Sigma(x-\bar{x})^2\times\Sigma(y-\bar{y})^2\}}$ =2665.86 (denominator)		

The formula for Pearson's correlation coefficient is

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{[\Sigma(x-\bar{x})^2 \times \Sigma(y-\bar{y})^2]}} = \frac{2498}{2665.86} = 0.94$$

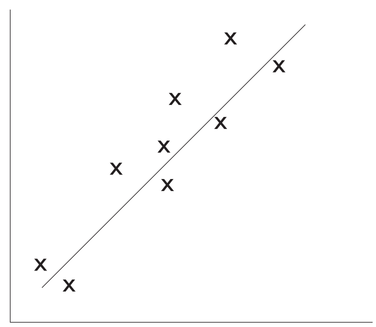
Hypothesis test
Degrees of freedom = number of paired observations minus 2 = 8 – 2 = 6
We are asked to test at significance levels 0.05 and 0.01.

The Pearson's coefficient critical value for 6 degrees of freedom and a significance level of 0.05 = 0.707
The Pearson's coefficient critical value for 6 degrees of freedom and a significance level of 0.01 = 0.834

As the Pearson's correlation coefficient is greater than both of the above figures then we can reject the null hypothesis and accept the alternative hypothesis: There is evidence that there is a strong relationship between % of people who work in industries and services and life expectancy.

Linear regression

Scattergraphs and linear regression
The next stage is to 'look' at the data. To do this we draw a scattergraph of the data – this is a graph of the relationship between the two variables.



Once a scattergraph has been drawn it is possible to draw a 'best-fit' line (a straight line) through the points on the graph. To do this by hand, we might try to draw a straight line through as many points as possible with roughly an equal number of points on either side of the line. It is quite difficult to be sure we have got the 'best fit' by hand so we use **regression** to do the job for us. It is because we are trying to get the best fit to a **straight** line that we call this technique **linear regression**, rather than just 'regression'.

By using the technique of linear regression we can calculate mathematically the best straight line to draw through points on a scattergraph.

From mathematics the equation for a straight line is $Y = a + bX$

Where 'X' and 'Y' represent the variables we are dealing with, i.e. Y = height and X = age, and 'a' and 'b' are as defined below.

'b' is the 'regression coefficient'

It is given by the formula: $b = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$

'a' is known as the 'intercept'. It is given by the formula: $a = \bar{y} - b\bar{x}$

Note, you must always calculate 'b' first.

Worked example

Draw a best-fit line to show the relationship between infant mortality and unemployment in the regions of Italy. (Worked example on page 140.) Our strong hunch is that infant mortality is in some way dependent on the level of unemployment.

- First write down the statement:
  Infant mortality is dependent on unemployment
- From this, write down the equation of the straight line and 'identify' X and Y:
  infant mortality = a + b (unemployment)

$$\mathbf{Y = a + b\,X}$$

- Now carry out a correlation analysis using the Pearson correlation coefficient method. (This has just been done!)
- Use this to calculate the value of b using the formula $b = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$

  [Hint, the top line is the sum of column 7 and the bottom line is the sum of column 5, from the Pearson correlation coefficient calculation.]

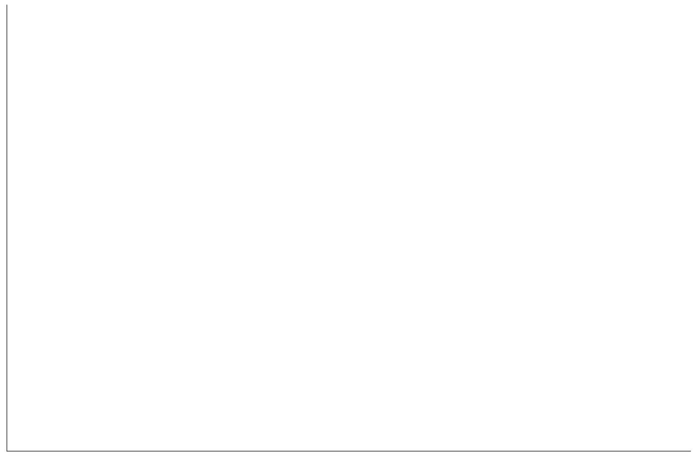
- And then find **a** using the formula $\mathbf{a = \bar{y} - b}$.
  [Hint: $\bar{\mathbf{x}}$ is the mean value of **x** from column 1 and $\bar{\mathbf{y}}$ is the mean value of **y** from column 2, from the Pearson correlation coefficient calculation.]

  You now have the equation of the straight line:
  Y (infant mortality) = a + b X (unemployment)

- To draw this straight line on the scattergraph we simply calculate values of y for chosen values of x using the above formula. For example, try using x = 5 and x = 20. These are values from the bottom and top of the range of x values. Sometimes x = 0 is a good value to use. (It is an easy one to use because when x = 0, y = a.) For this example let us use all three values:

  When x = 0, y = a =
  When x = 5, y =
  When x = 20, y =

- Plot these points on the scattergraph and draw a straight line between them. This is the **regression line**.



A11

Complete a correlation analysis using the Pearson correlation coefficient and draw a regression line to show the relationship between the Maths marks and the Physics marks of the following students. We have a strong hunch that being good at maths helps a student to be good at Physics; that should mean that a student's Physics mark is in some way dependent on his or her Maths mark.

Write this in terms of the null and alternative hypotheses. Use the table below to help you with your calculations. (Use the 0.05 and 0.01 levels of significance with 2 degrees of freedom as evidence of a relationship.)

Null hypothesis

____

Alternative hypothesis

____

STUDENT NO.	**MATHS (col x)**	**PHYSICS (col y)**
1	20	40
2	40	45
3	60	75
4	80	90

	Maths	**Physics**	**3**	**4**	**5**	**6**	**7**
Student no.	x	y	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
1	20	40					
2	40	45					
3	60	75					
4	80	90					
	$\sum x =$	$\sum y =$			$\sum(x-\bar{x})^2=$	$\sum(y-\bar{y})^2=$	**$\sum(x-\bar{x})(y-\bar{y})=$**
	$\bar{x} =$	$\bar{y} =$			$\sum(x-\bar{x})^2 \times \sum(y-\bar{y})^2 =$		**(numerator)**
					$\sqrt{\{\sum(x-\bar{x})^2 \times \sum(y-\bar{y})^2\}}=$ (denominator)		

A11 comment

Null hypothesis: There is no relationship between a student's Maths marks and his/her Physics marks.

Alternative hypothesis: There is a relationship between a student's Maths marks and his/her Physics marks.

STUDENT NO.	MATHS (col x)	PHYSICS (col y)
1	20	40
2	40	45
3	60	75
4	80	90

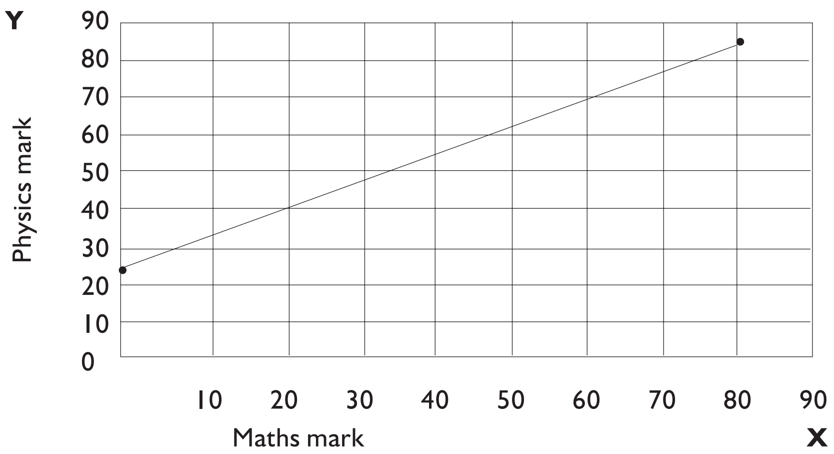
	Maths	Physics	3	4	5	6	7
Student no.	x	y	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
1	20	40	–30	–22.5	900	506.25	675
2	40	45	–10	–17.5	100	306.25	175
3	60	75	10	12.5	100	156.25	125
4	80	90	30	27.5	900	756.25	825
	$\Sigma x=200$	$\Sigma y=250$			$\Sigma(x-\bar{x})^2=2000$	$\Sigma(y-\bar{y})^2=1725$	**$\Sigma(x-\bar{x})(y-\bar{y})=1800$**
	$\bar{x}=50$	$\bar{y}=$**62.5**			$\Sigma(x-\bar{x})^2 \times \Sigma(y-\bar{y})^2=3{,}450{,}000$		**(numerator)**
					$\sqrt{\{\Sigma(x-\bar{x})^2 \times \Sigma(y-\bar{y})^2\}}$ $=1857.42$ **(denominator)**		

Pearson's correlation coefficient $= \dfrac{1800}{1857.42} = \mathbf{0.97}$

The critical values for 0.05 and 0.01 levels of significance are 0.95 and 0.99 respectively.

Therefore our value of 0.97 is significant at the 0.05 level, but not at the 0.01 level.

Thus we can reject the null hypothesis in favour of the alternative hypothesis (at the 0.05 level of significance) and say that there is a relationship between a student's Maths marks and his/her Physics marks.
But we cannot reject the null hypothesis in favour of the alternative hypothesis at the 0.01 level, so we cannot say that it is a 'strong' relationship.



T8

You should copy and submit this work to your tutor as it will be formally assessed. Success in this assignment means you have evidence that you can understand Pearson's correlation coefficient and linear regression.

Test the hypothesis that there is a relationship between the distance from the source of a river (in kilometres) and average pebble size (in millimetres). Comment geographically on your findings.
(Use the 0.05 level of significance as evidence of a relationship and the 0.01 level of significance as evidence of a strong relationship.)

Null hypothesis

____

Alternative hypothesis

____

Calculate the line of regression of distance from source versus pebble size. Plot the points and draw the line of regression.

	Distance from source (km)	Pebble size (mm)	**3**	**4**	**5**	**6**	**7**
Site no.	x	y	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
1	0.4	18					
2	0.8	15					
3	1.2	12					
4	1.6	10					
5	2.0	8					
6	2.4	4					
7	2.8	7					
8	3.2	5					
9	3.6	2					
	$\sum x =$	$\sum y =$					
	$\bar{x} =$	$\bar{y} =$					

SECTION 4

Introductory notes for Sections 4–6

The second skill area of Unit 1 is called ***the production and interpretation of maps and diagrams***, and has been divided into three sections:

Section 4: The design and layout of maps and diagrams
Section 5: The interpretation of Ordnance Survey maps and related data
Section 6: Topographic analysis based on Ordnance Survey maps

These sections have a dual purpose. First they aim to help you understand how to construct and present a variety of maps and diagrams which you may be able to use in your Geographical Study. Second, you are required to be able to interpret and analyse information from a 1:25000 OS map and Sections 4–6 of this pack give you hints and tips on how to do this properly. They do assume, however, that you have some experience and proficiency in the basic skills of OS map reading, so if you do not have these, read your Higher Geography notes and/or a quality Geography map-reading textbook.

The maps recommended in Sections 4–6 are the 1:25000 OS extracts used in past Standard Grade examinations as it was felt that these maps would be easily available from the Geography department in your school. The map you will be given in the external examination will also have a scale of 1:25000 but it will be the **1:25000 OS Explorer** map covering a wider area than the Standard Grade extracts. Ask your tutor if he /she has examples of these maps from past Advanced Higher Geography and Certificate of Sixth Year Studies Geography examinations.

There are tutor assignments located throughout Sections 4–6.

Sections 4–6 should take about 5 hours to complete.

The principles of linework, lettering and shading

All maps should be easy to interpret and understand, they should have good use of colour and /or shading and they should contain well presented information.

The following rules should be observed when constructing a map:

- The simpler the map, the clearer the information will be. Two simple maps or one map with an overlay will show information more clearly than a confused jumble on one map.
- Neither make your map too large with too much 'white space', nor too small, which will take a magnifying glass to read!
- Spatial distributions can be shown on a map (e.g. relief, the growth of a town, the distribution of different types of industry) with the appropriate use of shading, colour or symbols as long as a key is included.
- Space must be allowed on the map to include the key, which often takes up more room than anticipated by students!
- Do not forget to show the map's scale (and possibly a compass direction)!
- Important features must be included on the map particularly if referred to in the text or if they can be used as a framework for additional material.
- Annotation can provide a valuable source of information as long as the information is included outside the frame of the map, is concise, and is connected to the appropriate feature by a straight line.
- ***Linework***. When you are constructing your map you should work initially in pencil and colour pencil. This can easily be erased. You can draw over this work with fine-nib 'roller'-type felt pens when you are satisfied with your work or just leave well alone! Do **not** draw with a biro pen.
- ***Lettering***. Again work in pencil to make sure the words will fit comfortably onto the map. Make sure your writing is neat and legible. You could lightly pencil in guidelines to keep your writing horizontal. You can use stencils or 'Letraset' but these tend to be fiddly.
- ***Shading***. Shading or colouring is important to the visual appearance of a map. It is important to keep the tone or colour even and not patchy. Also the density of colour or tone should increase with increasing values. For example, try increasing blue colours with increasing depth of the oceans or increasing brown colours with increasing height of mountains. (Interestingly in lowland areas the lower the land the deeper the green colour!)
- Finally, don't forget the map's title – your reader needs to know what he/she is looking at!

