General Mathematics 2019 v1.2

IA1 high-level annotated sample response

April 2022

Problem-solving and modelling task (20%)

This sample has been compiled by the QCAA to assist and support teachers to match evidence in student responses to the characteristics described in the instrument-specific marking guide (ISMG).

This resource contains content that may be of a sensitive nature for students. Teachers should consult with school leaders and consider the suitability of the investigation in the school's context.

Assessment objectives

This assessment instrument is used to determine student achievement in the following objectives:

1. select, recall and use facts, rules, definitions and procedures drawn from Unit 3 Topics 1, 2 and/or 3
2. comprehend mathematical concepts and techniques drawn from Unit 3 Topics 1, 2 and/or 3
3. communicate using mathematical, statistical and everyday language and conventions
4. evaluate the reasonableness of solutions
5. justify procedures and decisions by explaining mathematical reasoning
6. solve problems by applying mathematical concepts and techniques drawn from Unit 3 Topics 1, 2 and/or 3.

Queensland Government
Queensland Curriculum & Assessment Authority
220251

Instrument-specific marking guide (ISMG)

Criterion: Formulate

Assessment objectives

1. select, recall and use facts, rules definitions and procedures drawn from Unit 3 Topics 1, 2 and/or 3
2. comprehend mathematical concepts and techniques drawn from Unit 3 Topics 1, 2 and/or 3
5. justify procedures and decisions by explaining mathematical reasoning

The student work has the following characteristics:	Marks
• documentation of appropriate assumptions • accurate documentation of relevant observations • accurate translation of all aspects of the problem by identifying mathematical concepts and techniques.	3–4
• statement of some assumptions • statement of some observations • translation of simple aspects of the problem by identifying mathematical concepts and techniques.	1–2
• does not satisfy any of the descriptors above.	0

Criterion: Solve

Assessment objectives

1. select, recall and use facts, rules, definitions and procedures drawn from Unit 3 Topics 1, 2 and/or 3
6. solve problems by applying mathematical concepts and techniques drawn from Unit 3 Topics 1, 2 and/or 3

The student work has the following characteristics:	Marks
• accurate use of complex procedures to reach a valid solution • discerning application of mathematical concepts and techniques relevant to the task • accurate and appropriate use of technology.	6–7
• use of complex procedures to reach a reasonable solution • application of mathematical concepts and techniques relevant to the task • use of technology.	4–5
• use of simple procedures to make some progress towards a solution • simplistic application of mathematical concepts and techniques relevant to the task • superficial use of technology.	2–3
• inappropriate use of technology or procedures.	1
• does not satisfy any of the descriptors above.	0

Criterion: Evaluate and verify

Assessment objectives

4. evaluate the reasonableness of solutions
5. justify procedures and decisions by explaining mathematical reasoning

The student work has the following characteristics:	Marks
• evaluation of the reasonableness of solutions by considering the results, assumptions and observations • documentation of relevant strengths and limitations of the solution and/or model • justification of decisions made using mathematical reasoning.	4–5
• statements about the reasonableness of solutions by considering the context of the task • statements of relevant strengths and limitations of the solution and/or model • statements about decisions made relevant to the context of the task.	2–3
• statement about a decision and/or the reasonableness of a solution.	1
• does not satisfy any of the descriptors above.	0

Criterion: Communicate

Assessment objective

3. communicate using mathematical, statistical and everyday language and conventions

The student work has the following characteristics:	Marks
• correct use of appropriate technical vocabulary, procedural vocabulary and conventions to develop the response • coherent and concise organisation of the response, appropriate to the genre, including a suitable introduction, body and conclusion, which can be read independently of the task sheet.	3–4
• use of some appropriate language and conventions to develop the response • adequate organisation of the response.	1–2
• does not satisfy any of the descriptors above.	0

Task

Investigate the phenomenon of ancestral heredity by focusing on the height of a parent and their biological child of the same sex, using data from students at your school.

The investigation should explore the dependence of a male's height on his father's height, or the dependence of a female's height on her mother's height.

Can a person's height be reliably predicted from their relative's height?

