LINEAR REGRESSION

A different means of obtaining a best fit line

José Brenes André,
Escuela de Física, University of Costa Rica

The best fit line is a common tool in metrological analysis, fundamentally during the calibration process at any level. It is regularly used by assuming that all the uncertainties can be loaded onto the independent variable, a practice that can now be considered as the standard method for its widespread use. A way to derive the least square fit is presented, and applied to several special cases. A comparison between the results thus obtained and those predicted by the ordinary least square are discussed.

Ordinary Least Square (OLS) from a physical point of view

The Ordinary Least Square (OLS) procedure is usually presented in a rather mathematical way, whereby the sum of the squares of the uncertainties of each point S(m,b) = Σ (y – mx – b)^2 is set to a minimum by the right choice of the slope **m** and the intercept **b** of a straight line, considered as the best fit line, i.e. the appropriate derivatives are taken and set equal to zero:

$$\frac{\partial S(m,b)}{\partial m} = \frac{\partial}{\partial m}\sum (y - mx - b)^2 = -2\sum (y - mx - b)(x) = 0$$

$$\frac{\partial S(m,b)}{\partial b} = \frac{\partial}{\partial b}\sum (y - mx - b)^2 = -2\sum (y - mx - b) = 0$$

If Δ = (y – mx – b) is considered as a deformation of a spring of constant unity, it would play the role of a force acting in an externally fixed predefined direction, the y direction in this case. Then the two conditions Σ Δ = 0 and Σ x Δ = 0 are recognized as the translational and the rotational conditions for a body to be in equilibrium. Based on these ideas the author has built a mechanical device that shows these analogies, which act as an analogue computer.

Derivation of Least Normal Squares (LNS)

This method can also be applied to deduce the best fit line obtained by the LNS method, and eventually expanded to deduce the structural line, which includes the other best fit methods as special cases.

To do that let us start by making the translation u = x – **X** and v = y – **Y** where (**X,Y**) are the coordinates of the centroid. Because of the translational equilibrium condition the best fit line in the (u,v) plane has to pass through the origin, and hence passes through the centroid **Y** = m**X** + b.

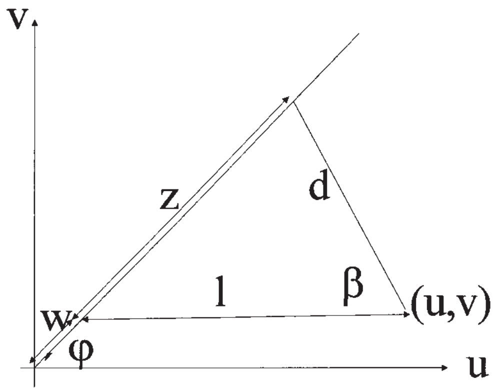

Figure 1

$$d = l\cos\beta = (u - w\cos\varphi)\, sin\varphi = u\, sin\varphi - v\cos\varphi$$

$$z + w = l\, sin\beta + {}^{v}\!/_{sin\varphi} = v\, sin\varphi + u\cos\varphi$$

Using Fig. 1 in the special case φ + β = π/2 we can deduce, by simple geometrical arguments, the lever arm and the torque respectively for each data point (u,v), which correspond to a rotation. This is the starting point the author used in the paper *Uncertainty of measurements of calibrated equipment* to deduce the LNS equation from other points of view.

The translational equilibrium condition gives, as expected:

$$\sum\Delta = \sum (u\, sin\varphi - v\cos\varphi) = 0 \qquad \sum u = 0 \qquad \sum v = 0$$

This is another way of saying that the intercept in the (u,v) plane is the origin. The rotational equilibrium condition gives Σ (u sinφ – v cosφ) (v sinφ + u cosφ) = 0 in turn, from which one easily finds:

$$tan 2\varphi = \frac{2\sum vu}{\sum (u^2 - v^2)}$$

The inversion property of the LNS can also be deduced from physical arguments. By definition, we took the force to be the distance from the data point to the best fit line. Hence we can define a force along the u axis F_u and another along the v axis F_v. This is the $\lambda = 1$ case, which allows the best fit line to be inverted, a property very useful from the metrological point of view.

Figure 2 shows that the components of both forces along the perpendicular to the best fit line have to be equal, and hence (taking note of the directions):

$$(m_1 y_i + b_1 - x_i)\, sin\varphi = (-m_0 x_i - b_0 + y_i) \cos\varphi \qquad m_0 = \frac{sin\varphi}{\cos\varphi}$$

from which $m_0\, m_1 = 1$.

