Become a Forensic Detective

Teacher Notes and Answers

7 8 9 **10** **11** 12

TI-Nspire Investigation Student 90 min

Introduction

A human skull and part of an arm have been found by a bushwalker in a national park. Although forensic pathologists can tell that the remains come from an adolescent, they would like to be able to determine the height of this person.

From the remains, the pathologists can measure the circumference of the head, the length of the forearm and the length of the middle finger. To help determine which of these would be the best indicator of height, your task is to analyse some relevant data.

The following table shows head circumference, forearm length, middle finger length and height measurements for 15 students.

Head circumference (cm)	Forearm length (cm)	Middle finger length (mm)	Height (cm)
57	25	80	163
55	25	69	156
56	25	80	171
56	26	90	185
54	22	75	150
58	27	89	169
57	25	80	150
57	27	82	172
59	26	79	175
59	23	80	169
57	24	78	166
61	28	85	188
58	27	95	179
59	23	74	162
57	24	85	170

Author: D. Tynan

Part 1: Analysing the data

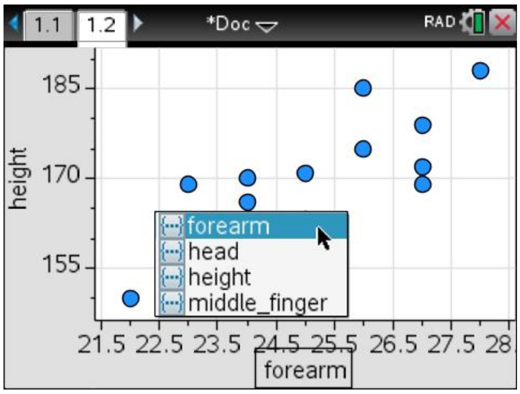
Question 1.

In the **Lists & Spreadsheet** application of a TI-Nspire CAS, enter the data from the table from the previous page.

The screenshot below shows the headings to use for each column.

- The column A heading is **head**.
- The column B heading is **forearm**.
- The column C heading is **middle_finger**. [Press **CTRL-SPACE** to type the underscore symbol.]
- The column D heading is **height**.

With a TI-Nspire CAS, a scatter plot can be created from data entered into the **Lists & Spreadsheet** application.



- Add a **Data & Statistics** page.
- Press **MENU**, select **Plot Properties** and then select **Add X Variable**.
- For each scatter plot, the 3 column headings, **head**, **forearm** and **middle_finger**, in turn, will be the horizontal axis variable.
- To add the vertical axis variable (**height**), select **Plot Properties** and then select **Add Y Variable**.

  [Note: If you need to change the window for a scatter plot, press **MENU**, select **Window/ Zoom** and then select **Window Settings**].

  Forearm length or middle finger length seem to be the best predictors of height, although all three appear to have moderately positive associations.

Question 2.

By drawing three different scatter plots, suggest whether head circumference, forearm length or middle finger length is the best predictor of a person's height.

To determine a line of best fit, the two-mean regression method will be used. The steps to follow for this method are outlined below.

- *Step 1*: Sort the data pairs from smallest to largest *x*-value. Note that data pairs with the same x-value are grouped and then ordered according to the corresponding y-value.
- *Step 2*: Identify the values that correspond to the sets $\bar{x}_L, \bar{x}_U, \bar{y}_L$, and $\bar{y}_U$ where, for example, x_L represents the lower half of the *x*-values and x_U represents the upper half of the *x*-values. As there are 15 data values, include the 8th value in both lower and upper halves.
- *Step 3*: Calculate the values of $\bar{x}_L, \bar{x}_U, \bar{y}_L$, and $\bar{y}_U$.
- *Step 4*: Identify the coordinates of the two mean value points $\left(\bar{x}_L, \bar{y}_L\right)$ and $\left(\bar{x}_U, \bar{y}_U\right)$.
- *Step 5*: Use these two points to calculate the equation of the regression line *y* = *mx* + *b* where and $m = \frac{\bar{y}_U - \bar{y}_L}{\bar{x}_U - \bar{x}_L}$ and $b = \bar{y}_L - m\bar{x}_L$. (Note that $b = \bar{y}_U - m\bar{x}_U$ could also be used.)

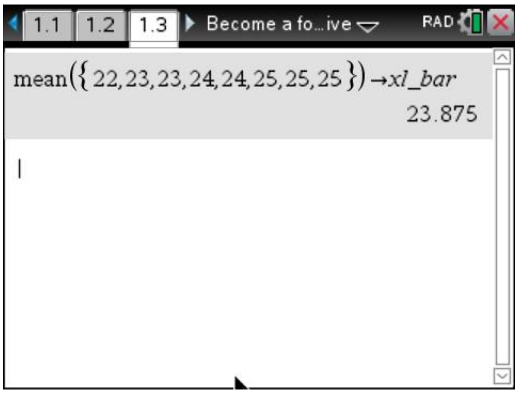
Author: D. Tynan

For the remainder of the investigation, assume that **forearm length** is the best predictor of height.