To complete this task, you must:

- respond with a range of understanding and skills, such as using mathematical language, appropriate calculations, tables of data, graphs and diagrams
- provide a response to the context that highlights the real-life application of mathematics
- respond using a written report format that can be read and interpreted independently of the problem-solving and modelling task sheet
- develop a unique response
- use both analytic procedures and technology.

See IA1 sample assessment instrument: Problem-solving and modelling task (20%) (available on the QCAA Portal).

Sample response

Criterion	Marks allocated	Provisional marks
Formulate Assessment objective/s 1, 2, 3	4	4
Solve Assessment objective/s 1, 6	7	6
Evaluate and verify Assessment objective/s 4, 5	5	5
Communicate Assessment objective/s 3	4	4
Total	**20**	**19**

The annotations show the match to the instrument-specific marking guide (ISMG) performance-level descriptors.

Table of contents

1 Introduction

Communicate [3–4]

coherent and concise organisation of the response, appropriate genre, including a suitable introduction

Sir Francis Galton (1822–1911) created the statistical concept of correlation. In 1903, assisted by Alice Lee, Pearson decided to supplement Galton's study on the inheritance of physical characteristics.

The results presented in this report are a product of the investigation into the ancestral heredity of stature with a focus on a son's height compared to their father's height. A sample of 100 male students from the current Year 12 cohort was selected for the study.

2 Considerations

2.1 Observations and assumptions

Formulate [3–4]

accurate documentation of relevant observations

While fathers and sons represent a very large population, a sample of Year 12 peers will be used to represent the sons. Year 12 boys will be sampled because most boys reach their full height by the age of 16.

A sample of 100 was randomly selected from various form classes as an appropriate size to investigate the stature of sons and their fathers. Since the Year 12 cohort has 213 male students, the decision to include approximately half of the cohort was deemed sufficient to allow for a variety of height data. Therefore, a valid comparison between father and son heights could be made. As the father's height will be used to predict the son's height in this investigation, it will be the explanatory variable (x cm) and the son's height will be the response variable (y cm).

Formulate [3–4]

documentation of appropriate assumptions and relevant observations

Additional assumptions and observations have been formulated:

- I saw that all the Year 12 heights were measured accurately and recorded correctly.
- I assume that students reported their father's height accurately. From discussions with each student, I am confident this is a reasonable assumption.
- According to *Medical News Today* (www.medicalnewstoday.com/articles/320676) most boys reach their full height by the age of 16. Therefore, I have assumed that all boys are at their full height for this investigation.
- I have assumed that the dataset represents a typical male population with respect to height variations because the data is generally around the average Australian male height of 175.6 cm (www.australian-population.com/what-is-the-average-australian-male-height).
- It was observed that many tall fathers had tall sons, and many short fathers had short sons. Therefore, it is assumed that there is a linear relationship between the heights of a son and his father.

2.2 Mathematical concepts and techniques

To investigate the phenomenon of the ancestral heredity of height, the following procedures were undertaken.

Student height was measured in class time using a measuring tape, with no shoes on and backs against a wall.

Father height data was collected independently by individual students.

Formulate [3–4]

accurate translation of problem by identifying mathematical concepts and techniques

Height data for father and son pairs was de-identified and aggregated by teachers, and a different random sample was provided to each student.

The heights of fathers and sons were graphed against each other to see if an association was apparent.

Once a linear association could be seen from the scatterplot of fathers' heights and sons' heights, a regression equation was developed using both the formulas from the formula sheet and spreadsheeting functions.

Using the spreadsheet trendline function, the regression line and the coefficient of determination for the heights were displayed on the scatterplot.

To evaluate if a linear model was appropriate to predict the sons' heights from the fathers' heights, residual analysis was used.

The residual values were calculated and plotted against father heights to observe the scatter pattern.

