



QUESTION WORKBOOK

Least squares regression and residuals

Further Statistics 2 · Year FM

7 questions · 26 marks

PRINT EDITION

SECTION A · Warm-up

- A1** Define a residual, and state the quantity the least squares line makes as small as possible. [2]

- A2** For the points $(2, 5)$, $(4, 9)$, $(6, 10)$, $(8, 14)$, $(10, 17)$, find the regression line of y on x . [3]

SECTION B · Standard

- B1** The regression line of y on x for the points $(2, 5)$, $(4, 9)$, $(6, 10)$, $(8, 14)$, $(10, 17)$ is $y = 2.3 + 1.45x$. Find the five residuals and the residual sum of squares, checking the total two ways. [4]

- B2** A test score y is recorded against x hours of practice for five students, giving $(2, 5)$, $(4, 9)$, $(6, 10)$, $(8, 14)$ and $(10, 17)$. The regression line is $y = 2.3 + 1.45x$. Interpret both coefficients, predict the score for 7 hours, and say why a prediction for 20 hours would be unwise. [4]

- B3** Describe what a plot of residuals against x should look like if a straight-line model is appropriate, and what two patterns would argue against it. [3]

- B4** Five points are recorded as $(2, 5)$, $(4, 9)$, $(6, 10)$, $(8, 14)$ and $(10, 17)$, and the regression line of y on x is $y = 2.3 + 1.45x$. The last point is then found to have been misrecorded, and should read $(10, 30)$. Recalculate the regression line and comment on the effect. [4]

SECTION C · Stretch

- C1** Five pairs of readings are coded using $p = (x - 6)/2$ and $q = y - 10$, giving the coded points $(-2, -5)$, $(-1, -1)$, $(0, 0)$, $(1, 4)$ and $(2, 7)$. Find the regression line of q on p , and hence find the regression line of y on x . [6]