1. Maria planned a statistical investigation into trees of a certain variety. She wished to test whether there is positive linear correlation between the height of a tree and the circumference of its trunk at the base.

(a) State, with a reason, whether a 1-tail or a 2-tail test is more appropriate. [1]

Maria recorded the height and circumference of a random sample of 10 trees of this variety in a wood near her home. She calculated the product-moment correlation coefficient for her sample and found that the value was 0.642.

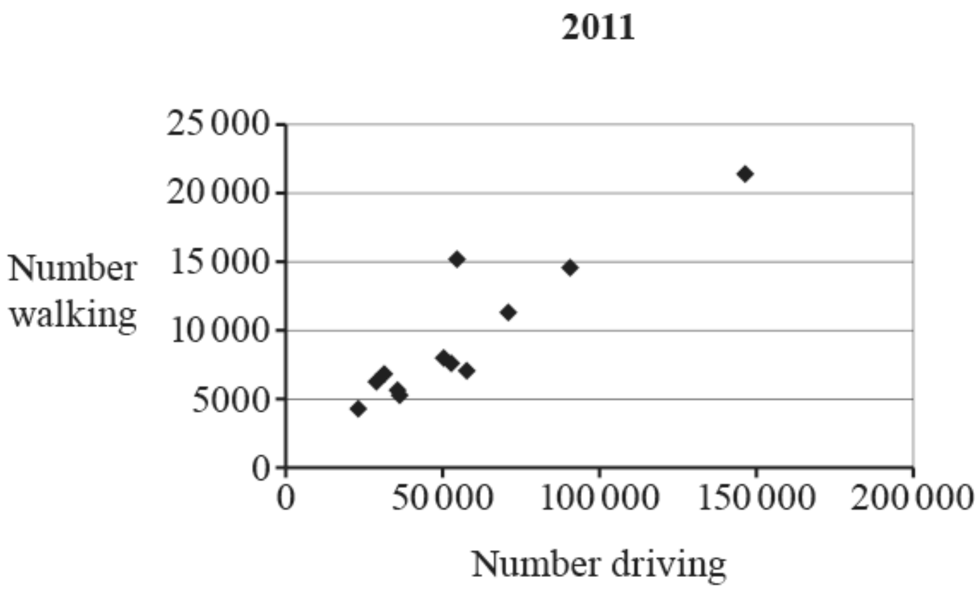
(b) Use the table below to carry out the test at the 2.5% significance level. [5]

(c) Give two reasons why it would not be appropriate to use Maria's results to draw a conclusion about all trees of this variety. [2]

Critical values of Pearson's product-moment correlation coefficient.

| | 1-tail test | 5% | 2.5% | 1% | 0.5% |
|---|---|---|---|---|---|
| | 2-tail test | 10% | 5% | 2.5% | 1% |
| $n$ | 9 | 0.5822 | 0.6664 | 0.7498 | 0.7977 |
| | 10 | 0.5494 | 0.6319 | 0.7155 | 0.7646 |
| | 11 | 0.5214 | 0.6021 | 0.6851 | 0.7348 |
| | 12 | 0.4973 | 0.5760 | 0.6581 | 0.7079 |

2. The scatter diagram shows data, taken from the pre-release data set (see http:/www.ocr.org.uk/Images/308727-units-h230-and-h240-large-data-set-lds-sample-assessment-material.xlsx), for several Local Authorities in one region of the UK in 2011. The diagram shows, for each Local Authority, the number of workers who drove to work, and the number of workers who walked to work.




(a) Four students calculated the value of Pearson's product-moment correlation coefficient for the data in the diagram. Their answers were 0.913, 0.124, –0.913 and –0.124. One of these values is correct. Without calculation state, with a reason, which is the correct value. [2]

(b) Sanjay makes the following statement.

"The diagram shows that, in **any** Local Authority, if there are a large number of people who drive to work there will be a large number who walk to work."