Formulate [3–4] accurate translation of problem by identifying mathematical concepts and techniques **Communicate [3–4]** coherent and concise organisation of the response, appropriate genre, including a suitable body **Solve [6–7]** accurate and appropriate use of technology **Formulate [3-4]** documentation of relevant observations **Solve [6–7]** accurate use of complex procedures to reach a valid solution, application of mathematical concepts and techniques relevant to the task	See main text below

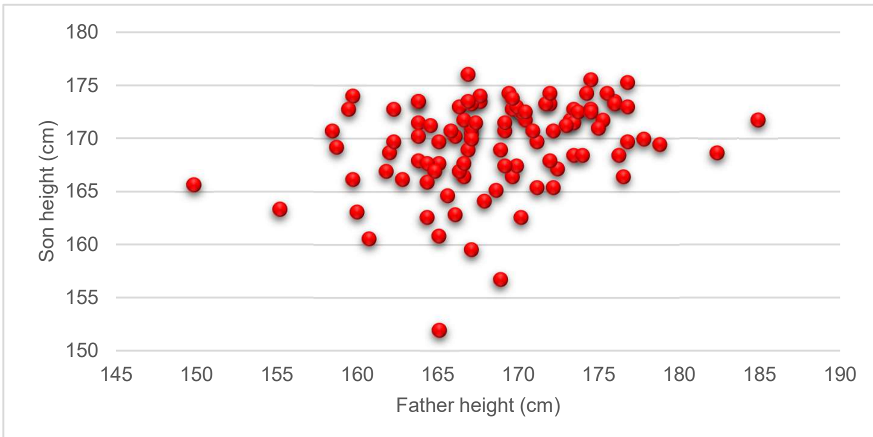
2.3 Use of technology

A spreadsheet program was used extensively during the investigation process to organise the father and son data, prepare graphs, confirm the regression equation and coefficient of determination, and calculate residuals. The program was also used to calculate the statistical measures of mean, standard deviation and the correlation coefficient, which are required to develop the least-squares regression equation analytically.

3 Developing a solution

The height data appears in Appendix 1. The data is presented below in Graph 1, using a scatterplot to identify a possible association between father and son heights.

Graph 1: Scatterplot of son height against father height



On first inspection there did not appear to be a strong association between the two variables. However, a weak positive linear relationship was identified as plausible. Based on this conclusion, a linear regression equation was developed using the least-squares method of regression.

The general form of the least-square regression line is given by:

$y = a + bx$

where $b = r \times \frac{s_y}{s_x}$ and $a = \bar{y} - b\bar{x}$, given r is Pearson's correlation coefficient,
s_x and s_y are the sample standard deviations, and $\bar{x}$ and $\bar{y}$ are the sample means.

As determined using the spreadsheet function CORREL: $r = 0.301815739$.

As determined using the spreadsheet function AVERAGE: $\bar{x} = 168.67$ and $\bar{y} = 169.41$.

As determined using the spreadsheet function STDEV: $s_x = 5.74252876$, and $s_y = 4.229316688$.

Refer to Appendixes 2 and 3 for spreadsheet functions used.

Solve [6–7]

accurate and appropriate use of technology

$b = r \times \frac{s_y}{s_x}$

$= 0.301815739 \times \frac{4.229316688}{5.74252876}$

$= 0.222284362$

$a = \bar{y} - b\bar{x}$

$= 169.41 - 0.222284362 \times 168.67$

$= 131.9172967$

∴ The least-squares regression line equation for the data is given by:

$y = 131.9172967 + 0.222284362x$

$y = 131.92 + 0.22x$ (correct to two decimal places).

Formulate [3–4]

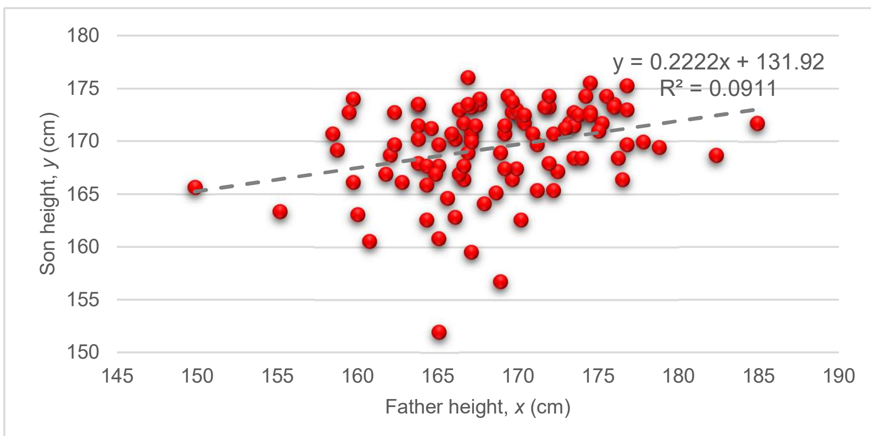
accurate translation of all aspects of the problem by identifying mathematical concepts and techniques