A1

Using the 1995 Standard Grade General examination map extract of Perth, draw a sketch map to show the site of the city of Perth.

A1 comment

Decide on the size of the area you need to draw. The map extract is 24 cm by 24 cm (6 km by 6 km^2.) A 16 cm × 16 cm frame would probably work here. Remember to work out the new scale (in this case, 1 cm represents 0.375 km).

Draw the main lines on your map. In this map these would include the River Tay, the major roads (M90, A9, A93, A94, A85, A90, A912), the railway and the roads in the central business district (CBD).

Draw in the main areas of housing (shade in grey), and upland to the south and around Kinnoull Hill (shade in light brown).

Include appropriate details, or you could annotate the map. In the case of the site of Perth you might wish to include the location of the CBD, the steep hillsides and higher ground to the south limiting Perth's growth, the flat land prone to flooding on the west side of the river, the bridges built over the river and the higher ground to the east of the river.

Add the map's title, north arrow, scale and key.

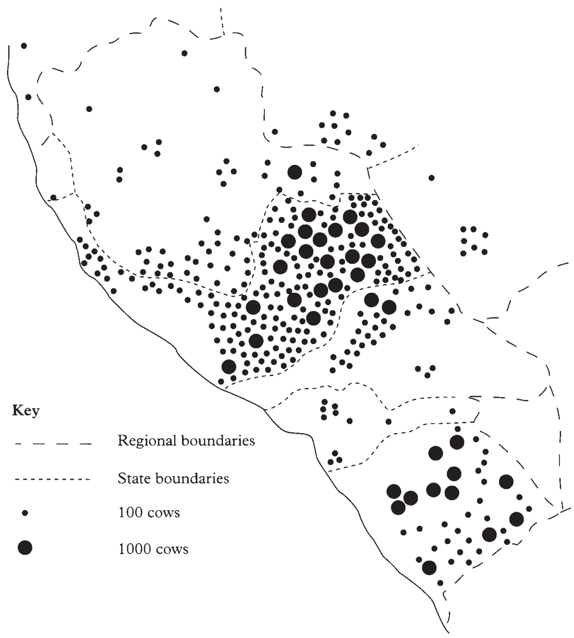
T9

You should submit this work to your tutor as it will be formally assessed. Success in this assignment means you have evidence that you can produce a sketch map.

Dot maps

Dot maps are used to show the distribution of features in an area. The simplest dot maps show the exact locations of a small number of features, e.g. power stations in Scotland. More advanced maps use the density of dots to show information relating to areas, e.g. the location of towns of more than 30,000 people in Scotland. The diagram below shows a typical dot map.

The distribution of cows in an area



Source: *Geographical Techniques*, Liz Taylor, p. 38

Method

- Decide on a suitable number of features for each dot to represent, e.g.
  - 1 dot = village,
  - 1 larger dot = a town, etc.

- Draw the correct number of dots for each area, distributing them evenly and up to the edge of the area.

- Give your map a suitable title and key.

- Obviously symbols other than dots can be used in the construction of simple distribution maps, and by using a number of different symbols you can show more than one distribution on the same map. For example, in a rural area arable farms could be shown as triangles, dairy farms as squares and mixed farms as dots.

A2

1. Refer to an atlas and describe the location of the population of Brazil in 1990.

The location of the population of Brazil (1990)

Source: adapted from *Situations in Human Geography*, J P Cole, p. 9

2. For which of the following distributions do you think the dot method would be the most appropate form of representation on a map? In each case explain your answer.

- Cities in England and Wales
- The location of shops in a town
- The location of hospitals in a health authority area
- Incidences of car thefts in a town
- Population density of the local council areas in Scotland

A2 comment

1. From the map it can be seen that the largest dots are located on the coast in the south-east of the country. These dots correspond to the cities of Rio de Janeiro and São Paulo. (Interestingly São Paulo is not located on the coast, but the size of the dot, corresponding to its population size, makes it look as if it is located there. This is a shortcoming of the dot method.) There are other large dots located along the coast in the north-eastern area, which correspond to cities such as Salvador, Recife, Fortaleza and Belém. There are numerous smaller dots in the south-east, showing that the majority of the population live there. Other larger dots show the location of Brasília, Belo Horizonte, etc. There are very few dots in the sparsely populated interior areas of the Brazilian Plateau and Amazon Basin, the only exception being Manaus.

2. *Cities in England and Wales*
   These could be shown as a dot map similar to the Brazil map in question 1 above, as their distribution could be clearly seen.

   The location of shops in a town
   This would be difficult to show as a dot map as the location of many shops in the CBD would make that part of the map quite busy.

   The location of hospitals in a health authority area
   This should work well as a dot map as, unlike the shops above, there will be fewer hospitals, they will be widely distributed and local health areas will be larger than the area of a town so it should be easy to see the dots.

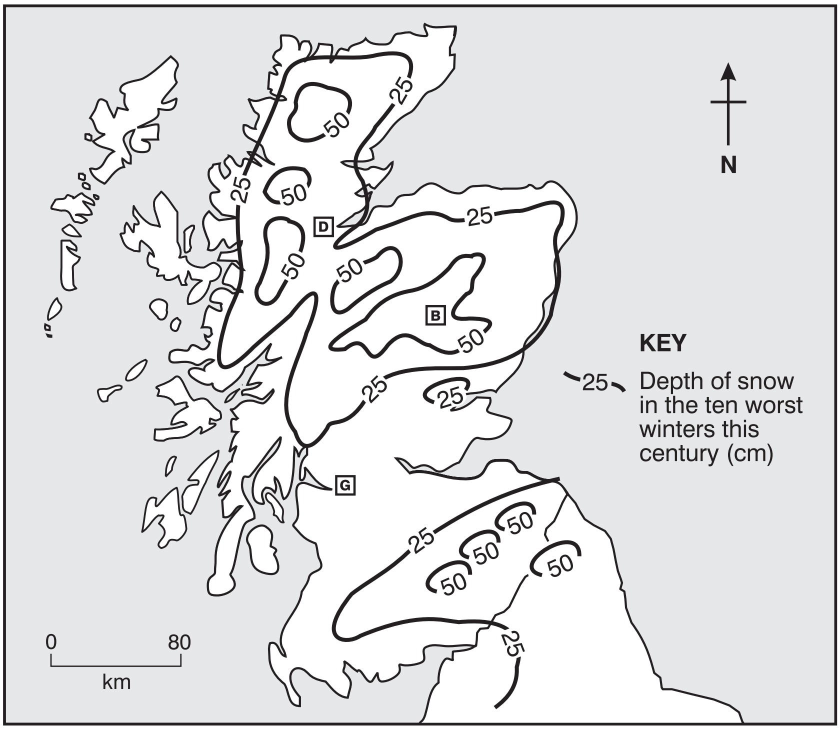
   Incidences of car thefts in a town
   Again this should be easy to show as a dot map for the same reasons as the distribution of hospitals.

   Population density of the local council areas in Scotland
   A choropleth map (shading map) would work much better here as deep shading would show the high population densities.

Isoline maps

Isolines are lines that join together places of equal value. The best known example is a contour line map joining all points of equal height above sea level. Weather data is often shown as isoline maps, e.g. isotherms, which join together all points of equal temperature. The map below shows the average snowfall depth in Scotland for the worst ten years of the 20th century.

Snowfall in Scotland



Source: adapted from *Geography Methods and Techniques*, HSDU

Isoline maps are drawn where data is available for a large number of points in an area. The lines are drawn to join up points of equal value, interpolating between the points where necessary. For example, if an isoline of value 20 has to be drawn between two points of values 18 and 27, it will be drawn nearer to 18 than to 27.

Areas of geography where isolines can be used:

1. Weather, e.g. temperature (isotherms), rainfall (isohyets), atmospheric pressure (isobars)
2. Height, e.g. contour lines
3. Rivers, e.g. speed of flow (isovels).

A3

An isotherm is a line joining together all points of equal temperature. Summer isotherms are usually drawn as red lines and winter ones as blue.

1. On Map 1 draw in the summer isotherms at 14, 15, 16 and 17 degrees centigrade. (Hint: this is the same idea as the join-up-the-dots puzzles you played when you were small. But none of the lines should cross over.)

2. On Map 2 draw the winter isotherms at 4, 5 and 6 degrees centigrade.

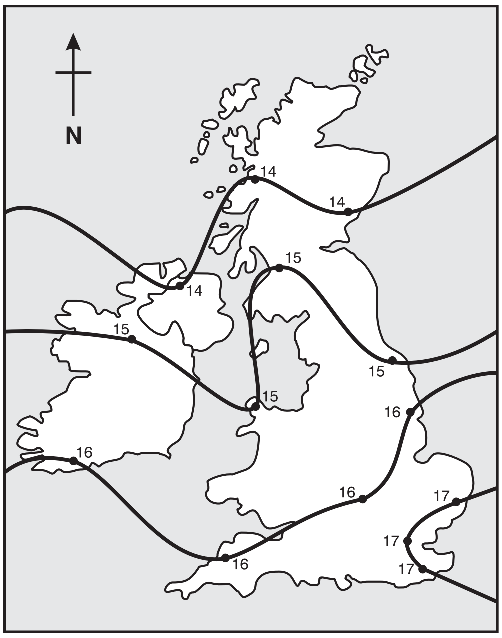
Map 1: Summer

Map 2: Winter

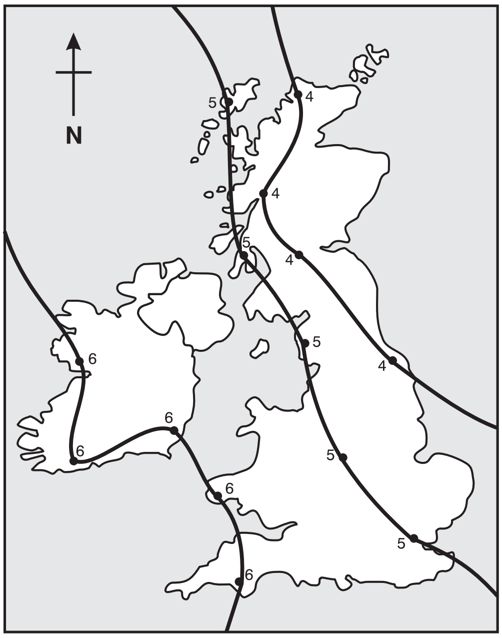
A3 comment

Your maps should look like this.

Map 1: Summer



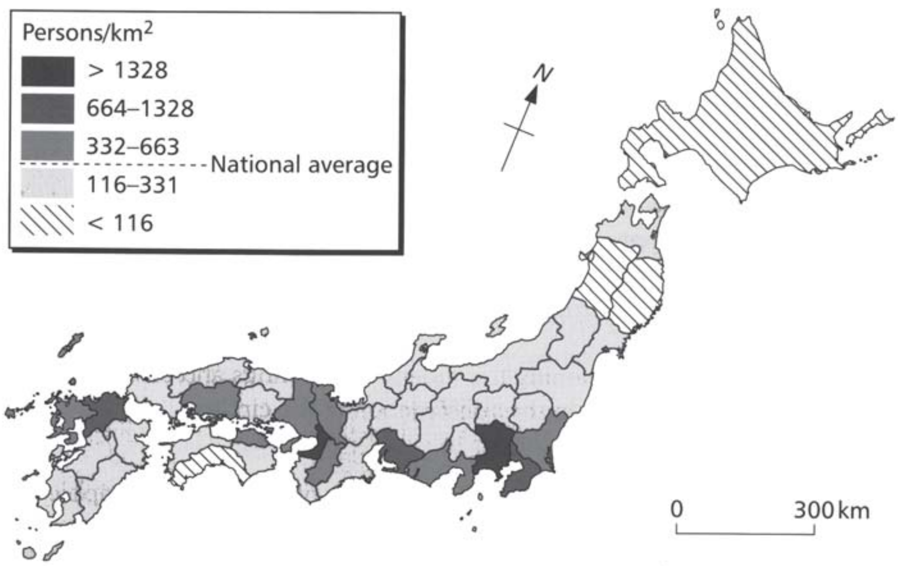
Map 2: Winter



Choropleth maps

A choropleth is a map that shows relative density per unit area. Choropleths are easy to construct and can give data a striking visual impact, for example, to convey population density.