Give a reason why this statement is incorrect. [2]

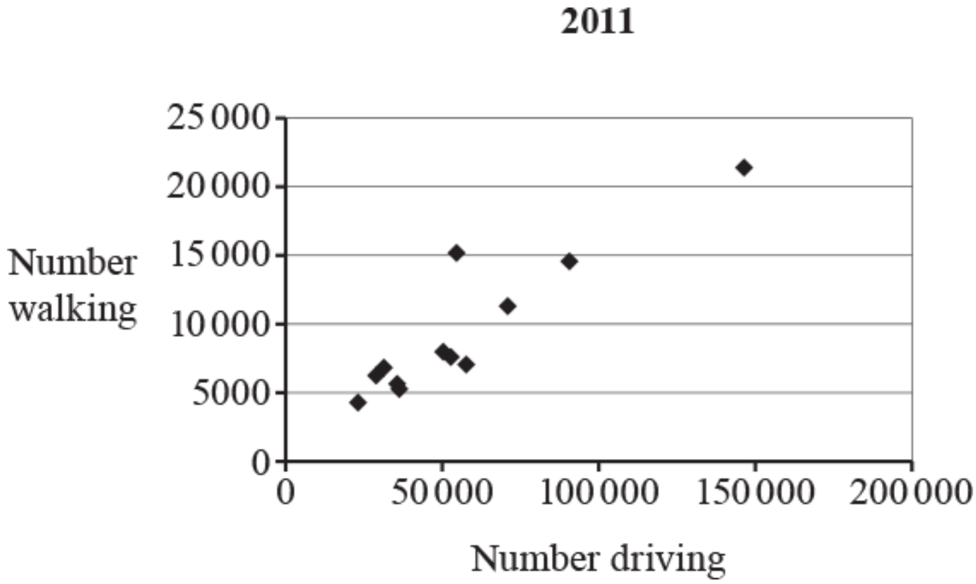
(c) Rosie makes the following statement.

"The diagram must be wrong because it shows good positive correlation. If there are more people driving to work, there will be fewer people walking to work, so there would be negative correlation."

[1]

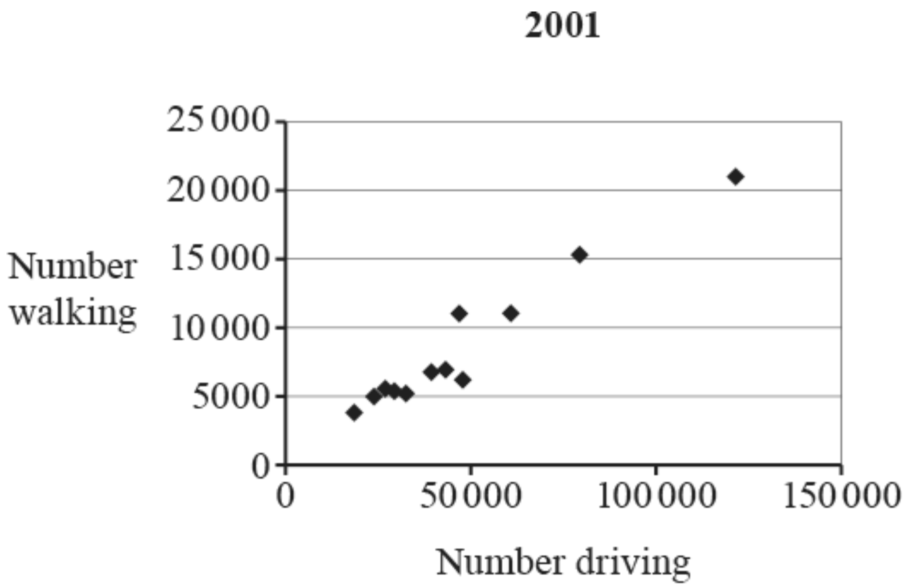
Explain briefly why Rosie's statement is incorrect.

(d) The diagram shows a fairly close relationship between the two variables. One point on the diagram represents a Local Authority where this relationship is less strong than for the others. On the diagram below, label this point A. [1]




(e) Given that the point A represents a metropolitan borough, suggest a reason why the relationship is less strong for this Local Authority than for the others in the region. [1]

The scatter diagram below shows the corresponding data for the same region in 2001.




(f) (i) State a change that has taken place in the metropolitan borough represented by the point A between 2001 and 2011. [1]

(ii) Suggest a possible reason for this change. [1]

3. Christa used Pearson's product-moment correlation coefficient, $r$, to compare the use of public transport with the use of private vehicles for travel to work in the UK.

