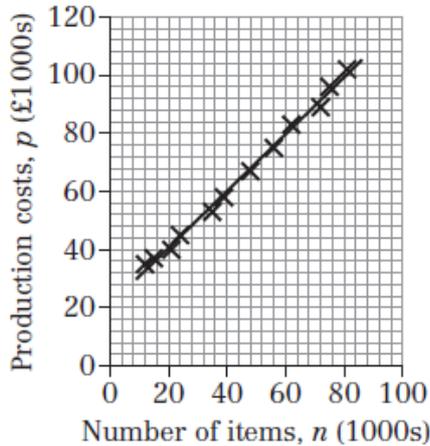


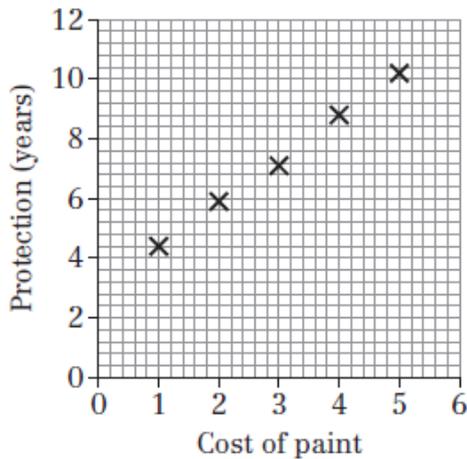
**Correlation 4B**

1 a and b



- c If the number of items produced is zero, the production costs will be approximately £21 000. If the number of items produced increases by 1000, the production costs increase by approximately £980.
- d The prediction for 74 000 is within the range of the data (interpolation) so is more likely to be reliable. The prediction for 95 000 is outside the range of the data (extrapolation) so is less likely to be reliable.

2 a

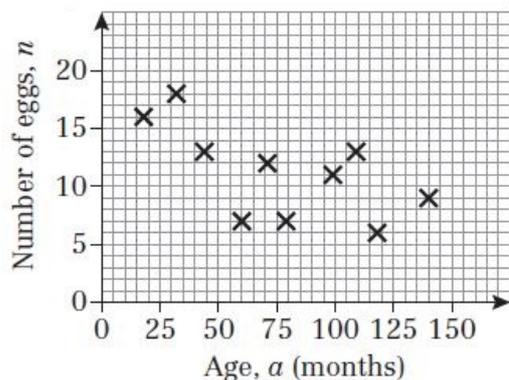


b There are two key problems with Helen’s statement:

First, 10 coats of paint is very far outside our range of given data, and we cannot assume that this linear relationship continues as we extrapolate, so using the regression line is not necessarily valid.

Second, even if we accept the extrapolation as valid, a gradient of 1.45 means that, for every extra coat of paint, the protection will increase by 1.45 years. Therefore, if 10 coats of paint are applied, the protection will be 14.5 years longer than if no paint were applied. Helen has, however, forgotten to include the constant 2.93 years, which is the weather resistance if no paint were applied. After 10 coats of paint the protection will last approximately  $2.93 + 14.5 = 17.43$  years.

3 a



b The scatter diagram shows weak negative correlation, therefore the gradient in the regression equation, given as 0.063, should be negative.

4 This is not a reasonable statement as there are unlikely to be any houses with no bedrooms, so she is extrapolating outside of the range of data, where the linear relationship is unlikely to continue.

5 a If the humidity increases by 1%, the decrease in visibility is approximately 106 m.

b High humidity (> 95%) can give rise to misty and foggy conditions, which in turn will cause lower visibility. Hence there is likely to be a causal relationship.

c i The prediction for 100% is outside the range of the data (extrapolation) so is less likely to be accurate.

ii This particular regression equation should only be used to predict a value for  $v$  given  $h$ .

d The data provided is only useful for analysing the first two weeks of September. Random values throughout September should be used and analysis made of the whole month. The sample size could also be increased across multiple months as data between May and October is available.