The regression line was added to the scatterplot using the trendline function of the spreadsheet program and the distribution of data points did not follow the trendline very closely (see Graph 2). The calculated correlation coefficient (r) value of 0.302 was confirmed using the coefficient of determination (R^2) value of 0.0911 generated by the spreadsheet program. The equation of the line, as determined by the spreadsheet trendline function, confirmed the regression equation constants calculated using formulas.

Solve [6–7]

accurate and appropriate use of technology

Graph 2: Scatterplot of son height vs father height with regression line



The result of 0.302 is close to zero, which indicates a (positive) weak correlation. The R^2 value of 0.0911 means that only 9% of the variation can be explained by the relationship between the heights of fathers and sons.

Communicate [3–4]

correct use of appropriate technical vocabulary, procedural vocabulary and conventions to develop the response

It is worth noting that if we expected a son to be the same height as his father, we would expect the slope (gradient) of the least-squares regression line to be exactly 1. The determined slope of 0.22 indicates that, in some instances, sons are taller than their fathers.

Evaluate and verify [4–5]

justification and explanation of decisions made using mathematical reasoning

4 Evaluation to verify results

The regression analysis confirmed a weak, positive association between a son's height and his father's. With such a low number of data points falling close to the regression line, the question was asked, is a linear equation most appropriate for the modelling of this data?

Testing the model with three friends' heights.

Friend	Father's actual height (x cm)	Predicted friend's height (y cm)	Friend's actual height (cm)
1	186	172.84	189
2	176	170.64	173
3	180	171.52	176

An example of the working for the estimation of a son's height:

$y = 131.92 + 0.22x$

$y = 131.92 + 0.22 \times 186$

$y = 172.84 \text{ cm}$

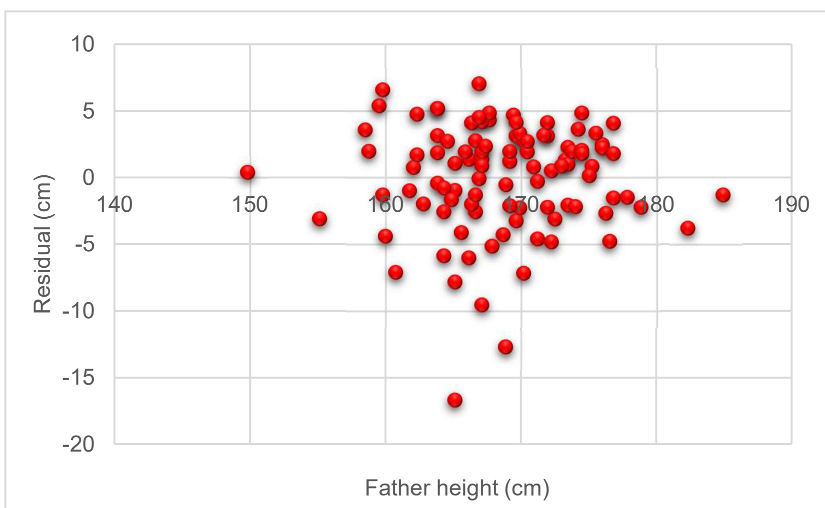
Using the friend's actual height data, it can be seen that the predicted heights are not accurate using this model based on the father's actual height.

Calculating residual values and producing a residual plot is a more reliable method of verifying if a linear relationship between two variables is a viable option.

