Determining the equation of the least squares line

Tuesday, 26 February 2019 5:59 PM

By the end of the lesson I would hope that you have an understanding and be able to apply to questions the following concepts:

- How to find the equation of the least squares line
- Know that it's very important to ensure you have the **explanatory** and **response variable** the correct war around
- Know how to use the CAS to find the equation of the least squares line.

RECAP:

In the last lesson we looked at how we use **residuals** to create a line of best fit which we call the least squares line. This is a specific way to get a line of best fit (but it's not the only way!).

We discovered that we can use the following formulae to help us find the equation of the least squares line:

Label of the x-axis

Label of the y-axis

$$y = a + bx$$

y-axis intercept

Gradient of the line

$y = mx + c$

$y = c + mx$

$y = a + bx$

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

r: Pearsons correlation coefficient
s_x: Standard deviation of x
s_y: Standard deviation of y

Slope of the line

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

$\bar{x}$: Mean of the x values
$\bar{y}$: Mean of the y values

y-axis intercept

Slope of line from previous equation

Finding the equation of the least squares line by hand

Let's do one question by hand, and then we can use the CAS.

The following question has been extracted, with permission, from the Cambridge Further Mathematics Units 3 and 4 textbook.

The heights (x) and weights (y) of 11 people have been recorded, and the values of the following statistics determined:

$\bar{x}$ =173.3 cm, s_x=7.444 cm, $\bar{y}$=65.45 cm, s_y=7.594 cm, r=0.8502

Use the formula to determine the equation of the least squares regression line that enable weight to be predicted from height. Calculate the slope and intercept correct to two significant figures.

Remember, we need to use the following formulae:

Weight /

predicted from height. Calculate the slope and intercept correct to two significant figures.

Remember, we need to use the following formulae:

$y = a + bx$

$b = \frac{rs_y}{s_x}$

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

$y = a + bx$

$b = \frac{0.8502 \times 7.594}{7.444} = 0.87$

$a = 65.45 - (0.867\ldots) \times 173.3$

$= \underline{\underline{-85}}$

$\text{Weight} = \underline{\underline{-85}} + 0.87 \times \text{Height}$

Finding the equation of the least squares using the CAS

Sorry ... but I'm using the CASIO Classpad at the moment. I'll add TI-Nspire instructions at a later date.

To be able to do this, you will need to be given the raw data (rather than the information shown above).

The following question has been extracted, with permission, from the Cambridge Further Mathematics Units 3 and 4 textbook.

The following data give the height (in cm) and weight (in kg) of 11 people.

Height (x)	177	182	167	178	173	184	162	169	164	170	180
Weight (y)	74	75	62	63	64	74	57	55	56	68	72

Determine and graph the equation of the least squares regression line that will enable weight to be predicted from height. Write the intercept and slope correct to three significant figures.

$a = -84.8$

$b = 0.867$

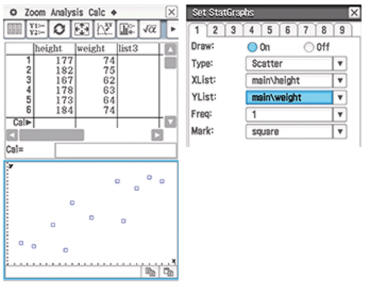
$\text{Weight} = -84.8 + 0.867 \times \text{Height}$

For your summary book

1 Open the **Statistics** application and enter the data into columns labelled **height** and **weight**.

2 Tap to open the **Set StatGraphs** dialog box and complete as shown.
Tap **Set** to confirm your selections.

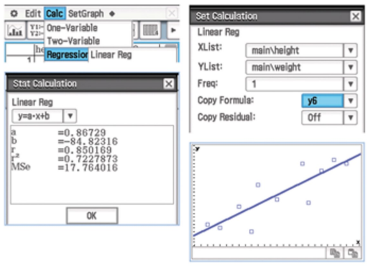
3 Tap in the toolbar at the top of the screen to plot the scatterplot in the bottom half of the screen.



4 To calculate the equation of the least squares regression line:

- Tap **Calc** from the menu bar.
- Tap **Regression** and select **Linear Reg**.
- Complete the **Set Calculations** dialog box as shown.
- Tap **OK** to confirm your selections in the **Set Calculations** dialog box. This also generates the regression results shown opposite.
- Tapping **OK** a second time automatically plots and displays the regression line.

Note: **y**6 as the formula destination is an arbitrary choice.



5 Use the values of the slope a and intercept b to write the equation of the least squares line in terms of the variables *weight* and *height*.

$Weight = -84.8 + 0.867 \times height$
(to three significant figures)
The coefficient of determination is $r^2 = 0.723$, correct to three significant places.