Question 3.

a) Calculate the values of $\bar{x}_L, \bar{x}_U, \bar{y}_L$, and $\bar{y}_U$ where y is the height and x is the forearm length.

Express your answer as $(\bar{x}_L, \bar{y}_L)$ and $(\bar{x}_U, \bar{y}_U)$



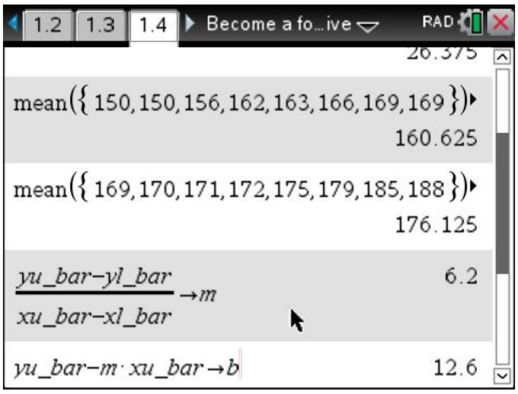
[Hints: To access the **Mean** command in the **Calculator** application of a TI-Nspire CAS press **MENU**, select **Statistics** and then select **List Math**. Store (**CTRL-VAR**) each of the four means as **xl_bar**, **yl_bar**, **xu_bar** and **yu_bar** respectively. The screenshots below show how the mean of the lower half of values for **forearm length** ($\bar{x}_L$) can be calculated and stored as **xl_bar**.

$(\bar{x}_L, \bar{y}_L)$ = (23.875, 160.625) and $(\bar{x}_U, \bar{y}_U)$ = (26.375, 176.125)

b) Use $m = \frac{\bar{y}_U - \bar{y}_L}{\bar{x}_U - \bar{x}_L}$ to show that m = 6.2.

Use *y* = *mx* + *b* to show that b = 12.6.

[Hint: In the **Calculator** application, store $\frac{\bar{y}_U - \bar{y}_L}{\bar{x}_U - \bar{x}_L}$ as *m* and either $\bar{y}_L - m\bar{x}_L$ or $\bar{y}_U - m\bar{x}_U$ as *b*.]



c) Hence state the two-mean regression line, *y* = *mx* + *b*, that models the relationship between these two variables.

y = 6.2 x + 12.6 or alternatively height = 6.2 × forearm length + 12.6

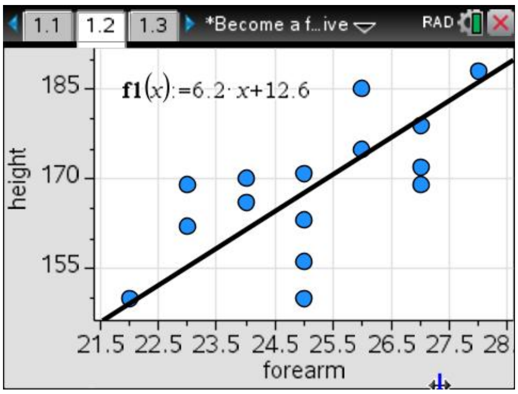
Question 4.

Plot the regression line from question 3 part c over the appropriate scatter plot from question 1, and answer the following questions.

[Hint: To plot the line on a **Data & Statistics** page, press **MENU**, select **Analyse** and then select **Plot Function**. Enter the equation of the line in the box that appears on the screen.]

a) Are the points 'close' to the line?

Approximately half the data points are close to the line.



b) How many of the points lie on either side of the line?

Six data points lie clearly above the line and five data points lie below the line. Four data points appear to lie on the line.

c) Do the points above the line have a similar spread to the points below the line?

Most of the data points above the line appear to be closer to the line than the data points below the line.

Author: D. Tynan

Part 2: Testing the rule

Now that you have developed a linear rule for linking the two variables *y* and *x* using the two-mean regression method, it would be useful to test how well the rule fits the given measurements.

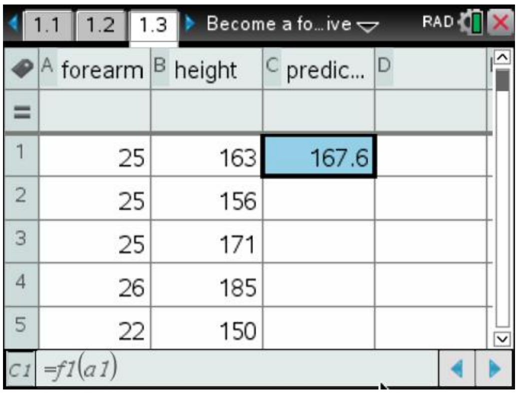
Use a new **Lists & Spreadsheet** page to set up the table below.

