

## Planning Investigations Mark Scheme

Q1.

Question	Answer	Additional guidance	Mark
	<p>B1 for each of five correct comments e.g.</p> <ul style="list-style-type: none"><li>• the hypothesis should be a statement not a question</li><li>• time intervals are not exhaustive (e.g. there is no option for 0 or 7+)</li><li>• time intervals need to be longer as many students are likely to spend more than 6 hours a week</li><li>• a scatter graph is not appropriate (as horizontal axis is not numerical data)</li><li>• it might be difficult for students to know the times as they might go on social media multiple times a day for short periods</li><li>• it is not necessary to know the name of the student</li><li>• he has not planned to collect the relevant information for a time series graph</li><li>• calculating estimates of the means is sensible</li><li>• finding the gender of the student is relevant</li></ul>	<p>B1 for each correct comment on the appropriateness of the hypothesis or plans for collecting, processing and presenting the data</p>	<p>(5)</p>

**Q2.**

Question number	Answer	Additional guidance	Mark
	<p><b>Collecting data</b>            B1 for identifying one appropriate thing that should be included in the plan for collecting data <b>and</b>            B1 for explaining why this aspect is appropriate            OR            B1 for deciding what data to collect and/or how to collect and record it <b>and</b>            B1 for an appropriate reason            OR            B1 for a strategy to process data <b>and</b>            B1 for an appropriate reason            OR            B1 for designing a collection method for primary/secondary data <b>and</b>            B1 for an appropriate reason            OR            B1 for appreciating the importance of acknowledging sources <b>and</b>            B1 for an appropriate reason            OR            B1 for recognising where issues of sensitivity may influence data availability <b>and</b>            B1 for an appropriate reason</p> <p><b>Processing and presenting</b>            B1 for planning to organise and/or process data <b>and</b>            B1 for an appropriate reason            OR</p>	<p>B1B1B1 for each of three planned elements and B1B1B1 for each of three appropriate reasons from their three things in the statistical enquiry cycle. Maximum 4 marks if only one aspect (from Collecting data, Processing and presenting, Interpreting and Evaluating) is referenced.            B1 for e.g. use amount of time measured to the nearest minute <b>and</b>            B1 for e.g. this is sufficient as there will be a large range of times</p> <p>B1 for e.g. collect data for 23 boys and 23 girls <b>and</b>            B1 for e.g. this will make the calculation of quartiles easier            B1 for e.g. use random sampling <b>and</b>            B1 for e.g. this reduces bias as Gary's friends/class/peers may generally watch the same programs</p> <p>B1 for e.g. use primary data <b>and</b>            B1 for e.g. this increases reliability as Gary will know how the data was collected</p> <p>B1 for e.g. A student (Gary) should collect the data <b>and</b>            B1 for e.g. students are more likely to give an honest answer to a fellow student (less threatening)</p> <p>B1 for e.g. use a grouped frequencies table for the data <b>and</b>            B1 for e.g. as this will enable a quick way of estimating the mean or this can be used to draw a histogram</p>	<p>(6)</p>

Question number	Answer	Additional guidance	Mark
<i>continued</i>	<p>B1 for planning to generate diagrams and/or visualisations to represent the data <b>and</b>            B1 for an appropriate reason            OR            B1 for planning to generate statistical measures to compare data <b>and</b>            B1 for an appropriate reason</p> <p><b>Interpreting</b>            B1 for planning to interpret diagrams and/or calculations/measures <b>and</b>            B1 for an appropriate reason</p> <p>OR            B1 for planning to make an inference and/or prediction <b>and</b>            B1 for an appropriate reason</p> <p><b>Evaluating</b>            B1 for planning to identify weaknesses in approach or representation <b>and</b>            B1 for an appropriate reason            OR            B1 for planning to refine the processes to elicit further clarification of the hypothesis <b>and</b>            B1 for an appropriate reason</p>	<p>B1 for e.g. use box plots <b>and</b>            B1 for e.g. these will enable the comparison of both the medians and the IQRs (i.e. the distributions) of the data</p> <p>B1 for e.g. interpret results for each individual school year <b>and</b>            B1 for e.g. as different years could have different watching habits</p> <p>B1 for e.g. by comparing means <b>and</b>            B1 for e.g. you can see whether the amount of time that boys spend watching TV is greater, in general, than the amount of time that girls spend watching TV</p> <p>B1 for e.g. use the results from the school to predict the results nationally <b>and</b>            B1 for e.g. as students in different parts of the country are likely to have the same watching habits</p> <p>B1 for e.g. choose not to display the information in histograms <b>and</b>            B1 for e.g. as “the target audience” may not know how to interpret them</p> <p>B1 for e.g. consider using more than one type of visual representation (for the same information) <b>and</b>            B1 for e.g. as different representations focus on different aspects of the data</p>	

