General Mathematics 2019 v1.2

IA1 high-level annotated sample response

November 2019

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).

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 Topic 1
2. comprehend mathematical concepts and techniques drawn from Unit 3 Topic 1
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 Topic 1.

Queensland Government
Queensland Curriculum & Assessment Authority
191345

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
5. justify procedures and decisions by explaining mathematical reasoning drawn from Unit 3, Topics 1, 2 and/or 3.

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 about 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 decisions and/or the reasonableness of solutions.	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

Context
The term 'regression' was first used by Sir Francis Galton, who was interested in the study of biometrics (the application of statistical analysis to biological data), in the late 1800s. Galton's law of ancestral heredity was concerned with comparing measurable characteristics of children and their parents.[1] In 1903, Karl Pearson, assisted by Alice Lee, further investigated Galton's findings and developed conclusions relating to the stature of fathers and their adult sons, and mothers and their adult daughters.[2] [1] Bulmer, M 2003, *Francis Galton: Pioneer of heredity and biometry,* JHU Press, Baltimore. [2] Pearson, K and Lee, A 1903, 'On the laws of inheritance in man: I. Inheritance of physical characteristics', in *Biometrika*, vol. 2, no. 4, pp. 357–462, www.jstor.org/stable/2331507?seq=1#page_scan_tab_contents.
Task
Investigate the phenomenon of ancestral heredity by focusing on the height of a parent and a child of the same gender, 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.

Sample response

Criterion	Allocated marks	Marks awarded
Formulate Assessment objectives 1, 2, 5	4	4
Solve Assessment objectives 1, 6	7	6
Evaluate and verify Assessment objectives 4, 5	5	5
Communicate Assessment objective 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. He was the first to apply statistical methods to the study of human differences and ancestral heredity (Goodreads n.d.). Karl Pearson (1857–1936) has been credited with establishing the discipline of mathematical statistics and was a protégé of Galton (Revolvy n.d.). 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 son's height compared to father's height. A sample of 100 male students from the current Year 12 cohort was selected for the study.

2 Considerations

The sample of 100 was selected as an appropriate size to investigate the stature of sons and their fathers. Generally, an effective sample size is determined by

Formulate [3–4] accurate translation of problem by identifying mathematical concepts and techniques	$\sqrt{population}$, and since the Year 12 cohort has 213 male students, it was concluded that approximately 15 (i.e. $\sqrt{213}$) students would not be adequate to inform this study. The decision to include approximately half of the cohort was made to allow for a variety of height data. Year 12 students were chosen as most students would have finished growing by Year 12; therefore, a valid comparison between father and son heights could be made. The variables allocated for this investigation are father's height (x) in centimetres and son's height (y) in centimetres.
	2.1 Observations and assumptions Prior to collecting and analysing data, the following assumptions and observations have been formulated:
Formulate [3–4] documentation of appropriate assumptions and relevant observations	• I saw that all heights were measured accurately and recorded correctly. • From discussions with my peers I am confident that all students were able to accurately report their biological father's height. • According to Medical News Today (https://www.medicalnewstoday.com/articles/320676.php,2018) most boys reach their full height by the age of 16. Therefore, I have assumed that all boys are their full height for this investigation. • I have assumed that the dataset represents a typical male population with respect to height variations because my experience with boys from other schools leads me to believe that the heights of our students seem comparable.
	2.2 The plan To investigate the phenomenon of the ancestral heredity of height, the following procedures were undertaken.

Formulate [3–4] accurate translation of problem by identifying mathematical concepts and techniques	• Data collection:
	  – 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.
	  – Height data for father and son pairs was de-identified and aggregated by teachers, and a different random sample was provided to each student.
	• Scatterplot: The heights were graphed against each other to see if an association was apparent.
	• Regression equation: Once a linear association was assumed, a regression equation was developed after calculating Pearson's correlation coefficient, sample means and standard deviations.
	• Regression line: The spreadsheet trendline function was used to display the regression line on the scatterplot with the coefficient of determination.
	• Residuals: To evaluate if a linear model was appropriate in this study, residual values were calculated.
	• Residual plot: The residual values were plotted against father heights to observe the scatter pattern.
	• Analysis: The resulting statistical information was reflected on to adjust assumptions and form conclusions.

	2.3 Use of technology

A spreadsheet program was used extensively during the investigation process to organise 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.