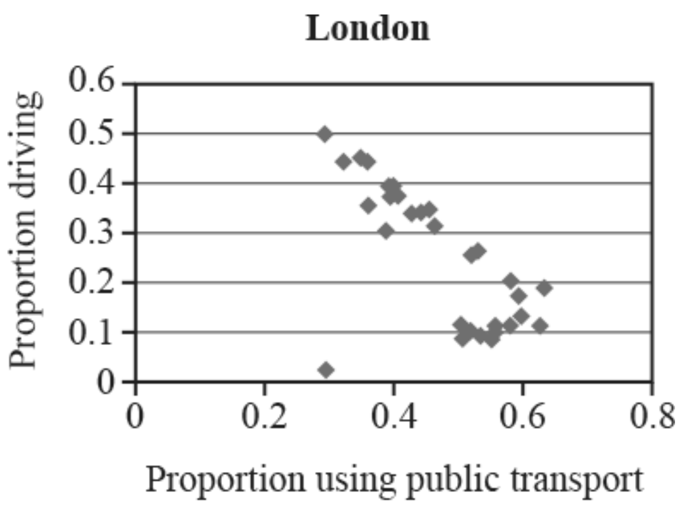
(a) Using the pre-release data set for all 348 UK Local Authorities, she considered the following four variables.

| | |
|---|---|
| Number of employees using public transport | $x$ |
| Number of employees using private vehicles | $y$ |
| Proportion of employees using public transport | $a$ |
| Proportion of employees using private vehicles | $b$ |

(i) Explain, in context, why you would expect strong, positive correlation between $x$ and $y$. [1]

(ii) Explain, in context, what kind of correlation you would expect between $a$ and $b$. [2]

(b) Christa also considered the data for the 33 London boroughs alone and she generated the following scatter diagram.




One London Borough is represented by an outlier in the diagram.

(i) Suggest what effect this outlier is likely to have on the value of $r$ for the 32 London Boroughs. [1]

(ii) Suggest what effect this outlier is likely to have on the value of $r$ for the whole country. [1]

(iii) What can you deduce about the area of the London Borough represented by the outlier? Explain your answer. [1]

4. In an experiment involving a bivariate distribution ($X$, $Y$) a random sample of 7 pairs of values was obtained and Pearson's product-moment correlation coefficient $r$ was calculated for these values.

(a) The value of r was found to be 0.894. Use the table below to test, at the 5% significance level, whether there is positive linear correlation in the population, stating your hypotheses and conclusion clearly. [5]

| 1-tail test | 5% | 2.5% | 1% | 0.5% |
|---|---|---|---|---|
| 2-tail test | 10% | 5% | 2% | 1% |
| $n$ | | | | |
| 1 | – | – | – | – |
| 2 | – | – | – | – |
| 3 | 0.9877 | 0.9969 | 0.9995 | 0.9999 |
| 4 | 0.9000 | 0.9500 | 0.9800 | 0.9900 |
| 5 | 0.8054 | 0.8783 | 0.9343 | 0.9587 |
| 6 | 0.7293 | 0.8114 | 0.8822 | 0.9587 |
| 7 | 0.6694 | 0.7545 | 0.8329 | 0.9745 |
| 8 | 0.6215 | 0.7067 | 0.7887 | 0.8343 |
| 9 | 0.5882 | 0.6664 | 0.7498 | 0.7977 |
| 10 | 0.5494 | 0.6319 | 0.7155 | 0.7646 |

Scatter diagrams for four sets of bivariate data, are shown.

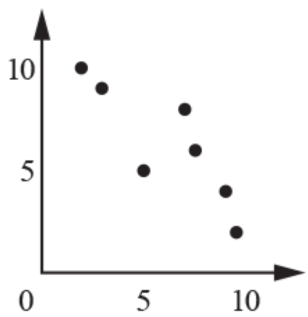


Diagram A

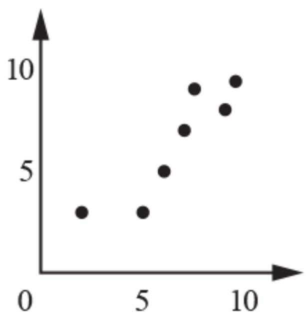


Diagram B

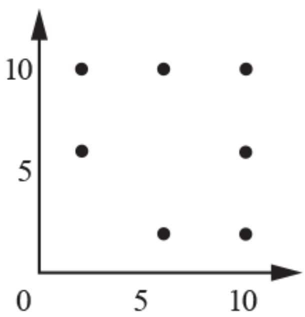


Diagram C

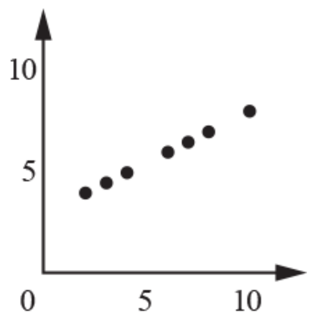


Diagram D

It is given that $r = 0.894$ for **one** of these diagrams.