The population density of Japan



Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 46

Method

1. Use a base map where a number of internal boundaries are drawn in.

2. Group the data into a number of categories.

3. Select a system of shading where the highest values are dark and the lowest values are light.

Limitations

- Since each of the areas shows only one value there is the temptation to think that conditions are uniform throughout each area.

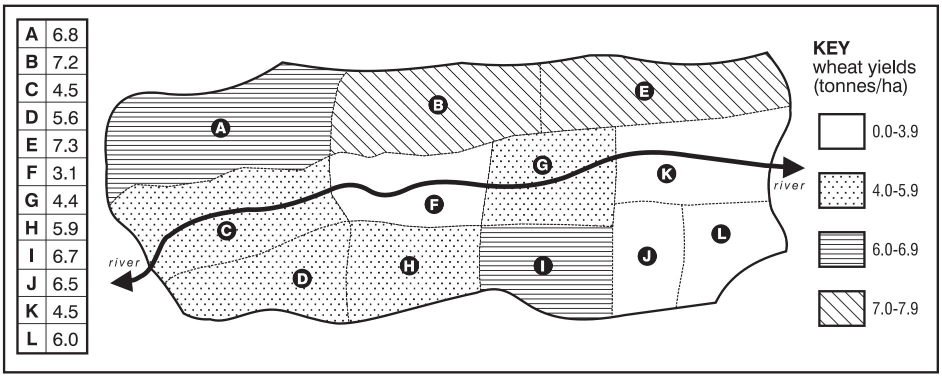
- Where adjacent areas show different values the boundary can appear to indicate a sharp contrast which almost certainly does not occur on the ground.

A4

1. **Wheat yields**

   Complete the map by correctly shading the areas J, K and L.

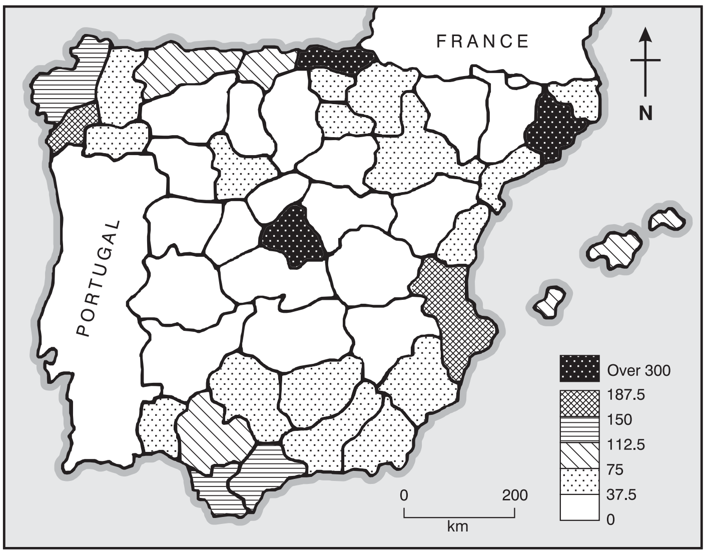
   Fields nearest the river give low yields. Is this statement true? Give a reason for your answer.



Source: adapted from *Geographical Enquiries*, Garrett Nagle and Kris Spencer, pp. 44–5

2. **Spain: population density**

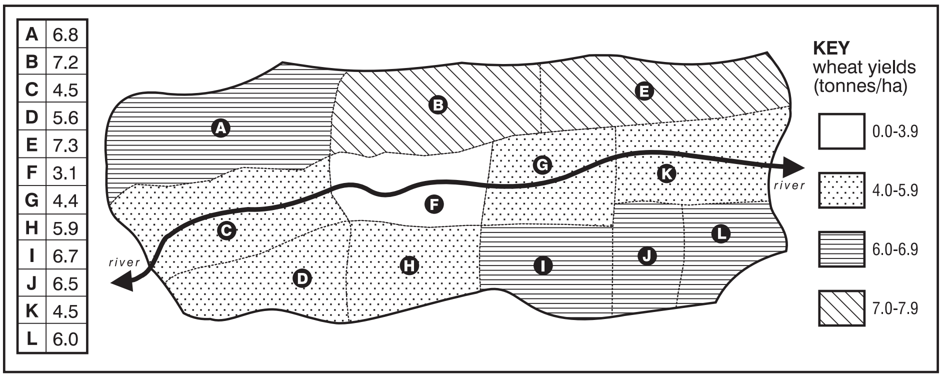
   Describe the population density of Spain from the map (key shows no. of people per km^2).



A4 comment

1. **Wheat yields**

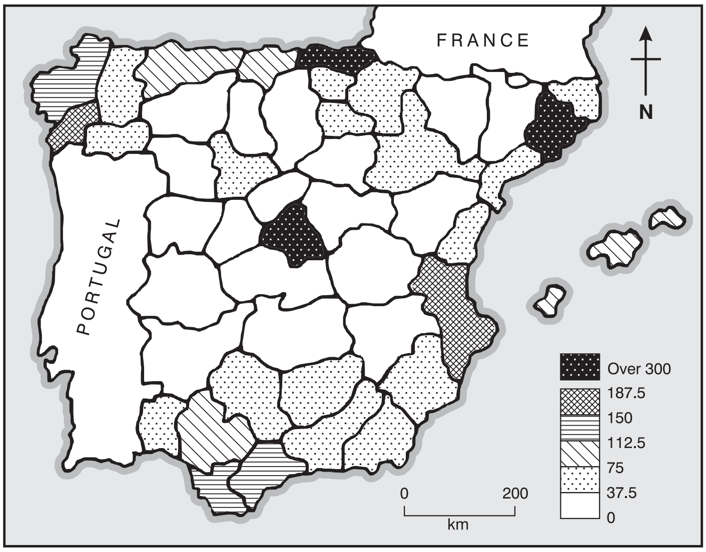
True, because if the land is badly drained or if flooding occurs, the dampness in the soil will reduce the yield.



Source: adapted from *Geographical Enquiries*, Garrett Nagle and Kris Spencer, pp. 44–5

2. **Spain: population density**

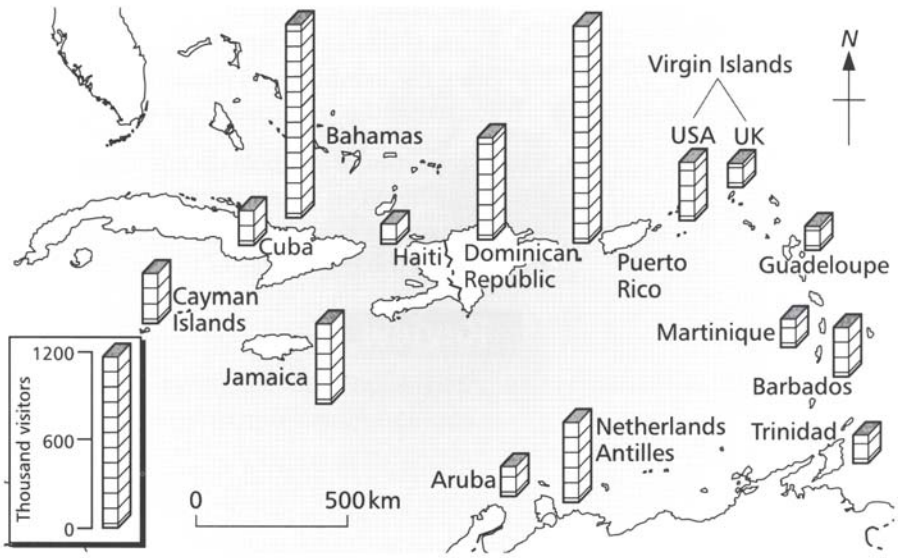
The most densely populated areas are in north-east Spain, northern Spain next to the western border with France, and central Spain around the Madrid area with over 300 people per square kilometre. The central east coast area and north-west Spain are next in density with over 150 people per square kilometre, and the rest of the coastal areas of the east and north-west have densities of between 37.5 and 149 people per square kilometre. Most of the interior of the country and the northern (Pyrenees mountains) areas are sparsely populated with under 37.5 people per square kilometre.



Proportional symbol maps

Proportional symbol maps use symbols such as bars or circles to represent data that varies from place to place. These symbols vary in size according to how large a figure is being represented. These maps can have a good visual impact.

Tourism in the Caribbean



Source: *Skills and techniques for Geography A-Level*, Garrett Nagle, p. 39

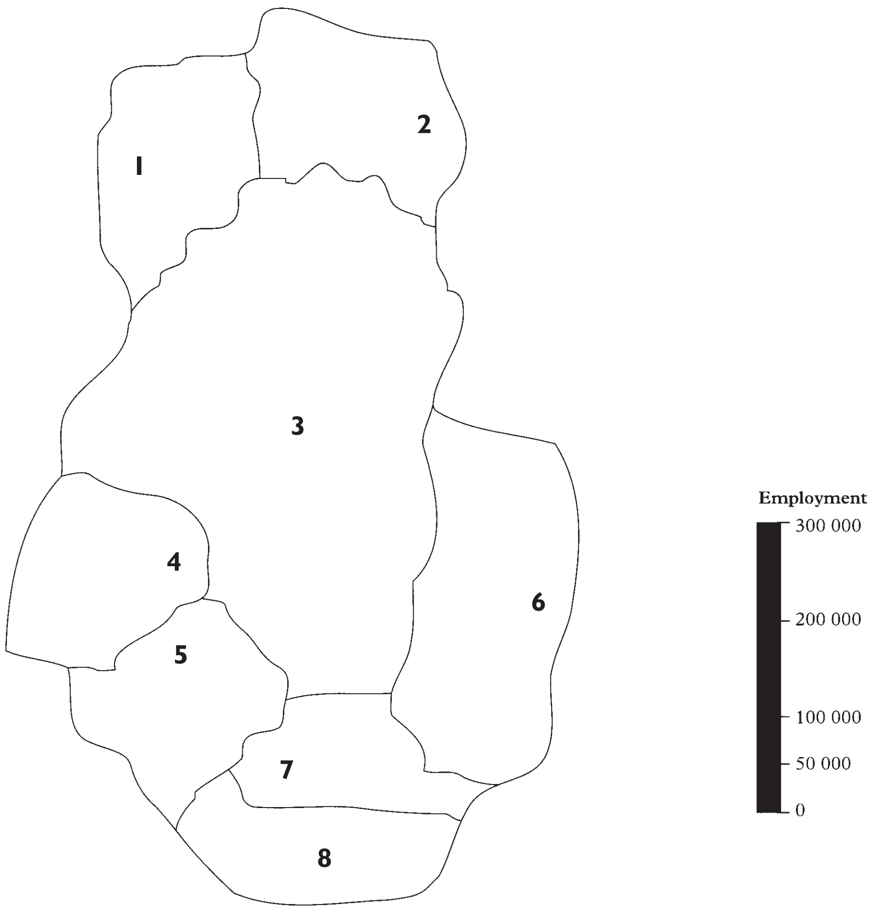
Method

- Draw a base map and choose the type of symbol you wish to use, e.g. bars or circles.
- Work out a suitable scale for the symbols. Spend time choosing this scale. If it is too big the symbols may not fit into the area available, and if it is too small all the symbols will look the same size.
- Draw the symbols to the correct scale and the correct size and colour them to make them stand out.
- Try to keep the symbols as near as possible to the location of the data, although sometimes they may have to be moved to stop them overlapping with some of the other symbols or features.
- To make circles proportional make the square roots of their radii proportional.
- Proportional circles can be subdivided to make pie charts.
- Remember to add the map's key, scale and title.

A5

On the map outline below, draw proportional bars to show the employment levels of each of the following areas.

