MAT4MDS — Practice 5 Worked Solutions

Model Answers to Practice 5

Question 1.

(a) The independent variable is $x =$ depth, and the dependent (or response) variable is $y =$ CO_2 concentration (it is the one we are estimating).

$$n = 8\,, \qquad \sum_{i=1}^{n} x_i = 4459, \qquad \sum_{i=1}^{n} x_i^2 = 3076083, \qquad \sum_{i=1}^{n} x_i y_i = 933719, \qquad \sum_{i=1}^{n} y_i = 1778,$$

so we solve for α and β in

$$\begin{bmatrix} 3076083 & 4459 \\ 4459 & 8 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 933719 \\ 1778 \end{bmatrix}$$

to get $\alpha = -0.097\,,\ \beta = 276.307\,$, that is,

$$y = -0.097x + 276.31.$$

You can solve this either by determing the inverse of the 2×2 matrix and then multiplying, or you can solve the 2 simultaneous equations

$$\begin{aligned} 3076083\alpha + 4459\beta &= 933719 \\ 4459\alpha + 8\beta &= 1778 \end{aligned}$$

by the usual method of elimination. Use whichever seems simplest for the problem at hand.

(b) When $x = 1050$ m, $y = -0.097 \times 1050 + 276.307 = 174.4$ ppm.

(c) The slope -0.097 means that for every 1 m increase in depth, the CO_2 concentration *decreases* by 0.097 ppm on average. You can also say that for every metre further below the surface, the CO_2 concentration *decreases* by an average of 0.097 ppm.

Question 2.

(a) In this question, we want to estimate the depth, therefore the independent variable x is CO_2 concentration and the dependent variable y is depth.

$$n = 8\,, \qquad \sum_{i=1}^{n} x_i = 1778, \qquad \sum_{i=1}^{n} x_i^2 = 403328, \qquad \sum_{i=1}^{n} x_i y_i = 933719, \qquad \sum_{i=1}^{n} y_i = 4459,$$

so we solve for α and β in

$$\begin{bmatrix} 403328 & 1778 \\ 1778 & 8 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 933719 \\ 4459 \end{bmatrix}$$

to get $\alpha = -7.015\,,\ \beta = 2116.424\,$, that is,

$$y = -7.015x + 2116.424.$$

When $x = 200$ ppm, $y = -7.015 \times 200 + 2116.424 = 713.5$ m.

(b) The least squares line of best fit minimises the sum of the squares of the residuals, and the residuals are the distances in the dependent (y) direction which in this question is along the "depth" axis. The line found in Question 1 minimises the sum of squares of residuals along the "CO_2 concentration" axis which is perpendicular to the "depth" axis.

Question 3.

(a) As always, first identify the independent variable x, and the dependent variable y. Then,

$$n = 4\,, \quad \sum_{i=1}^{n} x_i = 6, \quad \sum_{i=1}^{n} x_i^2 = 14, \quad \sum_{i=1}^{n} x_i y_i = 25, \quad \sum_{i=1}^{n} y_i = 11,$$

so we solve for α and β in

$$\begin{bmatrix} 14 & 6 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 25 \\ 11 \end{bmatrix}$$

to get $\alpha = 17/10 = 1.7$, $\beta = 1/5 = 0.2$, that is,

$$y = 1.7x + 0.2.$$

For parts (ii) and (iii) the residuals and sums of squares are

x x_i	y y_i	LLS best fit line $y = 1.7x + 0.2$	residual $y_i - y$	residual^2	other line $y = 1.6x + 0.3$	residual $y_i - y$	residual^2
0	1	0.2	0.8	0.64	0.3	0.7	0.49
1	1	1.9	−0.9	0.81	1.9	−0.9	0.81
1	3	3.6	−0.6	0.36	3.5	−0.5	0.25
3	6	5.3	0.7	0.49	5.1	0.9	0.81
			Sum	2.3		Sum	2.36

(b) See table above.

(c) The sum of squares of residuals is less for the LLS line of best fit.

(d) We expect the answer in (c) because the LLS line of best fit has the least sum of squares of residuals among *all* lines.

Question 4.

(a) As usual,

$$n = 6\,, \quad \sum_{i=1}^{n} x_i = 96.6, \quad \sum_{i=1}^{n} x_i^2 = 2071.62, \quad \sum_{i=1}^{n} x_i y_i = 7.89, \quad \sum_{i=1}^{n} y_i = 0.668,$$

so we solve for α and β in

$$\begin{bmatrix} 2071.62 & 96.6 \\ 96.6 & 6 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 7.89 \\ 0.668 \end{bmatrix}$$

to get $\alpha = -0.00555$, $\beta = 0.201$, that is,

$$y = -0.0055x + 0.201.$$

(b) The slope of -0.0055 (and the data) indicates that on average, as the snow cover increases the wind stress decreases. Turning this the other way around we see that the wind stress is *increasing* as the snow cover is getting less.
[We need to be careful here in how we interpret this statement. The data, and the line of best fit, do not by themselves prove that higher wind stress is caused by the decrease in snow cover. However, it is evidence in that direction, and is put together with other observations to establish the effects of global warming.]

Question 5.

(a) Because the we are modelling an exponential growth, we need to transform the variables. It will be convenient to use P = population (millions).

$$\text{independent variable:} \quad x = \text{year} - 1951$$

$$\text{dependent variable:} \quad y = \ln(P)$$

so that the linear relationship $y = \alpha x + \beta$ is equivalent to the exponential relationship $P = P_0\, e^{\alpha x}$, with $\beta = \ln(P_0)$ and $P_0 = e^{\beta}$.

The data set now looks like:

year	1951	1961	1971	1981	1991	2001	2011
x = time elapsed since 1951	0	10	20	30	40	50	60
P = population (millions)	361	439	548	683	846	1029	1211
$y = \ln(P)$	5.889	6.084	6.306	6.526	6.741	6.936	7.099

Proceeding as usual:

$$n = 7\,, \qquad \sum_{i=1}^{n} x_i = 210, \qquad \sum_{i=1}^{n} x_i^2 = 9100, \qquad \sum_{i=1}^{n} x_i y_i = 1425.155, \qquad \sum_{i=1}^{n} y_i = 45.582,$$

so we solve for α and β in

$$\begin{bmatrix} 9100 & 210 \\ 210 & 7 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 1425.155 \\ 45.582 \end{bmatrix}$$

to get $\alpha = 0.0206$, $\beta = 5.894$, that is,

$$y = 0.0206x + 5.894.$$

With $P_0 = e^{5.894} = 362.854$, this gives the exponential model

$$P = 362.854\, e^{0.0206x}$$

where x is the number of years since 1951.

(b) This model estimates that the population in 2011 ($x = 60$) is $362.854\, e^{0.0206\times 60} = 1248.9$ million, that is, 1.2489 billion. It is quite close to the actual population in 2011 even though the growth rate had decreased in the previous decade or two.