- The column A heading is **forearm**.
- The column B heading is **height**.
- The column C heading is **predicted_height**. [Press **CTRL-SPACE** to type the underscore symbol.]
- The column D heading is deviation.

Forearm (cm)	Height (cm)	Predicted height (cm)	Deviation (ignoring sign)
25	163		
25	156		
25	171		
26	185		
22	150		
27	169		
25	150		
27	172		
26	175		
23	169		
24	166		
28	188		
27	179		
23	162		
24	170		

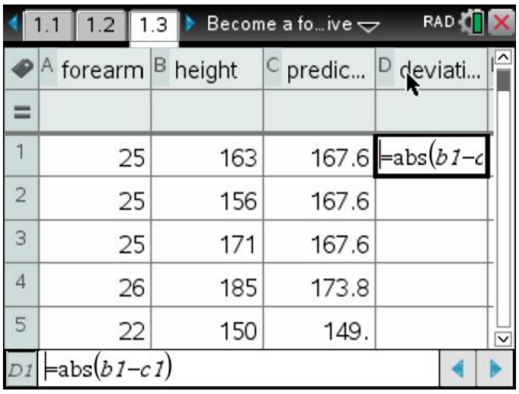
Question 5.

a) For each forearm length value, calculate the height, as predicted by your rule.
[Hint: In cell C1, enter the spreadsheet formula **= f1(a1)** (see the screenshot below) where f1(x) = 6.2x + 12.6 (found in question 3 part c). Use the Fill Down command (press **MENU**, select **Data** and then select **Fill**) to fill down from cell C1 to cell C15. Press **ENTER** to complete the calculation of the predicted heights.]



Author: D. Tynan

b) In the fourth column of the table, enter the deviation between the measured heights and the predicted heights.
Record the size of the difference only, not whether that difference is positive or negative. [Hint: In cell D1, enter the spreadsheet formula **= abs(b1 – c1)** where **abs** (absolute value found in the **Catalog**) allows us to record the size of the difference. Use the **Fill** command to fill down from cell D1 to cell D15. Press **ENTER** to complete the calculation of the deviations. See the screenshot.]

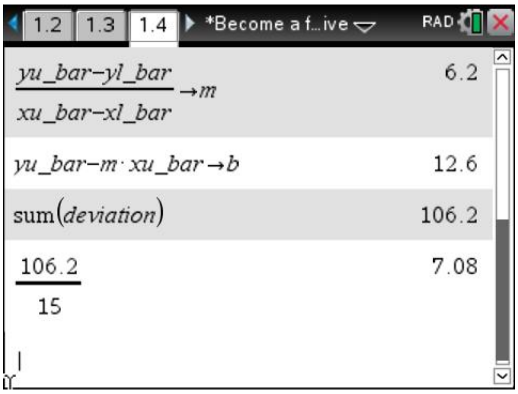


Predicted Heights: 167.6, 167.6, 167.6, 173.8, 149, 180, 167.6, 180, 173.8, 155.2, 161.4, 186.2, 180, 155.2, 161.4
Deviations: 4.6, 11.6, 3.4, 11.2, 1, 11, 17.6, 8, 1.2, 13.8, 4.6, 1.8, 1, 6.8, 8.6

Question 6.

As a rough guide to how well your rule fits the data, calculate the average deviation as follows.

a) Find the sum of the deviations in the fourth column.
[Hint: One approach using the **Calculator** application is to use the **Sum** command to sum the deviations.]
The sum of the deviations is 106.2



b) Divide your answer to question 6 part a by the number of students in the sample.
The average (mean) deviation is 7.1
[Hint: One approach using the **Calculator** application is to use the **Sum** command to sum the deviations and then divide this answer by 15.]

Question 7.

From the bones found in the national park, the forearm length measurement obtained by the forensic pathologists was 25 cm. Use the two-mean regression rule found in question c to predict the person’s height. Express this predicted height correct to one decimal place.
Predicted height is 6.2 x 25 + 12.6 = 167.6 cm (correct to one decimal place).

Author: D. Tynan

Part 3: Finding a rule with a movable line

To determine a line of best fit using a movable line, follow these instructions using your scatter plot from question 1 on a Data & Statistics page.

- Press **MENU**, select **Analyse** and then select **Remove Plotted Function**.
- Press **MENU**, select **Analyse** and then select **Add Movable Line**.
- Press **MENU**, select **Analyse**, then select **Residuals** and then select **Show Residual Squares**.

The sum of the areas of these residual squares displayed on the screen gives a numerical measure of how good a fit the line is to the data points.