Area	Employment level
1.	50,000
2.	80,000
3.	500,000
4.	30,000
5.	10,000
6.	300,000
7.	60,000
8.	50,000



Source: adapted from *Geographical Techniques*, Liz Taylor, p. 57

A5 comment

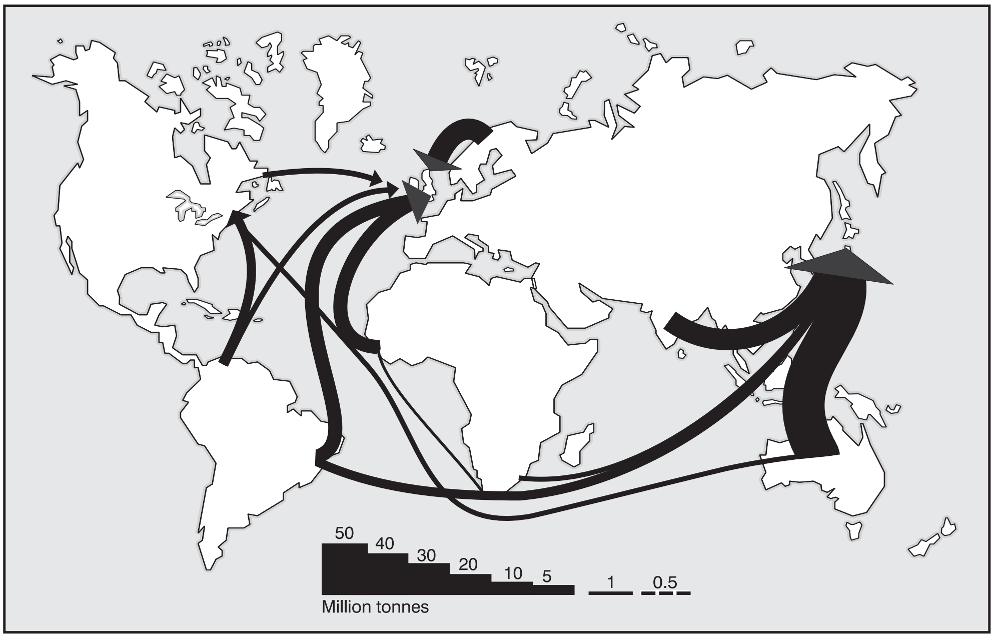
Your map should look like this.

Source: adapted from *Geographical Techniques*, Liz Taylor, p. 57

Flow maps

Flow maps are used to show movement between places, for example, traffic along a route or the movement of migrants or goods between countries. A line is drawn along the route proportional in width to the volume of flow and the direction can be shown as an arrow. The background of the map should be kept as simple as possible. A suitable scale should be chosen, ensuring the finished map is neither too empty nor too cluttered.

Iron ore trade



Method

1. Obtain or draw a base map of your area.
2. Decide on a suitable scale by examining the range of data.
3. Draw the flow lines in pencil first. Try not to let them overlap or cross over but this might be unavoidable.
4. Try to keep the arrows the same width between the source and the destination. You might have to take several measurements along the line to ensure this.

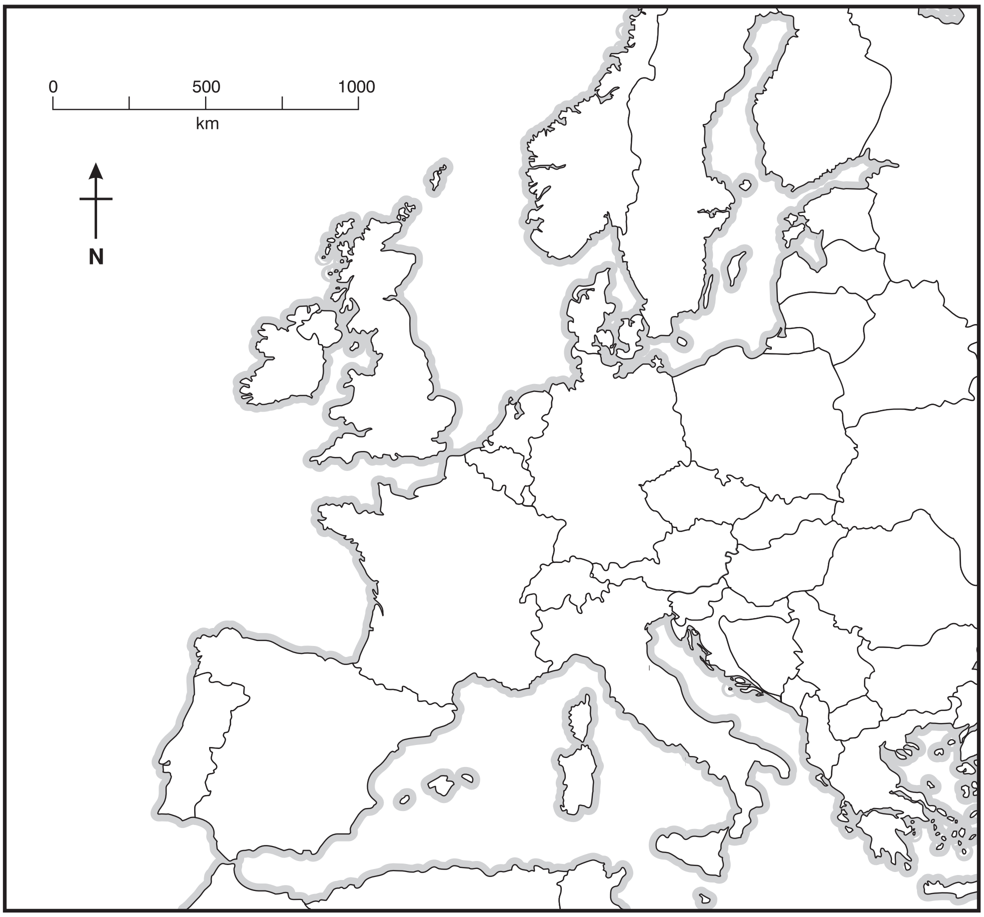
Areas of geography where flowlines can be used

1. Urban geography, e.g. traffic counts, pedestrian counts
2. Trade and aid, e.g. movement or flow of goods between countries
3. Population studies, e.g. migration patterns.

A6

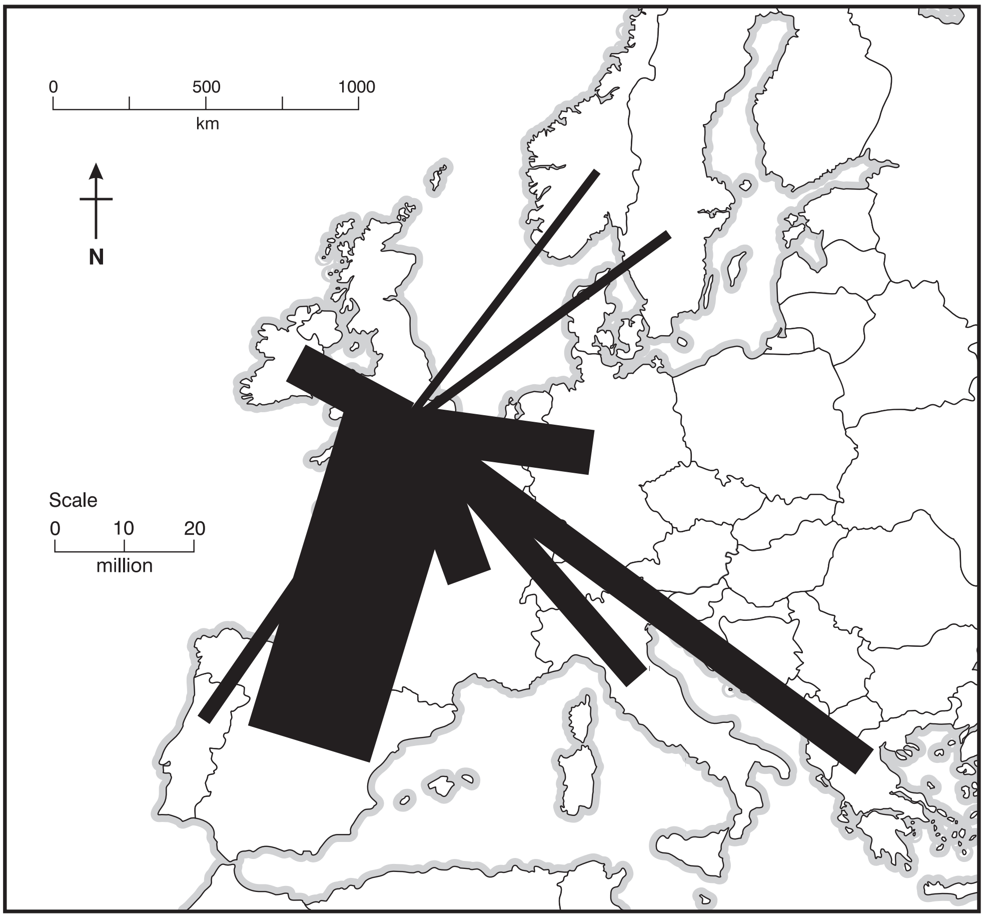
Draw a proportional flow map to show the air passenger movements between Britain and selected European countries. Choose a suitable scale for this map.

Country	Number of air passenger movements in 1995 (millions)
France	6.6
Germany	6.5
Greece	4.5
Irish Republic	6.0
Italy	4.0
Norway	1.2
Portugal	2.1
Spain	18.3
Sweden	1.4



A6 comment

Your map should look like this.



SECTION 5

Introduction

The map you will be given in the external examination will be an example from the OS 1:25000 Explorer map series and could be located anywhere in England and Wales. You will also be given an atlas. The location of the map area is found on the cover of the OS map. It is important to find that area in the atlas, as this will give you a better picture of the location of your area. You may also be able to get additional material from other pages in the atlas which can help you answer the examination question, e.g. from the pages of climategraphs or climate statistics, farming areas or land uses in Britain, nearby urban areas, population densities, nearby enterprise areas, National Parks, etc.

It is easy to think you are good at mapwork skills because you may have had plenty of practice interpreting them for Standard Grade, Higher Geography or similar examinations. However at a higher level, map skills involve an ability to look at the map and then be able to imagine what the area is like. For example, you should be able to tell what the housing is like from the street pattern or describe what the shape of the valley is like from the contour lines.

You can do two things to improve your use of OS maps:

1. Check you understand all the basic techniques:

   - 4-figure grid references for an area
   - 6-figure grid references for a point
   - Measuring distance and area
   - Height – contours, spot heights, shapes
   - Compass directions
   - Symbols (so you do not have to keep looking at the key!)

2. Look at maps – use any OS map – and think:

   - What is the land like in this square?
   - What is the shape of the land?
   - Can I describe the patterns of relief and drainage and other physical features seen on the map?
   - Can I describe settlement patterns, communications and other human features seen on the map?
   - Where would be a suitable site for a motorway, bypass, retail park, hotel with leisure facilities, industrial estate, etc.?

How to use Section 5

Section 5 has two purposes:

1. To help you describe and explain the distribution of physical and human features from an OS map and draw simplified sketch maps that can be used as NAB evidence.

2. To help you undertake a simple ‘planning’ exercise, e.g. find the best site for a motorway, bypass, retail park, hotel with leisure facilities, industrial estate, etc.

Practice in both of the above will be good preparation for your external examination.

The maps used in this unit are 1:25000 OS map extracts from past Standard Grade Geography papers and, although they are not identical to the Pathfinder and Explorer maps, they are still 1:25000 maps with the same map symbols. These maps have been chosen for availability purposes. Your tutor may not have the Pathfinder or Explorer series from previous CSYS or Advanced Higher examinations but will almost certainly have the examination map extracts that you can practise your mapping techniques on.

One word of caution: make sure you familiarise yourself with an urban area and a rural area on a Pathfinder and/or Explorer map before you sit your examination.

Where to locate an industrial estate

The following LOCATIONAL factors are required.

Site

- Plenty of flat land to build on easily. Remember each grid square is one kilometre square so take this into account when considering the size of your estate.
- Room for expansion, storage and parking.
- Good solid land to build on, e.g. no marshland or streams nearby.
- Cheap land, e.g. at edge of town, brown field site or derelict land (look for evidence that industry may have previously been located there – disused mines, spoil heaps, old railway sidings, etc.)
- Can you identify any other factories that may have industrial linkages with your estate?

Good communications

Roads. Name them.
- How far away are they from your site? (Compass direction and distance.)
- Are they dual carriageways or can they be widened to become dual carriageways?
- If there is a motorway nearby does it also have a slip road nearby or could one be built with ease? (You would have to take local authority planning permission into account for this.)
- Now look at your atlas.
- Do the roads in the map extract link up with main national trunk roads or motorways? If so which conurbations are they connected to?
- Is it possible to landscape the edge of the estate to hide it from the road?

Railways
- Is there a railway nearby?
- Even better, is there a railway station?
- How would goods get from the railway station to the industrial estate, i.e. are there good road links from the railway station to the industrial estate?

Airports
- Is there an airport nearby?
- Is it an international airport or a small local one?
- Are there good roads links between the industrial estate and the airport?

Ports

- Check your atlas.
- Is there a port nearby?
- Is it a large port or a small one?
- Are there good roads links between the industrial estate and the port?

Labour

- Is there a labour supply nearby to work on the estate, e.g. a suburban area?
- Are there good communications so that the workforce can travel to work easily?
- Now look at your atlas. Can you identify any areas nearby that have suffered from factory closures and could give a labour supply, or can you identify any area with high unemployment that might get government grants?
- Can you identify any other industrial estates that might be a pool of experienced labour for your industrial estate?

Market

- Is there a ready market nearby, i.e. a large conurbation? Look at your map or the atlas for this information.
- Are there good communications to the market?

