Erice School 2024 – linear least squares by hand

John S. O. Evans* and Ivana Radosavljevic Evans

Department of Chemistry, Durham University, Science Site, South Road, Durham DH1 3LE, U.K.

STUDENT PROBLEMS: LEAST SQUARES BY-HAND

It's worth trying to do a least squares problem by hand once in your life, if only to convince you of the power of using a computer. Answer for the problem can be checked with the Excel LINEST function and should be: $m = 5.3$, $c = -4.5$, σ_m=0.79, $\sigma_c = 2.17$. The correlation coefficient is $\mu_{mc} = -0.91$.

Q1. Find the best fit line that passes through the points (1,2), (3,5), (6,10). Calculate the standard uncertainty and correlation coefficient between the gradient and intercept. Comment on the sign of the correlation coefficient. Check your answer using, e.g. the excel LINEST function.

Q2. Add a constraint that the line should pass through the origin; how does the value of the gradient change under this constraint?

Q3. Repeat Q1 but using a restraint that the line should pass through the origin.

METHOD

You can solve this problem using the type of matrix expressions used in most least squares programs and applying a simple least squares recipe (the proof is given in many texts on statistics).

Express the experimental data as a series of observational equations, one for each of the three observations:

$$m \times x_1 + c = y_1$$

$$m \times x_2 + c = y_2$$

$$m \times x_3 + c = y_3$$

Where x_i is the independent variable and y_i the dependent variable. These equations can be expressed in matrix form as:

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

This equation has the form:

$$\mathbf{Ax} = \mathbf{b}$$

Where **A** is called the design matrix, **x** is the vector of unknowns (what we want to solve for) and **b** is the vector of observations. The least squares recipe to obtain the best-fit **x** is to pre-multiply each side of this equation by the transpose of $\mathbf{A^T}$:

$$\mathbf{A^T A x = A^T b}$$

then pre-multiply each side by $(A^T A)^{-1}$ to solve for **x**:

$$\mathbf{x} = \left(\mathbf{A^T A}\right)^{-1} \mathbf{A^T b}$$

The standard uncertainties and correlation coefficient can be calculated from the variance-covariance matrix **M**:

$$\mathbf{M} = \begin{pmatrix} \sigma_m^2 & \sigma_m \sigma_c \mu_{mc} \\ \sigma_m \sigma_c \mu_{mc} & \sigma_c^2 \end{pmatrix}$$

where σ_m^2 is the variance of m, $\sigma_m \sigma_c \mu_{mc}$ is the covariance of m and c, and μ_{mc} is the correlation coefficient. The values of each entry in this matrix can be calculated from the equation:

$$\mathbf{M} = \frac{1}{n-p} \sum_{1}^{n} w_i (y_{obs} - y_{calc})^2 \left(\mathbf{A^T A}\right)^{-1}$$

Where n is the number of observations and p the number of parameters.

To repeat the calculation using a constraint, adjust the problem so that the equation used is $y = mx$. i.e. you change the equations being used in the problem.

To add a restraint, you include (0, 0) as an extra observation. To increase the weighting on the restraint you could add the extra observation in twice, three times, or more. Alternatively, you could add a weighting matrix to the problem with zeros everywhere except the diagonal. The diagonal should contain either 1 or a higher number for each weight. The equations become:

$$\mathrm{W}\mathbf{Ax} = \mathrm{W}\mathbf{b}$$

$$(\mathrm{A^T W A})\mathbf{x} = (\mathrm{A^T W})\mathbf{b}$$

$$\mathbf{x} = (\mathrm{A^T W A})^{-1} (\mathrm{A^T W})\mathbf{b}$$

ANSWERS

Q1. For our specific example we can plug data values into these equations and express the observational equations as:

$$\begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 6 & 1 \end{pmatrix}\begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 10 \end{pmatrix}$$

Pre-multiplying both sides by the transpose $\mathbf{A}^T$:

$$\begin{pmatrix} 1 & 3 & 6 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 6 & 1 \end{pmatrix}\begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} 1 & 3 & 6 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 5 \\ 10 \end{pmatrix}$$

gives:

$$\begin{pmatrix} 46 & 10 \\ 10 & 3 \end{pmatrix}\begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} 77 \\ 17 \end{pmatrix}$$

Note that this represents two equations with two unknowns. The inverse matrix $(\mathrm{A}^T\mathrm{A})^{-1}$ is:

$$\frac{1}{38}\begin{pmatrix} 3 & -10 \\ -10 & 46 \end{pmatrix}$$

Pre-multiplying both sides gives:

$$\begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} 1.60526 \\ 0.31579 \end{pmatrix}$$

That is, m = 1.60526 and c = 0.31579.

To get the variance-covariance matrix we need to calculate the sum of $(y_{obs} - y_{calc})^2$:

x	y_{obs}	y_{calc}	$(y_{obs} - y_{calc})^2$
1	2	1.921	6.24×10^{-3}
3	5	5.131	0.0172
6	10	9.94	2.938×10^{-3}

The sum of the $(y_{obs} - y_{calc})^2$ column is 0.0263. We then calculate $\mathbf{M}$ as:

$$\boldsymbol{M} = \begin{pmatrix} \sigma_m^2 & \sigma_m\sigma_c\mu_{mc} \\ \sigma_m\sigma_c\mu_{mc} & \sigma_c^2 \end{pmatrix} = \frac{1}{1} \times 0.0263 \times \frac{1}{38} \times \begin{pmatrix} 3 & -10 \\ -10 & 46 \end{pmatrix} = \begin{pmatrix} 0.00208 & -0.0069 \\ -0.0069 & 0.0318 \end{pmatrix}$$

Giving σ_m=0.05, σ_c = 0.18, μ_{mc} = −0.86. The negative correlation coefficient is as we would expect: as the gradient m decreases (the line becomes less steep), the intercept c would increase.

Adding a constraint that the line should pass through the origin can be done by repeating the calculation with the observational equations $y = mx$. You can work through the problem in the same way and should get m = 1.674. Constraining the best fit line to pass through y = 0 rather than y = +0.31579 makes the gradient steeper as we would intuitively expect.

Adding an extra observation that the line goes through (0,0) leads to:

$$\begin{pmatrix} 1 & 3 & 6 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 6 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} 1 & 3 & 6 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 5 \\ 10 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 46 & 10 \\ 10 & 4 \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} 77 \\ 17 \end{pmatrix}$$

$$\begin{pmatrix} 1/21 & -5/42 \\ -5/42 & 23/42 \end{pmatrix} \begin{pmatrix} 77 \\ 17 \end{pmatrix} = \begin{pmatrix} 1.64 \\ 0.14 \end{pmatrix}$$

This gives m = 1.64 and c = 0.14. The line is closer to going through the origin and the gradient is steeper. Adding a weight matrix with a weight of 10 on the line going through the origin changes the problem to:

$$\begin{pmatrix} 46 & 10 \\ 10 & 14 \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} 77 \\ 17 \end{pmatrix}$$

$$\begin{pmatrix} 7/272 & -5/272 \\ -5/272 & 23/272 \end{pmatrix} \begin{pmatrix} 77 \\ 17 \end{pmatrix} = \begin{pmatrix} 1.669 \\ 0.02 \end{pmatrix}$$

The higher the weighting, the closer the line is to going through the origin and the steeper the gradient. The gradient of m = 1.67 is approaching the value with a constraint that the line passes through the origin. As we applied an arbitrary weighting scheme it doesn't make sense to calculate standard uncertainties.