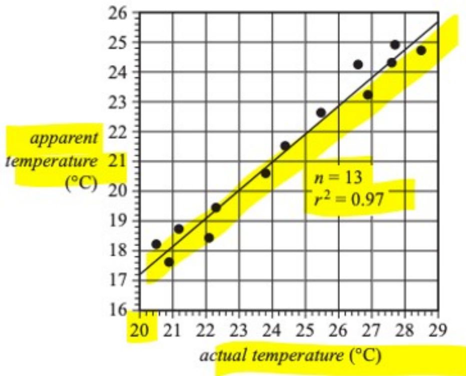
VCAA Exam Question on this concept
2016 Paper 2

Question 3 (8 marks)

The data in the table below shows a sample of actual temperatures and apparent temperatures recorded at the weather station. A scatterplot of the data is also shown.

The data will be used to investigate the association between the variables *apparent temperature* and *actual temperature*.

Apparent temperature (°C)	*Actual temperature* (°C)
24.7	28.5
24.3	27.6
24.9	27.7
23.2	26.9
24.2	26.6
22.6	25.5
21.5	24.4
20.6	23.8
19.4	22.3
18.4	22.1
17.6	20.9
18.7	21.2
18.2	20.5



$r^2 = 0.97$

$r = 0.98$

a. Use the scatterplot to describe the association between *apparent temperature* and *actual temperature* in terms of strength, direction and form. 1 mark

Strong, positive, linear

Strong positive association: r between 0.75 and 0.99
Moderate positive association: r between 0.5 and 0.74
Weak positive association: r between 0.25 and 0.49
No association: r between −0.24 and +0.24
Weak negative association: r between −0.25 and −0.49
Moderate negative association: r between −0.5 and −0.74
Strong negative association: r between −0.75 and −0.99

b. **i.** Determine the equation of the least squares line that can be used to predict the *apparent temperature* from the *actual temperature*.

Write the values of the intercept and slope of this least squares line in the appropriate boxes provided below.

Round your answers to two significant figures. 3 marks

apparent temperature = $\boxed{-1.7}$ + $\boxed{0.94}$ × *actual temperature*

ii. Interpret the intercept of the least squares line in terms of the variables *apparent temperature* and *actual temperature*. 1 mark

When my actual temp is 0°c, apparent temp is −1.7°c

c. The coefficient of determination for the association between the variables *apparent temperature* and *actual temperature* is 0.97

Interpret the coefficient of determination in terms of these variables. 1 mark

97% of the variation in apparent temp can be explained by the variation in actual temp.

0.97×100

97%

VCAA Exam Question on this concept
2017 Paper 2

Question 3 (6 marks)

The *number of male moths* caught in a trap set in a forest and the *egg density* (eggs per square metre) in the forest are shown in the table below.

Number of male moths	35	37	45	49	65	74	77	86	95
Egg density **(eggs per square metre)**	471	635	664	997	1350	1100	2010	1640	1350

a. Determine the equation of the least squares line that can be used to predict the *egg density* in the forest from the *number of male moths* caught in the trap.

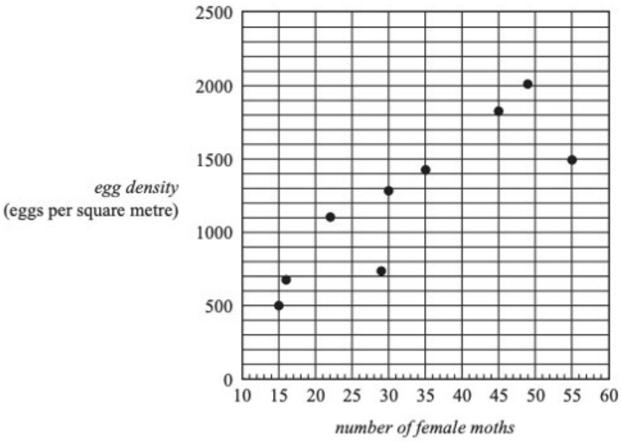
Write the values of the intercept and slope of this least squares line in the appropriate boxes provided below.

Round your answers to one decimal place. 2 marks

egg density = $\boxed{-46.8}$ + $\boxed{18.9}$ × *number of male moths*

b. The *number of female moths* caught in a trap set in a forest and the *egg density* (eggs per square metre) in the forest can also be examined.

A scatterplot of the data is shown below.



The equation of the least squares line is

egg density = 191 + 31.3 × *number of female moths*

i. Draw the graph of this least squares line on the **scatterplot on page 6**. 1 mark

(*Answer on the scatterplot on page 6.*)

ii. Interpret the slope of the regression line in terms of the variables *egg density* and *number of female moths* caught in the trap. 1 mark

For each increase by one of num of female moths egg density increase by 31·3

iii. The *egg density* is 1500 when the *number of female moths* caught is 55.

Determine the residual value if the least squares line is used to predict the *egg density* for this number of female moths. 1 mark

−412·5

Res = act − pred
= 1500 − 1912·5
= −412·5

iv. The correlation coefficient is $r = 0.862$

Determine the percentage of the variation in *egg density* in the forest explained by the variation in the *number of female moths* caught in the trap.

Round your answer to one decimal place. 1 mark

74·3%

$r^2 = (0{\cdot}862)^2$

$r^2 = 0{\cdot}74$