Environment

- Is it a pleasant environment that will attract workers into the industrial estate?
- Is it a pleasant environment that will attract managers into the locality, e.g. golf courses, open countryside, good housing (look at housing density and location of housing estates in nearby towns), or does it have good schools and other amenities?

Using the 1995 Standard Grade General examination map extract of Perth, write down the locational factors involved in siting the Inveralmond industrial estate at 0926. You can also list any disadvantages you see.

AI comment

Site

- Plenty of flat land to build on easily. This area covers approximately one kilometre square with room for expansion, storage and parking to the south-west. But expansion is restricted to the east by the dual carriageway running north at 096262. (Give lots of map evidence!)
- Good solid land to build on but there is a stream nearby to the north which may limit expansion and the site may be prone to flooding as it is low-lying – only about 10 metres above sea level. (Give lots of heights!) The burn also restricts northward expansion.
- Cheap land, e.g. at edge of town? Perhaps true, but there is no map evidence to give you a clue to this.
- Other factories that might have industrial linkages with the estate are the distillery at 097260, the other industrial estate at 100260, and the works at 100250 and 113240.

Good communications

Roads

- The A9 trunk road and dual carriageway marks the southern and eastern edge of the estate and a roundabout connects with the estate at 098262. There is also a smaller A-class road which runs through the estate and connects it to the A9.

- Now look at your atlas. The A9 trunk road is the main arterial road running south to north through eastern Scotland from Stirling to Inverness and the north of Scotland. This road links up with the M9 and the central belt towns and cities of Stirling, Glasgow and Edinburgh, as well as with Inverness in the north.

Railways

- There is a railway nearby although there is no railway station and no evidence to suggest that a railway would be useful here.

Airports

- There is only an airport for light aircraft nearby but the international airports of Glasgow and Edinburgh are well connected by motorway links to Perth and are about 50 kilometres away. Dundee airport is about 30 kilometres away.

Ports

- Check your atlas.
- Perth is a small port although there is no evidence of this from the map. It probably would not be convenient to transport light goods by water.

Labour

- There is a labour supply nearby that could work on the estate, e.g. a suburban area of Tulloch (095254) and Muirton (100255). Both these area are well connected by road (A85 and A912 respectively) to the estate.

- Now look at your atlas. The central belt to the south and the Dundee area to the north-east have suffered many factory closures and there are likely to be grants and incentives available to open up factories in this general area. Also Perth has a number of other industrial estates that might be used as a pool of experienced labour for a new industrial estate.

Market

- Again there are good communications in the shape of the M9 to the central belt market.

Environment

- This is a pleasant environment to attract workers into. The industrial estate has open land to the south-west and a river with wooded banks to the north. Perth itself has a pleasant environment to attract people into the locality, e.g. golf courses at 1222, open countryside to the south, the river for fishing (1026), Scone Palace at 1126, good housing (look at the housing density and location of housing estates in the nearby towns), e.g. Muirton at 1025 and at 1326. Both of the latter areas have schools, low housing densities and open countryside surrounding the housing estates.

Where to locate a retail park

Retail parks are a recent phenomenon. They are characterised by large retail outlets selling a variety of goods such as houseware, electrical goods and DIY , as well as by car showrooms. They generally rely on the customer visiting by car, as the goods purchased are often bulky and difficult to transport, so easy car parking is essential.

Many of the locational factors for a retail park are similar to those required for an industrial estate.

Site

- Plenty of flat land to build on easily. Remember each grid square represents 1 km^2, so take this into account when considering the size of your retail park.
- Room for expansion, storage and parking.
- Good solid land to build on, e.g. no marshland or streams nearby.
- Cheap land, e.g. at edge of town, brown field site or derelict land; or government help available (development area?).

Good communications

Roads
- How far away are they from your site? (Compass direction and distance.)
- Are they dual carriageways or can they be widened to become dual carriageways?
- Is there a motorway nearby?
- Now look at your atlas.
- Do the roads in the map extract link up with main urban areas? If so which ones are they connected to?

Railways
- Is there a railway nearby with a railway station?
- How would customers get from the railway station to the retail park, i.e. are there good road links from the railway station to the retail park?

Customers

- Are there customers nearby, e.g. living in a suburban area?
- Are there good communications so that the customers can travel easily to the retail park?
- Can you identify any other urban areas that might provide a pool of experienced labour for your retail park?

Environment

- Is it a pleasant environment that will attract customers?
- Will the local residents complain about the noise, litter and increased traffic caused by the retail park? In other words, is the retail park on the doorstep of a large housing area?
- Is it possible to landscape the edge of the estate to hide it from the road?

A2

Obtain the Standard Grade Credit examination map extract of Paisley. The 'works' at 4564, 4563 and 4463 was the Linwood car factory. Today it is the Phoenix Retail Park. Explain why it is located here.

A2 comment

The land is good for building on. It is flat, with room for parking and expansion and it was a brown field site from the former car factory, so would be cheap to buy.

The A740 dual carriageway is located about half a kilometre to the north of the retail park and a major road enters the park via a roundabout from the A740. This road links onto the M8 motorway, which on looking at the atlas you will see connects with Glasgow 10 kilometres to the east and Greenock 20 kilometres to the west. The retail park is also on the edge of the major urban area of Paisley. A railway runs to the south of the retail park but there are no stations nearby (the nearest is in Paisley at GR 484642, about 3 kilometres away).

As previously stated the retail park is ideally located for a large pool of potential customers. The Paisley housing estates of Ferguslie Park, Elderslie and Foxbar are located within 2 kilometres surrounding the park. All these areas are well connected to the park.

The CBD of Paisley is located at GR4864, about 3 kilometres away, and is not likely to offer any competition for the park. This is due to the likelihood that lack of space and car parking in the town centre means Paisley's CBD will have high-order chain and department stores rather than the large shopping units found in retail parks.

The closing of the car factory would have resulted in high unemployment but would also have created a potential pool of labour for the retail park. Further, the government would have offered incentives for big retail outfits to locate here in order to reduce unemployment.

The fact that the site was formerly a car factory means that the residents would have few complaints about changing the function to a retail park as this would be cleaner and quieter than the factory. The roads were probably already built for the factory, so local residents would be used to the traffic. Indeed, if the area had been landscaped like many retail parks are, the environment would probably have been substantially improved since the factory closure.

Where to locate a hotel and leisure centre

A huge range of people visit hotels that have leisure facilities. They may be holidaymakers visiting the area for a week or fortnight, business people spending a few days in the area returning to the hotel in the evenings or entertaining clients, or delegates attending business conferences. The hotel will wish to cater for all these people.

Site

- Pleasant surroundings, e.g. open countryside, location beside a river, woodland, scenic area, etc.
- Cheap land is not necessarily important here, the emphasis being on the quality of the site. Extra land values tend to get passed on to the customer in their hotel bills!
- Near to sites of interest, e.g. local villages, large cities with interesting museums, shopping facilities, theatres, castles, monuments, country parks, National Parks, etc. Look at your atlas for these.
- Near to centres of activity, e.g. a river for fishing, golf courses, footpaths, etc.

Communications

Roads, railways (remember there must be a station nearby!) and an airport (especially important for business people and conference delegates).

Labour

A pool of local labour to work in the hotel.

A3

Obtain the 1998 Standard Grade General examination map extract of Manchester Airport. A hotel and leisure centre has been built on the Bank House Farm site at GR831829. Explain why it has been located there. You may also list any disadvantages.

A3 comment

This is a pretty location for a hotel. It is at the top of a steep slope that drops down to a small river. Next to the hotel and on the opposite bank of the river is a wooded area. The river meanders gently through the area. Open fields surround the hotel on the other sides and there is a long driveway of about 250 metres separating the hotel from Altrincham Road (GR827828).

National Trust property surrounds the hotel and there are local places of interest, e.g. Giant's Castle at 825835 and Styal Country Park at GR 833834. There are many footpaths running through the area, e.g. at Morley and Morley Green (GR 8282).

Manchester city centre, with all the attractions of a large city, is located about 7 kilometres to the north of the hotel and Altrincham Road and the M56 connect the hotel to it.

It is relatively easy to travel to the hotel. Altrincham Road is connected to the M56, the Manchester Ring Road (look at your atlas), approx. 5 kilometres to the west. There is a railway station at GR841835, a short taxi ride to the north-east of the hotel. Manchester Airport is very close to the hotel, located 4 kilometres away and well connected to it. An added advantage is that the hotel is not on the flightpath of the planes so will not get aircraft noise. (The flightpath is along the line of the runway – in this case NE–SW.)

Employees to the hotel will probably come from Pownall Park (GR 8381) and Lacey Green (GR8482).

Where to locate a bypass

When a town centre becomes too crowded with vehicles, a bypass has to be built to take away some of the traffic. Bypasses are normally longer than the original route but the faster moving vehicles ensure that the journey time is shorter. Building a bypass is not easy. Money has to be found, suitable routes planned out and discussions held with people whom the road might affect.

The following factors are involved:

Site
- Is the land cheap to buy?
- Is the bypass route reasonably short? The shorter the bypass the more inexpensive it is to construct.
- Does it avoid steep slopes or do embankments or cuttings have to be built?
- Does it cross rivers so that bridges might have to be built?
- Is the ground firm or does it have marshland that needs to be drained?
- Can it be built on the site of an existing route, e.g. an old railway or drained canal?

Conservation issues
- Do many buildings have to be demolished to build the bypass?
- Does it avoid the best farmland?
- Does it involve cutting down trees?
- Does it avoid parkland and beauty spots?
- Does it avoid footpaths or bridleways?
- Is it close to built-up areas?
- Will it have many slip roads and roundabouts that waste land?
- Will the new road scar the landscape?

Economic issues
- Are there hotels on the bypass route that will benefit from the new road?
- Is the bypass so far away from the town that the town centre will be affected by the reduction in passing trade?
- Is the bypass near to industrial estates or retail parks that could take advantage of it?

A4

Using the 1995 Standard Grade General examination map extract of Perth, write down the advantages and disadvantages involved in building the road around Perth from the roundabout at GR083227 to the roundabout at GR 098262.

A4 comment

Advantages

- The bypass is reasonably short, about 6 kilometres long, around the outskirts of Perth.
- It crosses only one major river.
- The road is separated from residential areas by open field, e.g. GR 085240.
- The hotel at Letham farm (GR 088245) might win extra business.
- The road runs next to Inveralmond industrial estate (GR 0926) and a distillery (GR 097260).

Disadvantages

- Although the land is gently sloping a number of cuttings (GR 083230 and GR 083236) and embankments (GR 087255) were required.
- One major bridge and a few small bridges have had to be built, e.g. at GR 082238 and GR 086248.
- The road cuts through farmland (GR 088245) at Letham farm.
- The road cuts through woodland (GR 088254).
- The road runs beside the site of an ancient building (GR 084251) called Huntingtower Castle.

T10

Your tutor will give you an assessment based on a 1:25000 OS Pathfinder or Explorer map. You should submit this work to your tutor as it will be formally assessed. Success in this assignment means you have evidence that you can extract and interpret information from 1:25000 OS maps.

SECTION 6

How to draw a cross-section

A cross-section is a view of a landscape as it would appear if it was sliced open or seen sideways on.

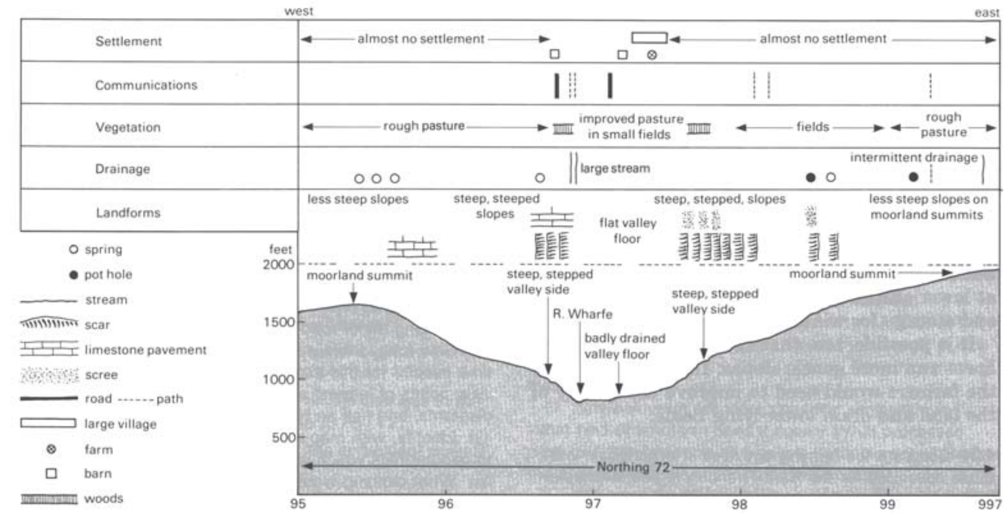
Method

- Place the straight edge of a piece of paper between the two end points.
- Mark off on the paper every contour and include any important geographical features.
- Align the straight edge of the paper against the horizontal line on graph paper which is exactly the same length as the line of section.
- Use a vertical scale of 1 cm:100 m or 1 cm:50 m.
- Mark off with a small dot each of the contours and the geographic factors.
- Join up the dots with a freehand curve.
- Label the features. Remember to label the horizontal and vertical scales, title, etc.