Solve [6–7]

accurate and appropriate use of technology

Graph 3: Residual plot[1]

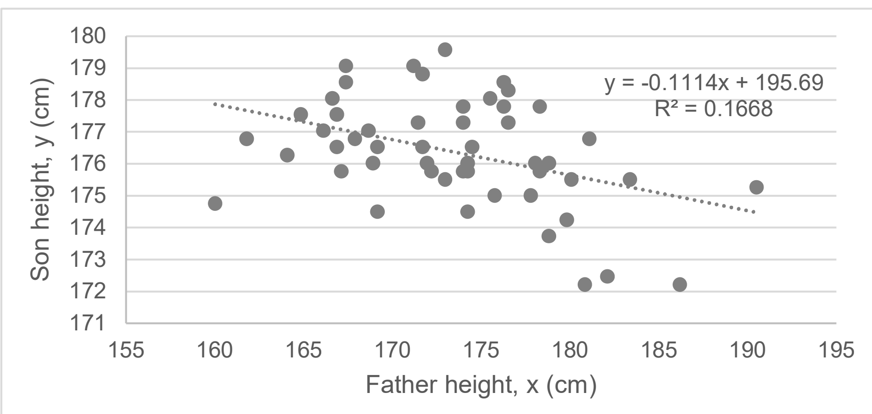


Evaluate and verify [4–5]

evaluation of the reasonableness of solutions by considering the results, assumptions and observations

The plot of the residuals against the explanatory variable determines if the linear model is an appropriate model to use in the regression analysis. If the distribution of points shows a random scatter across the x-axis, the

[1] Residual plot was produced using a spreadsheet program and summary output is provided in Appendix 4.

	relationship is most likely linear. It can be seen in Graph 3, the residual plot, that the points are randomly scattered across the x-axis. Therefore, it can be concluded that a linear regression model was appropriate for this study. In trying to explain the unexpected low correlation between son and father heights, the initial assumptions and observations were revisited. The following could explain the weak association:
Evaluate and verify [4–5] evaluation of the reasonableness of solutions by considering the results, assumptions and observations	• Errors may have occurred during the collection of data. • Self-reported data (father height) lacks reliability. • Some students may not have reached their full adult height. • Personal information about health and family circumstances is unknown and could be affecting the data.
	4.1 Improving the model
Evaluate and verify [4–5] documentation of the strengths and limitations of the solution	To improve the study, data collection of son heights and father heights could result in a stronger association. This would eliminate the assumption of growth. Directly measuring father height rather than using self-reported data could also improve reliability. Further, it must not be ignored that a son has two genetic pools — mother and father. Only the father has been studied in this investigation, with no acknowledgement of the influence the mother may have on the son's height. It would be an interesting study to also include the mothers' heights against their sons' heights. As an additional presentation of father–son height analysis, Pearson's data was accessed (The University of Alabama in Huntsville n.d.).
Communicate [3–4] correct use of technical and procedural vocabulary to develop the response	Randomly selecting 50 pairs (out of a total of 1078) of father and son heights from Pearson's 1903 dataset produced the scatterplot and regression equation shown in Graph 4. **Graph 4: Pearson's data, random sample n = 50, scatterplot and regression line**  Unexpectedly, this sample produced a negative moderate linear association with a stronger relationship between the variables than the father–son height in this study. However, when graphing Pearson's entire dataset, a positive linear association was found. This leads to the conclusion that the

	analysis is very dependent on the sample selected, and a sufficient sample size is required. A larger sample size with greater diversity, or the entire Year 12 male cohort, may result in a stronger association.
Evaluate and verify [4–5] documentation of the strengths and limitations of the solution and/or model	**4.2 Strengths and limitations** Some strengths of the model are that: • the method is generalisable and will work for other datasets or other hereditary characteristics, • it can tell us how much the father's height explains the son's height since the coefficient of determination measures how much of the variation is explained by the explanatory variable • the model's gradient can also tell us about the dependence of inherited characteristics on the value of the characteristic (e.g. if the gradient is 1, we expect father and son to be the same height; less than 1 and we might expect shorter fathers to have taller sons and vice-versa). Some limitations of the model include: • the sample size may be insufficient to make general claims about the predictions of heights; however, there is some evidence from the initial data analysis that is worth pursuing with additional data samples to produce a more reliable solution • there are other factors that affect height, e.g. mothers' and grandparents' heights, nutrition, general health, exercise (Danish 2017) The accuracy of the recorded data (father's height) may have been unreliable.
Communicate [3–4] Appropriate genre, including a suitable conclusion; coherent and concise organisation of the response, which can be read independently of the task sheet.	**5 Conclusion** The relationship between a father's height and his son's height is not very strong, according to this study. While there was a positive linear association, as demonstrated by the linear regression and correlation coefficient, the coefficient of determination (R^2) value of 0.0911 means this association is very weak. Given limitations such as the relatively small dataset, the self-reported heights of fathers, and the assumption that the students had reached their full adult height, it is not possible to draw a strong conclusion about whether a son's height is dependent on his father's height, from this study. Further research could be undertaken to investigate female students' heights compared to their mothers' heights. **Word count** Word count excluding, contents page, reference list, appendixes, data and tables is approximately 1743.