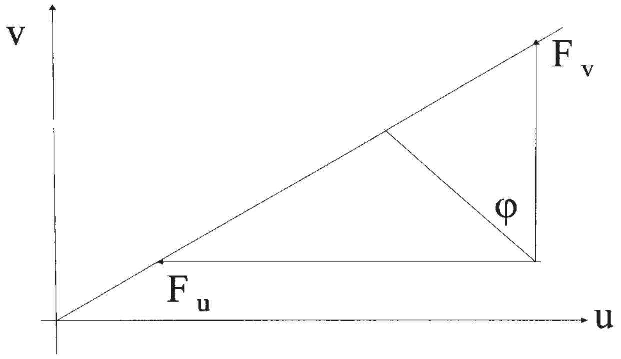

Figure 2

We now apply the main ideas to the most general $\varphi + \beta \neq \pi/2$ case, for which $l_0 = w + z$, $v = w \sin\varphi$ and $l = u - w \cos\varphi$.

Translational equilibrium requires:
$\Sigma\, l \sin\varphi = \Sigma\, (u \sin\varphi - v \cos\varphi) = 0$ as obtained in the special case of LNS, whereas rotational equilibrium leads to $\Sigma\, l_0\, l \sin\varphi = 0$.

Since this is a general deduction, the equation $z = l\omega$ is introduced, where ω is a factor to be defined afterwards. With the help of this equation one finds:

$$\sum (u\, sin\varphi - v\cos\varphi)\left(l\,\omega + {}^{v}\!/_{sin\varphi}\right) = 0$$

A little algebraic manipulation leads to:

$$\sum v^2 - \sum u^2 \frac{(\omega\ sin^2\varphi)}{(l - \omega\cos\varphi)\cos\varphi} = \sum uv \frac{(l - 2\omega\ \cos\varphi)\, sin\varphi}{(l - \omega\cos\varphi)\cos\varphi} \qquad (1)$$

Aside from having an impressive form, this equation includes the factor ω which still has to be defined. We can give it a more "friendly" form by defining:

$$\lambda = \frac{(\omega\ sin^2\varphi)}{(l - \omega\cos\varphi)\cos\varphi} \qquad (2)$$

and re-write equation (1) as:

$$\frac{\sum y^2 - \lambda \sum u^2}{2\sum uv} = \frac{tan^2\varphi - \lambda}{2\ tan\varphi} \qquad (3)$$

$$tan^2\varphi + \left(\frac{\lambda\sum u^2 - \sum v^2}{\sum uv}\right) tan\varphi - \lambda = 0 \qquad (4)$$

where $\tan\varphi$ is the required slope.

At this point it may be interesting to look back at the OLS case. Let us start with the relationship $\tan\varphi = \Sigma\, v^2 / \Sigma\, uv = r\, \sigma_y / \sigma_x$ and use it to calculate:

$$tan2\varphi = \frac{2tan\varphi}{1 - tan^2\varphi} = \frac{2\left({}^{r\sigma_y}\!/\!_{\sigma_x}\right)}{1 - \left({}^{r\sigma_y}\!/\!_{\sigma_x}\right)^2}$$

which is a quadratic that needs to be solved for tan φ. Not surprisingly the root with the plus sign reproduces $r\ \sigma_y / \sigma_x$, but that for the minus sign gives $-\sigma_x / r\sigma_y$, which is the slope of the best fit if we take the y axis as having null uncertainty.

Statistics behind equation (4)

Although this is a more friendly looking formula one still has the problem of how λ (or ω for that matter) relates to the statistical part of the problem. To overcome this objection one can, following and using Kendall´s notation and numbering [1], start with:

$\xi_i = x_i + \delta_i$

$\eta_i = y_i\ + \varepsilon_i$ (29.12)

$y_i = \alpha_o + \alpha_1\, x_i$ (29.13)

(a) $E(\xi) = \mu$

(b) $E(\eta) = \alpha_o + \alpha_1\, \mu$ (29.19)

(c) $s^2_\xi\ = \sigma^2_\delta + \sigma^2_x$

(d) $s^2_\eta\ = \sigma^2_\varepsilon + \alpha^2_1\ \sigma^2_x$

(e) $\alpha_1\, \sigma^2_x =\ s_{\xi\eta}$

Kendall, using maximum likelihood arguments and restricting σ^2 to be non-negative, obtained the set of six inequalities (29.20), from which he finds (29.21):

$(s^2_\eta\ - \sigma^2_\varepsilon\ ) = \alpha_1\, s_{\xi\eta} \quad \alpha_1\, (s^2_\xi\ - \sigma^2_\delta) =\ s_{\xi\eta}$

These correspond to equations (1) and (2) in Madansky´s [2] paper, who warns the reader that “neither (1), (2), (3), nor (4) are maximum likelihood estimates of β”. It is interesting to read that “If $\sigma_\delta{}^2$ is known then: $\alpha_1 = s_{\xi\eta} / (s_\xi^2 - \sigma_\delta^2)$”. Note these are different values obtained depending if one knows $\sigma_\varepsilon{}^2$ or λ. But how does mathematics know which case is involved?