How to draw a transect

Land-use transect
In order to be able to compare the height and shape of a slope with land use to see if there is a relationship, a land-use transect can be drawn. Drawing a cross-section and adding information about the slope underneath it does this.

The diagram below shows how land use and additional information have been taken from a completed land-use survey map and added to the cross-section to make it into a transect.



Source: *Ordnance Survey Mapwork*, B D R Worthington and Robert Gant, p. 33

Shading the cross-section and colouring or adding the information about the land use below the cross-section shows the direct relationship between the height and slope of the land and the land use.

A1

Is there a relationship between relief and land use?

Diagram 1 on the next page is an extract from an OS map. The contour lines are shown at 50-metre intervals.

- Draw a cross-section from A to B in the frame provided in Diagram 3 (on page 189).

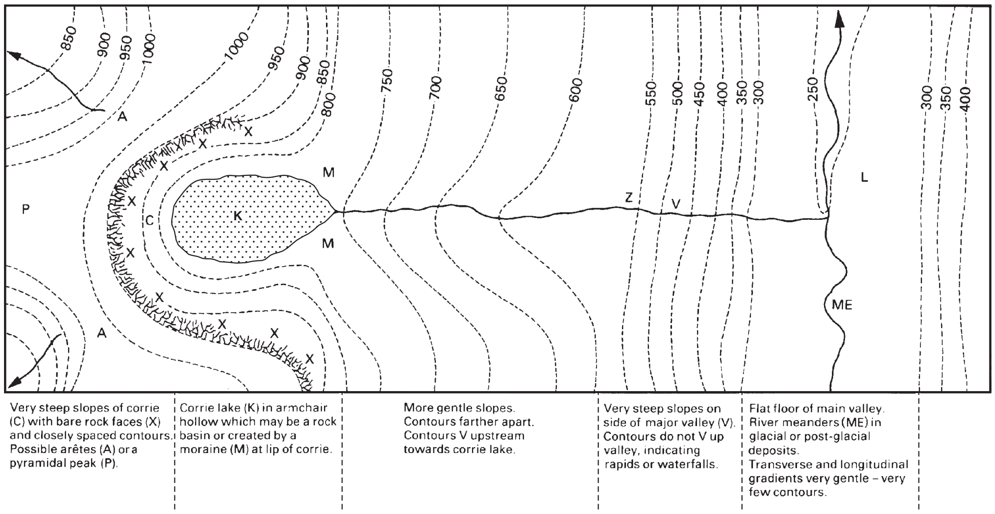
Underneath Diagram 1 is information about the glacial features found at different parts of the slope.

- Add the glacial features (letters) from the information given to the first row underneath the cross-section.

Diagram 2 is a land-use survey taken of the extract. The different land uses are shown in the key.

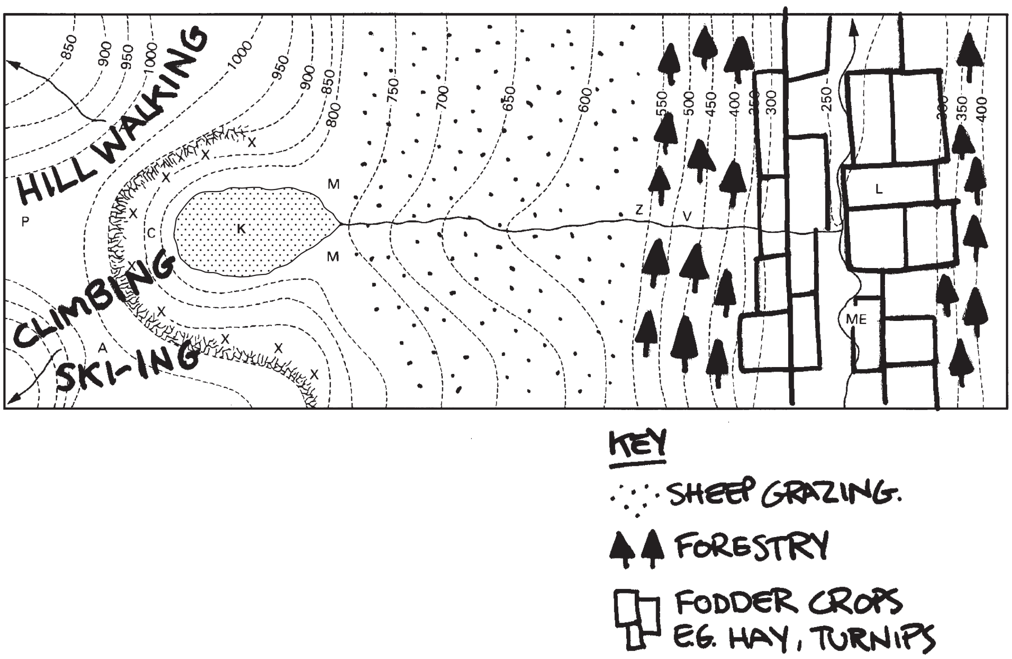
- Add this information to the second row underneath the cross-section.
- Explain the relationship between relief (and where appropriate, glacial feature) and land use in the space provided below.

Diagram 1



Source: *Ordnance Survey Mapwork*, B D R Worthington and Robert Gant, p. 35

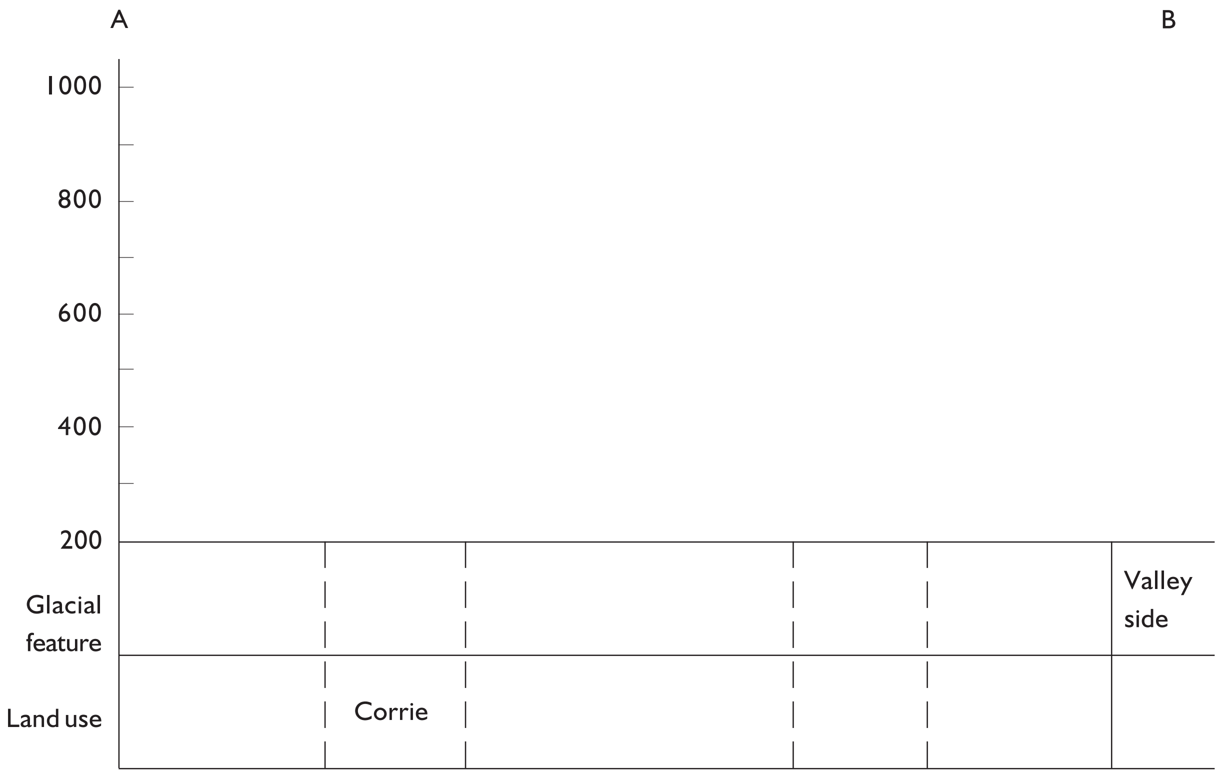
Diagram 2



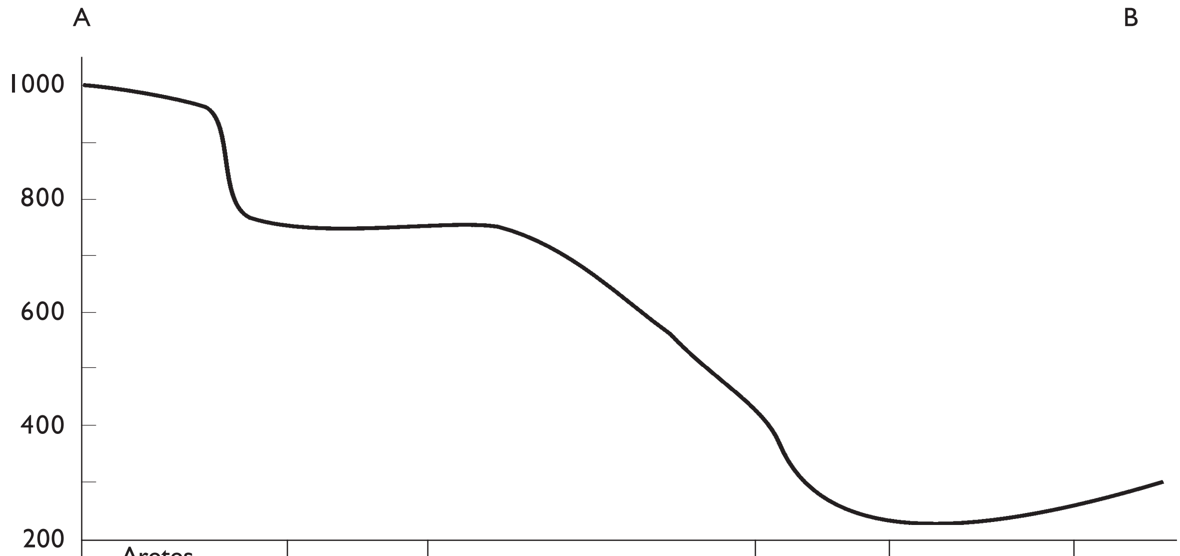
Source: adapted from *Ordnance Survey Mapwork*, B D R Worthington and Robert Gant, p. 35

Diagram 3

Land-use transect exercise



AI comment



Glacial feature	Aretes pyramidal peaks	Corrie	Alp	Hanging valley	U-shaped valley	Valley side
Land use	Hill walking climbing skiing	Corrie	Sheep grazing	Forestry	Fodder crops	Forest

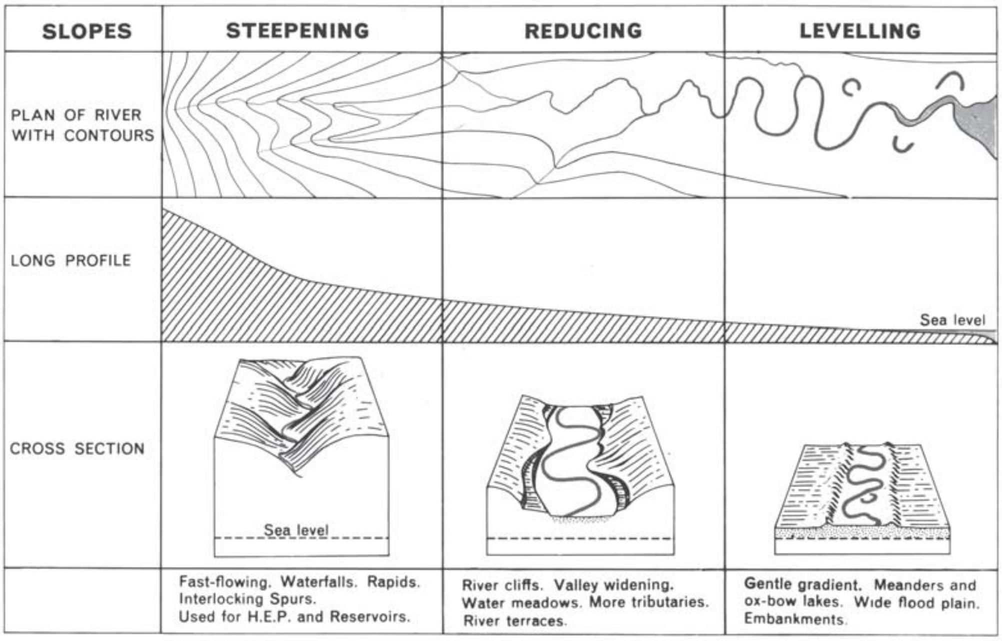
The steep slopes and land above 850 m to the west of the area are used for recreational purposes, such as hill walking, climbing and skiing. As this area has bare rock surfaces and consists of corries, aretes and a pyramidal peak, it cannot be used for any economic purpose. It would however provide an interesting scenic area for the above activities. The slopes between 550 and 800 m are used for sheep. Sheep can graze on these more gentle slopes and are able to withstand the colder weather experienced at this height. The steeper slopes on the U-shaped valley between 300 and 550 m are used for forestry. Conifer trees can be planted on the thinner soils found on these slopes. Fodder crops are grown on the sheltered, flat valley floor which probably has the best soils.

