

Ex 4D Notes

We have been looking at bivariate data and estimating or calculating (using q_r) how strong the linear relationship is. In other words, how close to a straight line do the data points form.

Now we are going to look at how to work out what the actual linear equation is for the data.

This process is called **LINEAR REGRESSION**

REMEMBER: Straight lines have the general equation of:

$$y = mx + c$$

OR

$$y = a + bx$$

we used this
← in our linear
topic (Ch 3)

y intercept

gradient
or slope of
the line

We are trying to find the

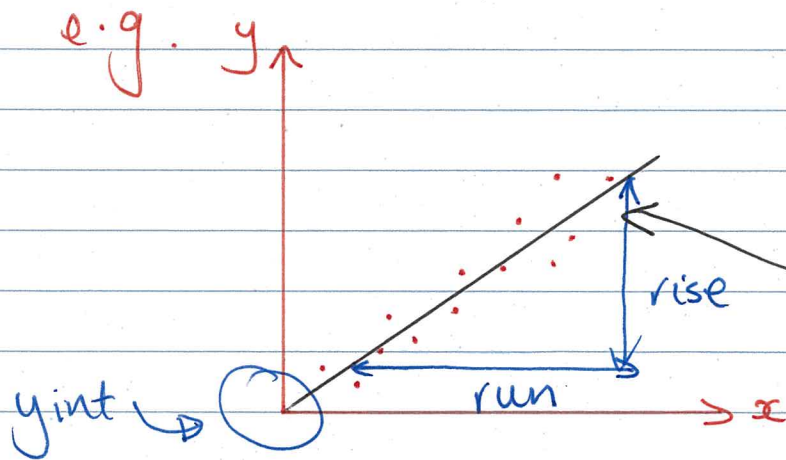
LINE OF BEST FIT

for the

data points

(2)

The line of best fit can be found 'by eye'. This means by looking at the points on the graph and ruling a line that suits the trend.



There is not only one answer

Try and have roughly the same amount of points either side of the line

You then use this line to work out the equation

Method

$$\text{gradient} = \frac{\text{rise}}{\text{run}}$$

(use points on the line !!)

(1)

yint = where the line crosses y axis (in my graph this is 0)

Method

(2)

Sometimes you can't get a good reading for the yintercept

so choose 2 points on the line then use CAS to find equation

Remember how??

(3)

when you write the equation, you must write using the actual variables
NOT x and y

Fitting line of Best Fit by eye gives a wide variety of equations and is NOT overly accurate when making future predictions.

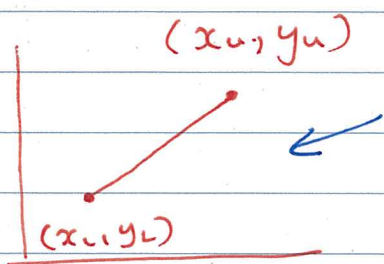
THE TWO MEAN METHOD is one way to improve accuracy.

steps involved: ① Rewrite data pairs in increasing numerical order

② Divide data into 2 halves (use 2 new tables)

$(x_L, y_L) \leftarrow$ ③ calculate mean (\bar{x}) for x and y in lower half

$(x_U, y_U) \leftarrow$ ④ calculate mean (\bar{x}) for x and y in upper half



⑤ Plot the 2 mean pts on the scatter plot

CAS \rightarrow ⑥ Use these points to work out equation.

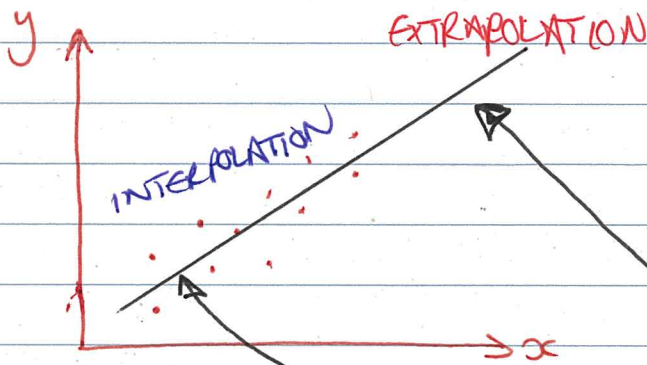
Ex 4E Notes

REMEMBER

Linear Regression

- this just means fitting a straight line to the data

The reason we do this is because it is easier to make predictions if you have a line to follow.



If we have line of best fit equation we can predict what will happen here or here

- If using the linear regression model to predict outside of the original data points this is **EXTRAPOLATION**
- If using to predict within the range of data it is **INTERPOLATION**