The lower the value the better the fit.

The aim in question 8 is to minimise this sum of squares value by moving the line.

To move the line either:

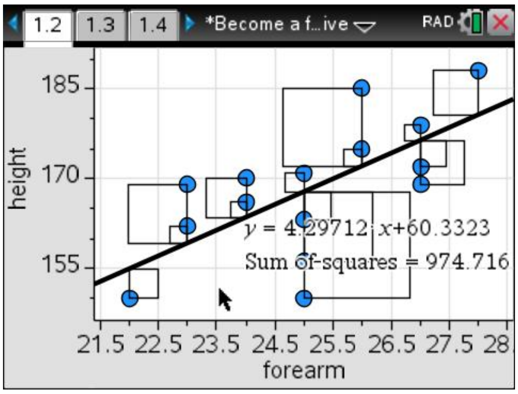
- Move the cursor over the line near one of the ends to obtain the rotational symbol, press **ENTER** to grab the line and rotate it, or:
- Move the cursor somewhere near the middle of the line to obtain the translational symbol, press **ENTER** to grab the line and move it up or down.

Question 8.

Use a combination of the two ways to move the line to try to obtain the smallest sum of squares value. Record the equation of your manual line of best fit and the sum of squares value.

Answers will vary.

Question 9.



Use the **Show Linear (mx+b)** command in the **Data & Statistics** application (select **Analyse** and then select **Regression**) to find the equation of the line of best fit and the minimum sum of the residual squares (select **Analyse**, then select **Residuals** and then **Show Residual Squares**). Express the value of *m* correct to three decimal places, the value of *b* correct to two decimal places and the sum of squares value correct to two decimal places. Compare your answer with the answer obtained in question 8.

y = 4.297x + 60.33 or alternatively height = 4.297 × forearm length + 60.33

The minimum sum of squares value is 974.72 (correct to two decimal places).

Question 10.

Compare the predicted height values obtained using the two-mean regression method and the least squares method.

Predicted height is 4.297 x 25 + 60.33 = 167.8 cm (correct to one decimal place).

Author: D. Tynan

Teacher notes

- This task is best suited to Year 10–11 students and is designed for them to explore which of the three body measurements (head circumference, forearm length or middle finger length) might be the best predictor of height. Students should find the context an interesting one as they are placed in the role of a forensic pathologist.
- Part 1 of the task requires students to enter the data into the Lists & Spreadsheet application with appropriate column headings. Ensure that students can generate scatter plots of height against the three independent variables (head circumference, forearm length and middle finger length) in the Data & Statistics application. Ensure also that students understand the concept of association, and that the pattern of points in a scatter plot indicates the strength of the association. Students also need to be familiar with terms such as positive, negative, strong and weak, when describing the strength of the association between two variables.
- Students are directed to use the two-mean regression method (the required steps are detailed in the task) to determine a line of best fit for height (dependent variable) against forearm length (the assumed best predictor of height). The Store command (press CTRL-VAR) in the Calculator application can be used to store the values of $\bar{x}_L, \bar{x}_U, \bar{y}_L$, and $\bar{y}_U$ as *xl_bar*, *xu_bar*, *yl_bar* and *yu_bar* respectively. The values of *m* and *b* in the two-mean regression equation $y = mx + b$ can then be calculated from $m = \frac{\bar{y}_U - \bar{y}_L}{\bar{x}_U - \bar{x}_L}$ as *m* and either $b = \bar{y}_L - m\bar{x}_L$ or $b = \bar{y}_U - m\bar{x}_U$ in the Calculator application with no loss of precision.
- In Part 2 of the task, students are required to undertake an informal approach to regression, in which students use spreadsheet formulas to calculate and average the absolute differences between the predicted height values and the actual height values for each forearm length. Ensure that students can construct the required spreadsheet by following the instructions offered in question 5 parts a and b. To fill down a column (say column C) from cell C1 to C15, highlight the starting cell (e.g. cell C1), press and hold down on the **DOWN** key for two seconds until a dashed border is seen around the cell. Use the Navigation Pad to fill down to cell C15 and then press **ENTER**. Note that the mean can also be used, as described in the task. To sum values in cells D1 to D15, enter **=sum(D1:D15)**.
- In Part 3 of the task, students are required to use the Movable Line command in the Data & Statistics application to determine a line of best fit by a combination of rotating and translating a movable line. The aim for students is to move the line in such a way as to minimise the sum of residual squares value displayed on the screen. Students are then required to compare the sum of squares value obtained with the movable line to that obtained in conjunction with the least squares regression command. Encouraging students to use the movable line approach should help them visualise and better understand the concept of least squares regression, a numeric routine that attempts to minimise the vertical distances between the points and the fitted line.

Author: D. Tynan