**Q3.**

Question	Answer	Additional guidance	Mark
	<p><b>Data collection</b>            B1 for identifying one appropriate thing that should be included in the plan for data collection  <b>and</b>            B1 for explaining why this aspect is appropriate</p> <p>B1 for deciding what data to collect and/or how to collect and record it <b>and</b>            B1 for an appropriate reason</p> <p>OR</p> <p>B1 for designing a collection method for primary/secondary data <b>and</b>            B1 for an appropriate reason</p> <p>OR</p> <p>B1 for recognising where issues of sensitivity may influence data availability            B1 for an appropriate reason</p> <p><b>Presenting and interpreting diagrams</b>            B1 for identifying an appropriate method for presenting data in a diagram  <b>and</b>            B1 for explaining how this will be interpreted in the context of the investigation</p>	<p>B1 for e.g. use distance to holiday to the nearest mile <b>and</b>            B1 for e.g. this is sufficient as there will be a large range of distances</p> <p>B1 for e.g. use quota sampling <b>and</b>            B1 for e.g. this ensures that you have data for each price range</p> <p>OR</p> <p>B1 for e.g. collecting data for prices of only one type of holiday, for example all inclusive <b>and</b>            B1 for e.g. this means that the prices aren't affected by factors other than distance</p> <p>B1 for e.g. find prices from a holiday company website rather than asking people how much they paid for their holiday <b>and</b>            B1 for e.g. as people may not want to say how much they spent on their holiday</p> <p>B1 for e.g. use a scatter diagram for the data <b>and</b>            B1 for e.g. as this will allow you to see if there is a correlation (between the distance travelled and the price of the holiday)</p>	<p>(6)</p>
	<p><b>Presenting and interpreting calculations</b>            B1 for identifying an appropriate method to generate statistical measures to compare data <b>and</b>            B1 for explaining how this will be interpreted in the context of the investigation</p>	<p>B1 for e.g. calculate the mean point to plot an accurate line of best fit <b>and</b>            B1 for e.g. as you could use the line of best fit to make predictions about the cost of holidays of a certain distance</p> <p>OR</p> <p>B1 for e.g. calculate the equation of the line of best fit <b>and</b>            B1 for e.g. to be able to see the relationship between distance travelled and price</p> <p>OR</p> <p>B1 for e.g. use a statistical package to calculate Spearman's rank correlation coefficient <b>and</b>            B1 for e.g. to measure the strength of any correlation between distance travelled and price</p>	

**Q4.**

Question	Answer	Additional guidance	Mark
(a)	B1 9.8		(1)
(b)	<p>B1B1 for considering the data eg</p> <ul style="list-style-type: none"> <li>• 'small data set'</li> <li>• 'data set may not be representative (of all adult cats)'</li> <li>• 'not a random sample'</li> </ul> <p>B2 for naming an appropriate average with a supporting reason eg</p> <ul style="list-style-type: none"> <li>• 'she should use the mean as it uses all the data'</li> <li>• 'she should use the median as it is less affected by extreme values'</li> </ul> <p>B1 for considering the effect of the outlier eg</p> <ul style="list-style-type: none"> <li>• 'outliers affect the mean'</li> <li>• 'outlier will have less of an effect the median'</li> <li>• 'including the outlier may lead to an overestimate'</li> <li>• 'it would be appropriate to include the outlier as it may be a real data value'</li> </ul>	<p>B1 for each correct consideration of the data (up to B2)</p> <p>B2 for correct of average and supporting reason (B1 for naming an appropriate average and an attempt at a reason)</p> <p>B1 for an appropriate comment about the effect of the outlier</p>	(5)