Communicate [3–4]

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

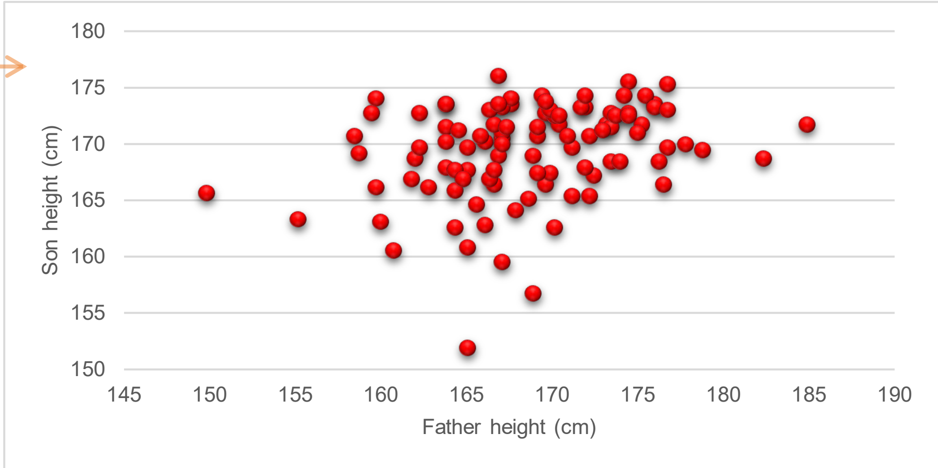
3 Developing a solution

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.

Solve [6–7]

accurate and appropriate use of technology

Graph 1: Scatterplot of son height against father height



Formulate [3-4]

documentation of relevant observations

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.

Solve [6–7]

accurate use of complex procedures to reach a valid solution, application of mathematical concepts and techniques relevant to the task

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

$\boldsymbol{y = a + bx}$

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.

Solve [6–7]

accurate and appropriate use of technology

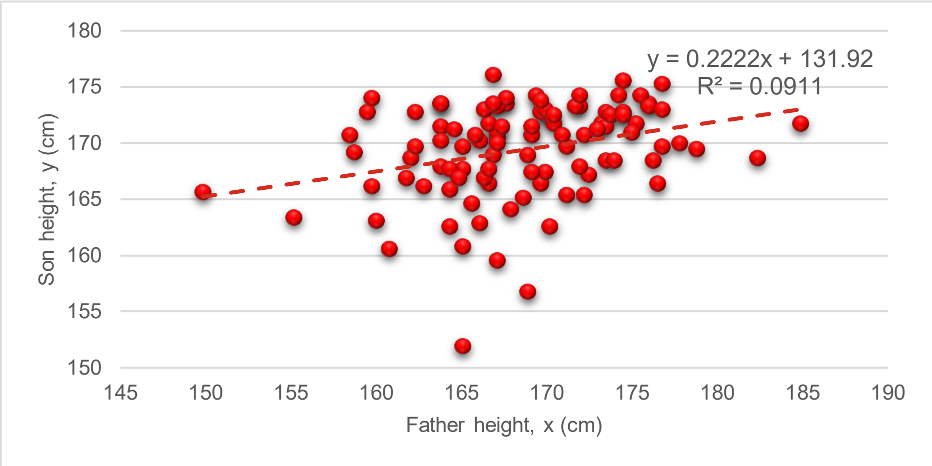
As determined using the spreadsheet function CORREL: $r = 0.301815739$.

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 Appendix 2 and 3 for spreadsheet functions used.