River profiles

One method of investigating how a river changes from its source downstream is to construct a river profile or long profile. This is done in a similar way to drawing a cross-section.

Method

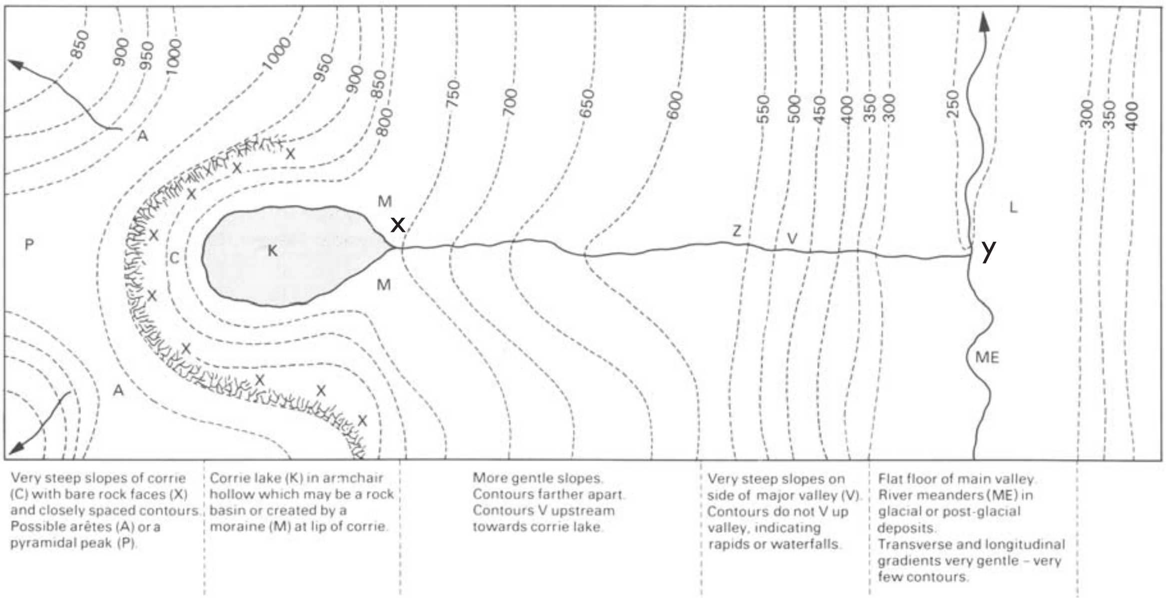
- Place the straight edge of a piece of paper along the first stretch of the river.
- Mark off on the paper each time a contour crosses the river and include any important geographical features.
- Twist the paper around so that the straight edge will follow the curvature of the river, still making sure you are marking off the contour lines.
- Align the straight edge of the paper against the horizontal line on a sheet of graph paper that is exactly the same length as the river profile.
- Use a vertical scale of 1 cm:100 m or 1 cm:50 m.
- Mark off with a small dot each of the contours and the geographic factors.
- Join up the dots with a freehand curve.
- Label the features. Remember to label the horizontal and vertical scales, title, etc.



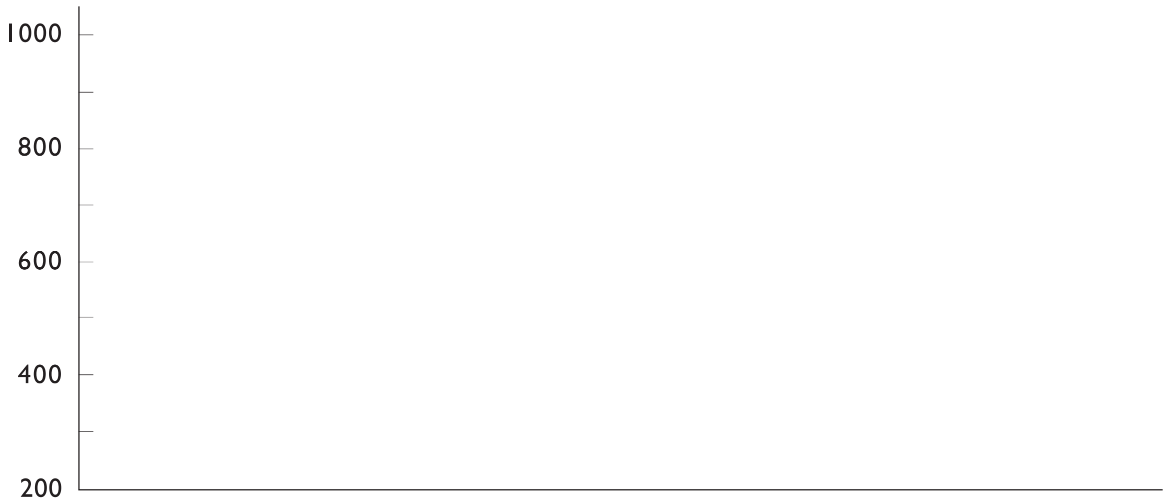
Source: *Map Reading and Interpretation*, P Speak and A H C Carter, p. 23

A2

Draw a long profile of the river from x to y.



Source: adapted from *Ordnance Survey Mapwork*, B D R Worthington and Robert Gant, p. 35



A2 comment

Your diagram should look like this.

SECTION 7

Introduction

After working through Skill areas, *Statistical awareness* and *The production and interpretation of maps and diagrams,* you should now be able to attempt the third Skill area, *Fieldwork survey/ measurement and recording techniques*.

As fieldwork is an essential element in the study of Geography, there are obviously many high-quality textbooks on this topic. Ask your tutor to recommend some to you. These books will give you an idea of the different techniques to use in the field. While undertaking your fieldwork you should bear in mind the elements of fieldwork which are of particular interest to you and on which you can readily obtain data; these may be good foundations upon which to build your Geographical Study.

How to go about fieldwork

1. Plan your investigation, where you want to go, what you want to find out about, why you want to find out about this, etc.
2. Read a fieldwork textbook for the appropriate techniques to use.
3. Consult your tutor about your site and the techniques to use.
4. **Ensure you will be safe when you undertake your fieldwork**. Never go out alone, make sure other people know where you are going and when you will be back. Do not take any risks. **Ask your tutor for advice on safety issues**. He/she will be aware of local policies regarding fieldwork. Do not embark on fieldwork without notifying the appropriate authority.
5. Undertake the fieldwork.
6. Analyse and present the information collected. This can be done either as several A3-size sheets of paper so that your work will be similar to the CSYS Geography analysis sheets or as a mini-project on folio or A4 paper. It is best to store your fieldwork recordings and classwork analysis in a separate ring binder so you have evidence that you have undertaken and analysed the fieldwork for external moderation.

Obviously it is impossible to undertake 'distance learning' fieldwork! So what we have here are some suggestions about fieldwork excursions that you can undertake. Each of these suggestions attempts to cover several fieldwork, mapping, statistical and graphical techniques so that you will have plenty of evidence to show that you are familiar with as many fieldwork survey/ measurement and recording technique as possible.

Each of the suggestions has a summary sheet based on the '*Arrangements*' document, which shows the number of techniques covered. Store the summary sheets in your fieldwork binder with your fieldwork projects as evidence of the techniques used. Again this is important for external moderation.

The study of a retail environment

A1 Investigation summary sheet

Skill area: Fieldwork survey/measurement and recording techniques

Physical topics

- ❑ Morphological mapping
- ❑ Vegetation sampling
- ❑ Slope analysis
- ❑ Stream analysis
- ❑ Meteorology in the local setting
- ❑ Soil profiles and characteristics
- ❑ Pebble analysis

Human topics

- ❑ Rural land-use mapping
- ✓ ***Urban land-use mapping***
- ✓ ***Traffic, pedestrian and environmental quality surveys***
- ✓ ***Questionnaire design and implementation***
- ✓ ***Use of secondary sources***
- ✓ ***Reilly's law of retail gravitation***
- ❑ Huff's probability law
- ❑ Nearest neighbour analysis

Skill area: Statistical awareness

- ❑ Sampling – random, systematic, stratified
- ✓ ***Handling different data types*** – ***nominal***, ordinal, interval, ratio
- ✓ ***Graphical presentation of data*** – systems diagrams, logarithmic graphs, kite diagrams, scattergraphs, polar graphs, triangular graphs, dispersion diagrams, ***bipolar analysis***
- ❑ Descriptive statistics – mean, median, mode, standard deviation
- ❑ Statistical testing – nearest neighbour, student's t-test, chi-square, Spearman's rank correlation, Pearson's product moment correlation coefficient, linear regression

Skill area: The production and interpretation of maps and diagrams

- ✓ ***Design and layout of maps*** – principles of lettering, linework, ***shading***, dot maps, isoline maps, choropleth maps, proportional circles, divided proportional circles, ***flow maps***
- ✓ ***Interpretation of OS maps and related data***
- ❑ Topographical analysis – cross-sections, transects, river profiles

AI techniques

Fieldwork technique	Information to collect	Production/ interpretation of maps	Statistical/ graphical techniques
Urban land-use survey. Obtain a base map of the local high street shopping centre. Walk along street taking notes of the names and types of shops.	Classify shops according to shop type and order of service. Type includes food & drink; pubs & catering; clothes & shoes; financial (banks & building societies); specialist (chemist, bookshop, furniture); and services (opticians, hairdressers).	Land-use map of shop type classification and order of service.	Production and analysis of bar graphs and/or pie charts to show shop type and order.
Pedestrian count. Traffic count.	Tally tables of pedestrian flows in centre and on both edges of high street. Traffic counts at morning/evening rush hour and midday.	Flow maps on base maps to illustrate the pedestrian and traffic count.	
Customer questionnaire.	Questionnaire should include: Where do you live? How often visit? Method of travel? Type of services used? Amount of money spent?	Sphere of influence map of customers' houses. Analysis of this map.	Bar graphs/pie charts of customer responses. Description and explanation of findings.
Bipolar analysis.	Questionnaire should include questions for bipolar analysis.	Completed bipolar graph table for all the customer responses. Analysis of responses.	

A river study

A2 Investigation summary sheet

Skill area: Fieldwork survey/measurement and recording techniques.

Physical topics

- ❑ Morphological mapping
- ❑ Vegetation sampling
- ❑ Slope analysis
- ✓ ***Stream analysis***
- ❑ Meteorology in the local setting
- ❑ Soil profiles and characteristics
- ✓ ***Pebble analysis***

Human topics

- ❑ Rural land-use mapping
- ❑ Urban land-use mapping
- ❑ Traffic, pedestrian and environmental quality surveys
- ❑ Questionnaire design and implementation
- ❑ Use of secondary sources
- ❑ Reilly's law of retail gravitation
- ❑ Huff's probability law
- ❑ Nearest neighbour analysis

Skill area: Statistical awareness

- ✓ ***Sampling – random***, systematic, stratified
- ❑ Handling different data types; nominal, ordinal, interval, **ratio**
- ✓ ***Graphical presentation of data*** – systems diagrams, logarithmic graphs, kite diagrams, ***scattergraphs***, polar graphs, triangular graphs, dispersion diagrams, bipolar analysis
- ✓ Descriptive statistics – ***mean***, median, mode, ***standard deviation***
- ✓ Statistical testing – nearest neighbour, student's t-test, chi-square, Spearman's rank correlation, ***Pearson's product moment correlation coefficient, linear regression***

Skill area: The production and interpretation of maps and diagrams

- ✓ ***Design and layout of maps*** – principles of lettering, linework, shading, dot maps, isoline maps, choropleth maps, proportional circles, divided proportional circles, ***flow maps***
- ✓ ***Interpretation of OS maps and related data***
- ✓ ***Topographical analysis*** – cross-sections, transects, ***river profiles***

A2 techniques

Fieldwork technique	**Information to collect**	**Production/ interpretation of maps**	**Statistical/ graphical techniques**
Measurement of river at different locations.	Width, depth, speed, etc.	Description of stream and its valley from 1:25000 OS map. Construction of long profile from this map. Additional information from 1:50000 solid geology maps.	Does the river become wider, deeper, slower downstream? Line graphs to show this information. Horizontal axis – distance from source (use the horizontal axis from long profile). Vertical axis speed/ depth width.
Measurement of the river's discharge.	Draw cross-sections of river at sampling points. Calculate areas and multiply by river's speed at the different sampling points.	Flow chart to show varying discharge downstream.	Pearson's product moment correlation coefficient. 'There is no relationship between distance from source and discharge of river.' Use the information collected at the sample points for this exercise. Draw a regression line.
Analysis of river's bedload material.	Collect a random sample of at least 10 stones at each of the sampling points. Measure their long, horizontal and vertical axes.		Calculate the mean and standard deviation (SD) stone size for the 10 stones collected at each sample point. It is expected that the stones will become smaller downstream as the river loses its energy and the SD becomes smaller as sorting takes place.