**Q5.**

Question	Answer	Additional guidance	Mark
	<p>B1B1B1 for planning for each of 3 uses of technology Rebecca can use technology to <b>Plan for Processing and Presenting</b></p> <ul style="list-style-type: none"> <li>• e.g. sort/order data columns by age/number of appointments</li> <li>• e.g. generate random numbers</li> <li>• e.g. remove extraneous data (phone number/address) columns</li> <li>• e.g. identify missing data</li> <li>• e.g. remove extraneous symbols</li> <li>• e.g. automate the calculation of summary statistics</li> <li>• e.g. automate the production of visual representations</li> <li>• e.g. remove outliers</li> </ul> <p>B1B1B1 for justifying why it is appropriate/advantageous <b>Advantages of using technology</b></p> <ul style="list-style-type: none"> <li>• e.g. saves time</li> <li>• e.g. can reduce errors (in lengthy calculations)</li> <li>• e.g. facilitates use of all data (unbiased)</li> <li>• e.g. easier to correct a mistake</li> <li>• e.g. more visually appealing</li> </ul>	<p>B1B1B1 for each use of technology</p> <p>B1B1B1 for each advantage over processing data by hand</p> <p>Ignore extraneous non-contradictory comments</p>	(6)

**Q6.**

Question	Answer	Additional guidance	Mark
(a)	<p>B1 Scatter (diagram) is most suitable graph</p> <p>B1 ... as data is bivariate/paired</p> <p>B1 line of best fit drawn (through the (double) mean point)</p> <p>B1 gradient (of line of best fit) is 0.8 / will tell her the relative rates of change in gross pay</p> <p>B1 points will be close to the line of best fit, OR there is strong correlation (so evidence is strong)</p>	<p>B1 for correct choice of graph</p> <p>B1 for reason explaining appropriateness for graph choice. Allow equivalent reference to bivariate data eg she was expecting to see correlation</p> <p>B1 for recognising that the double mean point should be used in drawing a best fit line</p> <p>B1 for correctly recognising that statistical reasoning for the comparison is based on gradient. Accept equivalent, eg steepness</p> <p>B1 for recognising how the high value of pmcc supports the validity of the conclusion OR that points will be close to a straight line</p>	(5)
(b)	<p>B1 Correct reason from</p> <ul style="list-style-type: none"> <li>• Old data / out of date</li> <li>• Small sample</li> <li>• Secondary data so don't know how collected</li> <li>• May be affected by (varying) exchange rates</li> </ul>	<p>B1 for an equivalent reason for conclusion to be inappropriate.</p> <p>Ignore excess statements if not contradictory.</p>	(1)

**Q7.**

Question number	Answer	Additional guidance	Mark
	<p>B1 for each of five correct comments eg</p> <ul style="list-style-type: none"> <li>• the hypothesis should be a statement not a question</li> <li>• pre-test a good idea as it can identify issues/problems with the reaction time test</li> <li>• people might do better on the caffeine test because they have done the test before – biased</li> <li>• need to give more detail on test protocol e.g. how much coffee to drink, how long to wait after drinking coffee before doing second test</li> <li>• need to plan to control extraneous variables e.g. what time of day</li> <li>• people may not record data in spreadsheet accurately / may lie about their time</li> <li>• median/interquartile range are suitable measures to compare (as data may be skewed/have outliers)/needed to draw a box plot</li> <li>• good idea to consider outliers/may not be appropriate to remove outliers as they could be genuine data</li> <li>• box plots are suitable to e.g. show the distribution of the times</li> <li>• need to draw comparative box plots and not just one for after the coffee has been drunk</li> </ul>	<p>B1 for each correct comment on the appropriateness of the hypothesis or plans for collecting, processing and presenting the data</p>	(5)

