6.4 Correlation and regression

Name: _____ Class: _____ Date: _____ **Total: 22 marks**

KEY WORDS regression line • correlation coefficient

Correlation and regression

Objective

Compute and interpret the product moment correlation coefficient, fit and use a least-squares regression line, reverse coding on S statistics, and respect the interpolation range (objectives 4.1–4.3).

1 Key knowledge and worked examples

Machinery. $S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$ (same shape for S_{yy}, S_{xy}); $r = \dfrac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$; regression of y on x: $y = a + bx$ with $b = \frac{S_{xy}}{S_{xx}}$, $a = \bar{y} - b\bar{x}$. Coding $w = cd$ reverses as $S_{tw} = c\,S_{td}$, $S_{ww} = c^2 S_{dd}$. Predictions only INSIDE the data range.

Worked example – coded regression (WST01 June 2025 Q3). With $w = 11.5d$, $n = 12$, $S_{tt} = 26.43$, $S_{tw} = -91.55$, $\sum w = 339.25$, $\sum t = 9.1$, $\sum w^2 = 10036.45$: $S_{ww} = 445.57$; $r = \frac{-91.55}{\sqrt{26.43 \times 445.57}} = -0.844$ (warmer $\Longrightarrow$ less gas); w on t: $b = -3.46$, $a = 30.9$; dividing by 11.5, $d = -0.301t + 2.69$; at $t = 2$, $d = 2.09$ m^3, and each $+1$ °C cuts about 0.3 m^3.

2 Practice

2.1 Given $S_{xx} = 40$, $S_{yy} = 90$, $S_{xy} = -54$, find r and interpret it. [3]

2.2 With $\sum x = 60$, $\sum y = 84$, $\sum x^2 = 410$, $\sum y^2 = 750$, $\sum xy = 522$ and $n = 10$: find the equation of the regression line of y on x. [4]

2.3 For the regression line $y = 2.5x - 3$ fitted to data with x between 4 and 12, explain why predicting y at $x = 20$ is unreliable. [2]

3 Exam-style questions

3.1 *(Edexcel IAL WST01 June 2025 Q3, condensed.)* Madison models daily temperature t (°C) against daily gas consumption d (m^3) using coding $w = 11.5d$ on a 12-day sample, with $S_{tt} = 26.43$, $S_{tw} = -91.55$, $\sum w = 339.25$, $\sum t = 9.1$, $\sum w^2 = 10036.45$. (a) Show that $S_{ww} = 445.57$. (b) Find S_{td} and S_{dd}. (c) Find the pmcc between d and t. (d) Interpret your value in context. (e) Show that the regression line of w on t is $w = -3.46t + 30.9$. (f) Hence estimate the gas consumption when $t = 2$, and interpret the change per extra degree. [13]

Solutions

2.1 $r = \frac{-54}{\sqrt{40\times 90}} = \frac{-54}{60} = -0.9$: a strong negative linear correlation (as x rises, y falls fairly consistently).

2.2 $S_{xy} = 522 - \frac{60\times 84}{10} = 18$; $S_{xx} = 410 - \frac{60^2}{10} = 50$; $b = \frac{18}{50} = 0.36$; $\bar{x} = 6$, $\bar{y} = 8.4$; $a = 8.4 - 0.36(6) = 6.24$: $y = 6.24 + 0.36x$.

2.3 $x = 20$ lies outside the observed range 4–12: extrapolation assumes the linear pattern continues, with no evidence to support it.

3.1 (a) $S_{ww} = 10036.45 - \frac{339.25^2}{12} = 10036.45 - 9590.88 = 445.57$ ✓. (b) $S_{td} = \frac{S_{tw}}{11.5} = \frac{-91.55}{11.5} = -7.96$; $S_{dd} = \frac{S_{ww}}{11.5^2} = \frac{445.57}{132.25} = 3.37$. (c) $r = \frac{-91.55}{\sqrt{26.43\times 445.57}} = -0.844$ (coded or uncoded give the same r). (d) As average daily temperature increases, daily gas consumption decreases (a strong negative linear relationship). (e) $b = \frac{-91.55}{26.43} = -3.4638$; $a = \frac{339.25}{12} - (-3.4638)\frac{9.1}{12} = 28.271 + 2.627 = 30.897$: $w = 30.9 - 3.46t$ ✓. (f) $d = \frac{w}{11.5}$: at $t = 2$, $w = 30.9 - 6.92 = 23.98$, so $d = 2.09$ m^3; the gradient for d is $\frac{-3.46}{11.5} = -0.301$: each extra degree of average temperature cuts consumption by about 0.3 m^3.

Local review guidance: S statistics from the summary data with the n division; coding reversed by the right power of the factor; interpretations in context with units, and predictions flagged as within range only. This is authored classroom guidance.