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

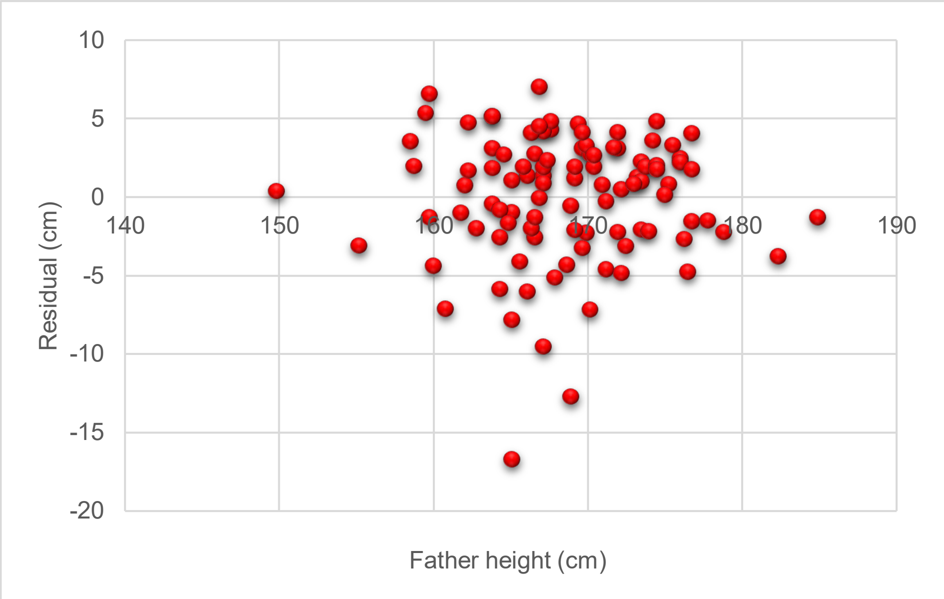
	$= 0.301815739 \times \frac{4.229316688}{5.74252876}$ $= 0.222284362$ $a = \bar{y} - b\bar{x}$ $= 169.41 - 0.222284362 \times 168.67$ $= 131.9172967$ ∴ The linear regression equation for the data is given by: $y = 131.9172967 + 0.222284362x$
Formulate [3–4] accurate translation of all aspects of the problem by identifying mathematical concepts and techniques	$y = 131.92 + 0.22x$ (correct to two decimal places). The regression line was added to the scatterplot using the trendline function of the spreadsheet program (see Graph 2). The equation of the line, as determined by the spreadsheet trendline function, confirmed the regression equation constants calculated using formulas. The distribution of data points did not follow the trend line very closely. The calculated correlation coefficient value of 0.302 was confirmed using the coefficient of determination (R^2) value of 0.0911 generated by the spreadsheet program.
Solve [6–7] accurate and appropriate use of technology	Graph 2: Scatterplot of son height against father height with regression line 
Communicate [3–4] correct use of appropriate technical vocabulary, procedural vocabulary and conventions to develop the response	A correlation coefficient value of one indicates a perfect positive linear association. The result of 0.302 indicates the initial description of the data distribution is correct, as the closer the value of the correlation coefficient is to zero, the weaker the association. The coefficient of determination (R^2) is used to describe how well the regression line represents the set of data. The closer R^2 is to one, the stronger the relationship between the variables. In the case of the height data presented, the R^2 value of 0.0911 is a weak correlation.
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 model appropriate for the modelling of this data?

Solve [6–7]

accurate and appropriate use of technology

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.

Graph 3: Residual plot[1]



The residual value is the distance the actual data value is from the value predicted using the regression equation. It can take on either negative or positive values, indicating if the actual value is below or above that of the predicted value. In this case, the residual value will represent the number of centimetres the son's height is below or above the height predicted by the regression equation.

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 independent 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 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 height, the initial assumptions and observations were revisited. The following could explain the weak association:

- 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.

Evaluate and verify [4–5]

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

4.1 Improving the model

To improve the study, data collection of adult son height and father height 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.

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

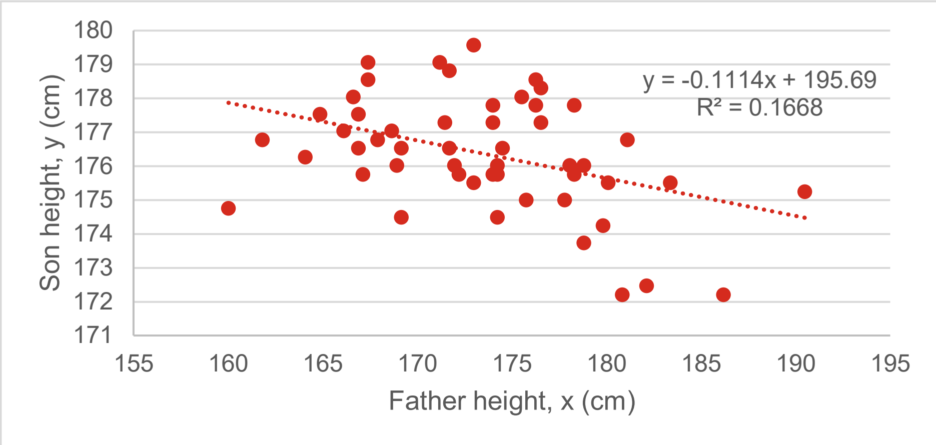
Evaluate and verify [4–5]

documentation of the strengths and limitations of the solution

As an additional presentation of father–son height analysis, Pearson's data was accessed (The University of Alabama in Huntsville n.d.).

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



Communicate [3–4]

correct use of technical and procedural vocabulary to develop the response

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 male Year 12 cohort, may result in a stronger association.

4.2 Strengths and limitations

A strength of the model is that a large sample from the whole cohort was used.

However, this investigation was limited to fathers' heights and many other factors (Danish 2017) and limitations could have been considered, such as:

- the mothers' and grandparents' heights
- nutrition
- general health
- exercise
- direct measurement of father heights.

Evaluate and verify [4–5]

documentation of the strengths and limitations of the solution

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 residual analysis, the coefficient of

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.	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.

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
172.47 173.48	167.13 168.40
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
163.83 163.83	173.48 173.48
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.