The study of a slope in a rural area

A3 Investigation summary sheet

Skill area: Fieldwork survey/measurement and recording techniques

Physical topics

- ❑ Morphological mapping
- ❑ Vegetation sampling
- ✓ ***Slope analysis***
- ❑ Stream analysis
- ❑ Meteorology in the local setting
- ✓ ***Soil profiles and characteristics***
- ❑ Pebble analysis

Human topics

- ✓ ***Rural land-use mapping***
- ❑ Urban land-use mapping
- ❑ Traffic, pedestrian and environmental quality surveys
- ❑ Questionnaire design and implementation
- ❑ Use of secondary sources
- ❑ Reilly's law of retail gravitation
- ❑ Huff's probability law
- ❑ Nearest neighbour analysis

Skill area: Statistical awareness

- ✓ ***Sampling*** – random, ***systematic***, stratified
- ✓ ***Handling different data types***, ***nominal***, ordinal, interval, ratio
- ❑ Graphical presentation of data – systems diagrams, logarithmic graphs, kite diagrams, scattergraphs, polar graphs, triangular graphs, dispersion diagrams, bipolar analysis
- ❑ Descriptive statistics – mean, median, mode, standard deviation
- ✓ ***Statistical testing*** – nearest neighbour, student's t-test, ***chi-square***, Spearman's rank correlation, Pearson's product moment correlation coefficient, linear regression

Skill area: The production and interpretation of maps and diagrams

- ✓ Design and layout of maps – principles of lettering, linework, shading, dot maps, isoline maps, **choropleth maps**, proportional circles, divided proportional circles, flow maps
- ✓ Interpretation of OS maps and related data
- ✓ ***Topographical analysis*** – ***cross-sections***, ***transects***, river profiles

A3 techniques

Fieldwork technique	**Information to collect**	**Production/ interpretation of maps**	**Statistical/graphical techniques**
Rural land-use survey. Obtain a base map (1:10000) of a gently sloping local rural area.	Walk down the slope making note of the land use observed.	Description of the physical features of the slope from a 1:25000 OS map. Completed land use map using Coleman's land use classification (see fieldwork books for the key).	**Cross-section** of slope. Add land-use information below the cross-section obtained from completed land-use map to **construct a transect.** Use point sampling techniques to calculate the area given over to each land use. Draw a **pie chart** from this information. **Chi-square test** to examine whether there is a relationship between land use and height.
	The examination of solid geology and drift geology maps (latter for information on soils).	Construct solid and drift geology overlays. Add in some important contours.	**Chi-square test** to examine whether there is a relationship between land use and geology. Add solid and drift geology information to transect.
Soil samples taken from various sample points on the slope (use drift geology map to help you).	Lab work on the samples. Dry out for moisture content. Burn off vegetation from dry sample for organic content, and test ph to find if there is a relationship between height, moisture content, organic material and ph.	Drift geology maps give an idea of soils. There are also soil maps available.	Draw another cross-section but at this time have proportional squares at each sample point to show the percentage of moisture and organic material at each sample point on the cross-section. Use 1cm square graph paper for this. You can also illustrate the ph by colouring in a square according to its ph colour.

APPENDICES

Appendix 1: Bibliography

Bowen, Ann and Pallister, John, *Tackling Geography Coursework*, Hodder & Stoughton, 1997

Broadley, E and Cunningham, R, *Core Themes in Geography: Physical*, Oliver & Boyd, 1991

Cole, J P, *Situations in Human Geography*, Blackwell, 1975

Doherty, A and Borradaile, L, *Geographical Methods and Techniques: Statistical Awareness, Advanced Higher*, Scottish CCC, 2000

Geographical Methods and Techniques, HSDU Support Pack (ref. 3826)

Nagle, G, *Skills and Techniques for Geography A Level*, Stanley Thornes, 1998

Nagle, G and Spencer, K, *Geographical Enquiries*, Stanley Thornes, 1997

Speak, P and Carter, A H, *Map Reading and Interpretation*, Longman, 1971

Swan, W and Williams, P, *Open University Statistics Guide*, Open University Press, 1985

Taylor, L, *Geographical Techniques*, Pearson, 1998

Waugh, D, *Geography: An Integrated Approach* (3rd edn), Nelson, 2000

Worthington, B and Gant, R, *Ordnance Survey Mapwork*, Macmillan/OS, 1983

Appendix 2: Statistical tables

Table 1: Critical Values of Student's t

(For TWO-TAILED TESTS)

Degrees of freedom	Significance level 0.05	0.01
1	12.71	63.66
2	4.30	9.93
3	3.18	5.84
4	2.78	4.60
5	2.57	4.03
6	2.45	3.71
7	2.36	3.50
8	2.31	3.36
9	2.26	3.25
10	2.23	3.17
11	2.20	3.11
12	2.18	3.06
13	2.16	3.01
14	2.14	2.98
15	2.13	2.95
16	2.12	2.92
17	2.11	2.90
18	2.10	2.88
19	2.09	2.86
20	2.09	2.85
21	2.08	2.83
22	2.07	2.82
23	2.07	2.81
24	2.06	2.80
25	2.06	2.79
26	2.06	2.78
27	2.05	2.77
28	2.05	2.76
29	2.05	2.76
30	2.04	2.75
40	2.02	2.70
60	2.00	2.66
120	1.98	2.62

To calculate the degrees of freedom where the two sample sizes are n_x and n_y, use:
Degrees of freedom $= (n_x - 1) + (n_y - 1)$

- Select the significance level you have chosen or been given (either 0.05 or 0.01).
- Go down the appropriate column until you get to the row with the correct degrees of freedom.
- Read the critical value from the table.
- For example, for 15 degrees of freedom, critical value = 2.13 (at 0.05 significance level).
- Reject the null hypothesis (NH) if the value of **t** you have calculated is greater than the critical value from the table.

Table 2: Critical Values of Chi-square, χ^2

Degrees of freedom	Significance level 0.05	0.01
1	3.84	6.63
2	5.99	9.21
3	7.81	11.3
4	9.49	13.3
5	11.1	15.1
6	12.6	16.8
7	14.1	18.5
8	15.5	20.1
9	16.9	21.7
10	18.3	23.2
11	19.7	24.7
12	21.0	26.2
13	22.4	27.7
14	23.7	29.1
15	25.0	30.6
16	26.3	32.0
17	27.6	33.4
18	28.9	34.8
19	30.1	36.2
20	31.4	37.6
21	32.7	38.9
22	33.9	40.3
23	35.2	41.6
24	36.4	43.0
25	37.7	44.3
26	38.9	45.6
27	40.1	47.0
28	41.3	48.3
29	42.6	49.6
30	43.8	50.9
40	55.8	63.7
50	67.5	76.2
60	79.1	88.4
70	90.5	100.4
80	101.9	112.3
90	113.1	124.1
100	124.3	135.8

To calculate the degrees of freedom where there are R rows and C columns use:

Degrees of freedom = (R – 1) + (C – 1)

- Select the significance level you have chosen or been given (either 0.05 or 0.01).
- Go down the appropriate column until you get to the row with the correct degrees of freedom.
- Read the critical value from the table.
- For example, for 15 degrees of freedom, critical value = 25.0 (at 0.05 significance level).
- Reject the null hypothesis (NH) if the value of χ^2 you have calculated is greater than the critical value from the table.

[*NOTE that the Chi-square test as it is used in this course is effectively a 2-tailed test*.]

Table 3: Critical Values of Spearman's rank correlation coefficient, r_s

(For TWO-TAILED TESTS)

Degrees of freedom	Significance level 0.05	0.01
5	1.000	-
6	0.886	-
7	0.786	1.000
8	0.738	0.881
9	0.683	0.833
10	0.648	0.794
11	0.623	0.818
12	0.591	0.780
13	0.566	0.745
14	0.545	0.716
15	0.525	0.689
16	0.507	0.666
17	0.490	0.645
18	0.476	0.625
19	0.462	0.608
20	0.450	0.591
21	0.438	0.576
22	0.428	0.562
23	0.418	0.549
24	0.409	0.537
25	0.400	0.526
26	0.392	0.515
27	0.385	0.505
28	0.377	0.496
29	0.370	0.487
30	0.364	0.478

Degrees of freedom = number of paired measurements in the sample.

- Select the significance level you have chosen or been given (either 0.05 or 0.01).
- Go down the appropriate column until you get to the row with the correct degrees of freedom.
- Read the critical value from the table.
- For example, for 15 degrees of freedom, critical value = 0.525 (at 0.05 significance level).
- Reject the null hypothesis (NH) if the value of $\mathbf{r}_s$ you have calculated is greater than the critical value from the table.

Table 4: Critical Values of the Pearson product moment correlation coefficient, r

(For TWO-TAILED TESTS)

Degrees of freedom	**Significance level**	
	0.05	**0.01**
1	0.997	0.9999
2	0.950	0.990
3	0.878	0.959
4	0.811	0.917
5	0.754	0.874
6	0.707	0.834
7	0.666	0.798
8	0.632	0.765
9	0.602	0.735
10	0.576	0.708
11	0.553	0.684
12	0.532	0.661
13	0.514	0.641
14	0.497	0.623
15	0.482	0.606
16	0.468	0.590
17	0.456	0.575
18	0.444	0.561
19	0.433	0.549
20	0.423	0.537
25	0.381	0.487
30	0.349	0.449
35	0.325	0.418
40	0.304	0.393
45	0.288	0.372
50	0.273	0.354
60	0.250	0.325
70	0.232	0.302
80	0.217	0.283
90	0.205	0.267
100	0.195	0.254

Degrees of freedom = n – 2 (where n is the number of paired observations).

- Select the significance level you have chosen or been given (either 0.05 or 0.01).
- Go down the appropriate column until you get to the row with the correct degrees of freedom.
- Read the critical value from the table.
- For example, for 15 degrees of freedom, critical value = 0.482 (at 0.05 significance level).
- Reject the null hypothesis (NH) if the value of **r** you have calculated is greater than the critical value from the table.

Table 5: Random Numbers

10	30	50	35	27	36	37	34	92	09	72	73	16	59	48	10	56	47	19	23	65	30	07	29	02
50	27	78	83	19	76	16	94	11	68	84	26	23	54	20	86	85	23	86	66	99	07	36	37	47
87	87	53	59	81	63	61	65	36	90	18	48	27	45	68	10	19	41	48	01	78	78	85	55	19
27	37	24	96	95	47	96	05	43	91	39	65	36	70	77	45	86	50	51	74	13	39	35	22	54
81	99	78	55	23	84	97	77	72	73	30	53	36	02	95	49	34	88	73	61	94	58	25	58	08
81	57	35	27	33	71	72	24	53	63	94	09	41	10	76	47	91	44	04	95	49	66	39	60	05
49	48	50	86	54	48	50	22	06	34	72	52	82	21	15	65	20	33	29	94	71	11	15	67	59
94	12	03	57	61	96	48	95	25	16	39	33	66	56	10	23	77	21	30	12	90	49	22	23	62
78	77	80	20	75	82	32	99	93	63	89	73	44	99	05	48	67	26	43	18	37	56	09	34	08
38	34	00	78	45	63	98	35	55	03	68	56	08	64	07	34	18	52	56	27	09	24	61	85	45
06	23	13	13	92	72	66	99	37	24	49	57	74	25	43	62	17	10	11	79	99	63	22	98	38
87	14	29	09	04	87	84	58	83	27	29	84	86	50	00	52	35	63	65	23	41	94	21	78	09
37	36	46	18	34	94	20	77	91	16	00	08	43	18	73	68	69	61	34	25	44	67	34	16	55
78	27	62	50	96	72	79	44	61	40	15	14	53	40	65	39	27	31	58	50	28	11	39	00	25
48	01	61	16	96	94	50	78	30	95	44	68	67	67	86	74	76	79	77	48	38	75	93	29	29