(b) For each of the **other** diagrams, state how you can tell that $r \neq 0.894$. [3]

5. Laxmi wishes to test whether there is linear correlation between the mass and the height of adult males.

[1]

(a) State, with a reason, whether Laxmi should use a 1-tail or a 2-tail test.

Laxmi chooses a random sample of 40 adult males and calculates Pearson's product-moment correlation coefficient, $r$. She finds that $r = 0.2705$.

(b) Use the table below to carry out the test at the 5% significance level. [5]

Critical values of Pearson's product-moment correlation coefficient.

| | 1-tail test | 5% | 2.5% | 1% | 0.5% |
|---|---|---|---|---|---|
| | 2-tail test | 10% | 5% | 2.5% | 1% |
| $n$ | 38 | 0.2709 | 0.3202 | 0.3760 | 0.4128 |
| | 39 | 0.2673 | 0.3160 | 0.3712 | 0.4076 |
| | 40 | 0.2638 | 0.3120 | 0.3665 | 0.4026 |
| | 41 | 0.2605 | 0.3081 | 0.3621 | 0.3978 |

END OF QUESTION paper

# Mark scheme

| Question | | | Answer/Indicative content | Marks | Guidance | |
|---|---|---|---|---|---|---|
| 1 | | a | 1-tail. Testing for "positive" linear correlation | E1(AO 3.1b)<br>[1] | or Expect larger circumference to go with larger height oe | |
| | | b | $H_0$: $\rho = 0$<br>$H_1$: $\rho > 0$ where $\rho$ is lin correlation coeff in pop<br>Comp 0.642 with 0.6319<br>Reject $H_0$.<br>There is evidence of +ve (linear) corr'n between height & circ of trees of this variety (in this wood) | B1(AO1.1)<br>B1(AO2.5)<br>M1(AO1.1)<br>M1(AO1.1)<br>A1(AO2.2b)<br>[5] | B1B0 for 1 error, eg undefined $\rho$ or 2-tail<br>In context, not definite. | Allow omission of "linear" throughout<br>Allow without "linear" and / or "in this wood" |
| | | c | eg sample is small<br>conditions in other areas may be different | E1(AO 3.5a)<br>E1(AO 3.5a)<br>[2] | | |
| | | | Total | 8 | | |
| 2 | | a | Points close to straight line with +ve gradient<br>Hence 0.913 is the correct value | B1(AO 2.4)<br>B1(AO 2.2b)<br>[2] | Dep on 1st B1 | |