6 Appendixes

Appendix 1

Raw data father–son heights for n = 100

Father (cm)	Son (cm)
165.10	151.89
160.78	160.53
165.10	160.78
167.13	159.51
155.19	163.32
160.02	163.07
166.12	162.81
164.34	162.56
167.89	164.08
170.18	162.56
149.86	165.61
159.77	166.12
161.80	166.88
162.81	166.12
164.34	165.86
165.61	164.59
168.66	165.10
166.62	166.37
171.20	165.35
169.67	166.37
172.21	165.35
176.53	166.37
158.75	169.16
162.05	168.66
163.83	167.89
165.10	167.64
164.34	167.64
166.88	168.91
166.37	166.88
166.62	167.64
169.93	167.39
169.16	167.39
168.91	168.91
172.47	167.13
173.48	168.40
171.96	167.89

173.99	168.40
176.28	168.40
182.37	168.66
158.50	170.69
163.83	171.45
163.83	170.18
162.31	169.67
166.12	170.18
165.10	169.67
166.88	176.02
174.50	175.51
168.91	156.72
164.85	166.88
176.78	175.26
164.59	171.20
165.86	170.69
167.13	170.43
167.13	170.94
166.62	171.70
167.39	171.45
167.13	169.93
170.43	171.70
169.16	170.69
171.20	169.67
170.94	170.69
169.16	171.45
173.23	171.70
172.21	170.69
173.48	171.45
172.97	171.20
175.26	171.70
175.01	170.94
176.78	169.67
178.82	169.42
177.80	169.93
184.91	171.70
159.51	172.72
159.77	173.99
162.31	172.72
163.83	173.48
163.83	173.48
167.64	173.48
167.13	173.23

167.64	173.99
166.37	172.97
166.88	173.48
169.67	172.72
169.42	174.24
170.18	172.47
169.93	172.97
169.67	173.74
170.43	172.47
171.96	173.23
173.48	172.72
171.70	173.23
173.74	172.47
171.96	174.24
174.50	172.72
176.02	173.48
174.50	172.47
175.51	174.24
176.02	173.23
174.24	174.24
176.78	172.97

Appendix 2

Statistical measures calculated using a spreadsheet program

r = 0.301815739	
mean (x) = 168.67	mean (y) = 169.41
std dev (x) = 5.74252876	std dev (y) = 4.229316688

Appendix 3

Statistical functions used in a spreadsheet program

r = =CORREL(A2:A101,B2:B101)	
mean (x) = =AVERAGE(A2:A101)	**mean (y) =** =AVERAGE(B2:B101)
std dev (x) = =STDEV(A2:A101)	**std dev (y) =** =STDEV(B2:B101)

Appendix 4

Residuals

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.301747042
R Square	0.091051277
Adjusted R Square	0.08177629
Standard Error	4.052583781
Observations	100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	161.2266157	161.2266	9.816863	0.002280937
Residual	98	1609.496659	16.42344		
Total	99	1770.723275			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	131.9225292	11.9707077	11.02045	7.5E-19	108.1670493
X Variable 1	0.222240363	0.07093108	3.133187	0.002281	0.081479944

7 Reference list

Danish, E 2017, 'Factors affecting children's height', www.healthguidance.org/entry/14999/1/Factors-Affecting-Childrens-Height.html.

Goodreads n.d., 'Francis Galton', www.goodreads.com/author/show/3191106.Francis_Galton.

Revolvy n.d., 'Karl Pearson', www.revolvy.com/topic/Karl%20Pearson&item_type=topic.

The University of Alabama in Huntsville n.d., 'Pearson's height data', www.math.uah.edu/stat/data/Pearson.html.