**Q8.**

Question number	Answer	Additional guidance	Mark
(a)	<p>B1 B1 for any two from</p> <p>Appropriate since...</p> <ul style="list-style-type: none"> <li>• they do not overlap</li> <li>• they are exhaustive (since no earthworm is greater than 36 cm)</li> <li>• unequal class intervals are suitable if data is closely grouped in one section of the range of values and more spread out in others</li> <li>• as normally distributed data would be more closely clustered in the middle of the range of values</li> </ul> <p>Not appropriate since...</p> <ul style="list-style-type: none"> <li>• it would be better to have larger class intervals towards 0 and 40 and smaller class intervals in the middle of the data range</li> <li>• it would be better to collect the data before deciding on the class intervals to use</li> </ul>	<p>B1 for each assessment (maximum 2) of the appropriateness of Kate's class intervals</p>	(2)
(b)	<p>B1 B1 B1 B1 for any four from</p> <ul style="list-style-type: none"> <li>• Histogram will allow you to see if the distribution is bell-shaped</li> <li>• Histogram requires continuous data</li> <li>• Calculation of frequency density is incorrect as class width should be <math>10.5 - 5.5 (= 5)</math></li> <li>• 68% of the data should lie within 1 standard deviation of the mean</li> <li>• 95% of the data should lie within 2 standard deviations of the mean</li> </ul>	<p>B1 for each assessment (maximum 4) of the appropriateness of using a histogram and calculating means and standard deviations</p> <p>Condone comments which refer to additional calculations that Bien could include to determine whether the data was normally distributed:</p> <ul style="list-style-type: none"> <li>• Finding the median would also be appropriate</li> <li>• Calculating the skew would be appropriate</li> </ul>	(4)

**Q9.**

Question number	Answer	Additional guidance	Mark
(a)	B1 A list of all the students at his school	<p>B1 for a suitable sampling frame</p> <p>Must have list or suitable alternative e.g. register/database and include the whole population</p> <p>Allow a list of students who have school meals</p>	(1)
(b)	B1 Quota (sampling)		(1)

(c)	<p>B1 B1 B1 B1 B1 for any 5 comments from</p> <p>A. Sampling method:</p> <ul style="list-style-type: none"> <li>• 70 students is a good sample size/appropriate sampling from every year group</li> <li>• The sampling method may generate more of one gender than another/take equal number of boys and girls from each year group</li> <li>• Sampling method is not random/year groups may be different sizes</li> <li>• Doesn't state how he will take his sample</li> </ul> <p>B. Question:</p> <ul style="list-style-type: none"> <li>• Year group will not necessarily give the age</li> <li>• Favourite meal is not relevant to either hypothesis</li> <li>• Most of the questions are closed which will be easier to analyse/Question 3 is an open question</li> <li>• Question 4 is a leading question</li> <li>• Question 5 is not exhaustive/there is no option for 0</li> <li>• Question 5 does not have a time frame</li> </ul> <p>C. Presenting data:</p> <ul style="list-style-type: none"> <li>• There may not be any vegetarians in the sample so a graph could not be drawn/ he hasn't asked if the students are vegetarian to draw a pictogram of the results</li> <li>• A pictogram is appropriate</li> <li>• He can't plot age on the axis as he only has year group</li> </ul>	<p>B1 for each correct comment (maximum 5) on the appropriateness of the plans. With at least one comment from each of sections A, B and C (with a maximum of 3 marks for comments from any one section)</p>	(5)
(d)	<p>B1 Advantage e.g. there are less groups to work with so data easier to handle/it's quicker to handle the data</p> <p>B1 Disadvantage e.g. less accurate</p>	<p>B1 for a suitable advantage of larger class intervals</p> <p>B1 for a suitable disadvantage of larger class intervals</p>	(2)