| | | | | | |
|---|---|---|---|---|---|
| | | b | Sample is from one area, hence not random oe | E2(AO 2.3 2.2b) [2] | Or might be different relationship elsewhere / Allow E1 for "Sample is small" |
| | | c | Both depend on the size (of the pop) of area. | E1(AO 2.4) [1] | |
| | | d | Correct point indicated (54200, 15300) | B1(AO 3.2b) [1] | |
| | | e | More local jobs (so higher proportion walk) | E1(AO 2.2b) [1] | Any sensible equivalent, |
| | | f | (i) Eg Fewer walk to work<br>(ii) Eg Some businesses within the borough have closed down or have moved to the outskirts | E1(AO 2.2b) [1]<br>E1(AO 2.2b) [1] | or any relevant comment |
| | | | Total | 9 | |
| 3 | | a | (i) Both the number of employees using public transport and the number of employees using private vehicles depend on the LA population.<br>(ii) Negative<br>If a large prop use public transport then a smaller | E1 (AO 2.1) [1]<br>E1ind (AO 2.2b) | or similar, but must be in context. Ignore all else / NOT No. using pt is prop to no. using pv<br>**Examiner's Comments**<br>Most candidates answered correctly, showing a good understanding of the difference between this part and part (a)(ii). A few, however, wrote that as the number of employees using public transport increases, the number using private will decrease.<br>Ignore "strong" or "slight" etc / NOT Inverse prop'n |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | prop drive (and vice versa) | E1ind (AO 2.4) [2] | or similar in context; NOT "as a increases $b$ decreases" unless in context<br><br>**Examiner's Comments**<br><br>Many candidates understood the point, although some of these worded their answers badly, referring to the "numbers" (rather than "proportions") using the two types of transport, without making it clear that they were discussing each individual LA rather than all LAs together. |
| | | b | (i) | Decrease the size of $r$ or Make $r$ less negative | E1 (AO 2.2b) [1] | Make (value of ) $r$ increase $r$ closer to 0; Ignore eg "greatly"; Ignore all else; NOT Make $r$ decrease; NOT Weaken the corr'n; NOT Make corr'n less<br><br>**Examiner's Comments**<br><br>Some candidates stated that $r$ would decrease, or that the value of $r$ would decrease, both of which are incorrect. It is possible that what they meant was that the size of $r$ would decrease, which is correct, but unfortunately these candidates could not be credited the mark. Some candidates ensured that there was no ambiguity by saying that $r$ would "become less negative" or "move closer to 0" or "decrease in magnitude". Some candidates gave inadequate answers such as "It will weaken the correlation" or "It will weaken the value of $r$". There were a few irrelevant answers such as "The outlier will skew the distribution." |
| | | | (ii) | Little effect (because the population of the LA is small compared with the whole population) | E1 (AO 2.2b) [1] | or No effect or similar; Ignore all else<br><br>**Examiner's Comments**<br><br>Many good answers were seen. |
| | | | (iii) | Ignore all reference to public transport<br>Type 1 answers<br>People don't travel far to work | E1 (AO 2.4) | Type 2 answers<br>Any suggested |

| | | | | | |
|---|---|---|---|---|---|
| | | | Jobs are close<br>High proportion walk (or cycle) | [1] | reason why few drive<br>eg Few garages;<br>Parking expensive<br>or similar in context — NOT just Few drive<br><br>**Examiner's Comments**<br><br>Most candidates recognised the key factor - that a tiny proportion drive to work. But some candidates mistakenly suggested that this is because there is a great deal of public transport available. Others merely stated that few people drive to work. This was not considered an adequate answer to the question. To gain the mark answers had to fall into one of two categories:<br><br>1. A sensible suggestion for a possible reason why in this particular area few people drive.<br><br>2. A statement that it is likely that a large proportion walk or cycle to work, or that jobs are generally close to home. |
| | | | Total | 6 | |
| 4 | | a | $H_0$: There is no linear correlation between $X$ and $Y$<br><br>$H_1$: There is positive linear correlation between $X$ and $Y$<br><br>Compare with 0.6694<br><br>Reject $H_0$<br><br>There is evidence of positive linear correlation between $X$ & $Y$ | B1 (AO1.1)<br>B1 (AO2.5)<br>B1 (AO1.1)<br>M1 (AO1.1)<br>A1 (AO2.2b)<br>[2] | B1B0 for one error,<br><br>eg omission of "linear" OR "+ve"<br><br>Or $\rho = 0$<br><br>Or $\rho > 0$<br><br>In context, not definite |
| | | b | A: Negative (linear) correlation<br><br>C: No (linear) correlation<br><br>D: $r = 1$ | E1 (AO1.2)<br>E1 (AO2.2a)<br><br>E1 (AO1.2)<br>[3] | Allow without "linear"<br><br>or points not close to straight line, or $r = 0$ — Allow without "linear" |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | Not “r small” or “poor correlation |
| | | | Total | 8 | |
| 5 | | a | 1-tail, because expect mass and height to be positively correlated | B1 (AO 1.1)<br>[1] | 1-tail because generally expect taller people to be heavier \| or equivalent |
| | | b | $H_0$: $\rho = 0$<br>$H_1$: $\rho > 0$ where $\rho$ is pmcc for population<br>comp 0.2638<br>Reject $H_0$<br>There is evidence that the mass and height of adult males are positively correlated | B1 (AO 1.1)<br>B1 (AO 2.5)<br>B1 (AO 1.1)<br>M1 (AO 2.2b)<br>A1 (AO 3.5a)<br>[5] | Omit definition of $\rho$: B1B0 |
| | | | Total | 6 | |