Dividing the numerator and the denominator of equation (3) by N, the number of data, one can associate:

$$\frac{\sum v^2}{N} = s_\eta^2 = \sigma_\varepsilon^2 + \alpha_1^2\ \sigma_x^2$$

$$\frac{\sum u^2}{N} = s_\xi^2 = \sigma_\delta^2 + \sigma_x^2$$

$$\frac{\sum uv}{N} = s_{\xi\eta}$$

$$\alpha_1 = tan\varphi$$

Because $\alpha^2 - \lambda_- \neq 0$ equation (4) can be re-written as:

$$\theta = \frac{\alpha^2 - \lambda}{2\alpha} = \frac{\left(\sigma_\varepsilon^2 + \alpha_1^2\ \sigma_x^2\right) - \lambda\left(\sigma_\delta^2 + \sigma_x^2\right)}{2s_{\xi\eta}} = \frac{\sigma_x^2\left(\alpha_1^2 - \lambda\right) - \left(\lambda\sigma_\delta^2 - \sigma_\varepsilon^2\right)}{2s_{\xi\eta}}$$

Several alternatives are possible. For instance if:

$$\left(\lambda\sigma_\delta^2 - \sigma_\varepsilon^2\right) = 0 \qquad (5)$$

one finds the usual definition of λ, for which case equation (4) turns into:

$$\alpha_1^2\, s_{\xi\eta} + \alpha_1(\lambda s_\xi^2 - s_\eta^2) - \lambda\, s_{\xi\eta} = 0 \text{ and } \alpha_1\, \sigma_x^2 = s_{\xi\eta}$$

If on the other hand $\sigma_\varepsilon^2 = R\, \alpha_1^2\ \sigma_x^2$ and $\sigma_x^2 = R\, \sigma_x^2$, then $\alpha_1\, \sigma_x^2\, (1 + R) = s_{\xi\eta}$, no definition of λ is possible and $\alpha_1^2 = \sigma_\varepsilon^2 / \sigma_x^2$, which is equation (4) of Lindley [3], obtained from maximum likelihood arguments (and heavily objected to by him).

Several comments can now be made referring to the way in which the structural equation are deduced here. First, equation (4) was obtained from physical and geometrical arguments, without imposing any restrictions on any of the estimators used, avoiding the objection posed by Kendall (equation 29.8), Lindley [3], and Solari [4] that “in fact no maximum likelihood solution exists for this problem” (Robertson [5], page 357). Second, there is no need to study all the different possibilities between the estimators, as was to objected by Kendall (equation 29.17) and by Birch [6]. Third, rather than be amazed as Birch was for cases (v) and (vi) of his paper where he writes “It is notable that the formula for α_1 is the same as that for α_1 in case (i)”, with this derivation it is only natural that it has to be so.

Last but not least, the fact that equation (29.27) can only be obtained from geometrical considerations, that (29.19) (c) and (d) leads to (29.19) (e) and to an expression for λ, suggests that the structural line has a deeper meaning than initially supposed.

Several special cases are reproduced if one uses geometry to find the value of ω, equation (2) to find the value of λ, and equation (5) to evaluate σ. Such three special cases are presented in Table 1.

It is worth noting that Lindley writes: “In many applications it will be known that the two errors (in x and y) are about the same order of magnitude. This might imply that λ lies between k^{-1} and k for suitable k, when quite strong results about the posterior distribution of θ can be made by the methods of Section 4”, which seems to be in the line of ω, which in turn is related to λ.

Description	ω	λ	σ	Tan φ
OLS along y-axis	1/ cosφ	∞	$\sigma_\varepsilon{}^2 = 0$	Σ uv /Σ u^2
LNS, lines perpendicular	Cosφ	1	$\sigma_\delta{}^2 = \sigma_\varepsilon{}^2$	
OLS along x-axis	0	0	$\sigma_\delta{}^2 = 0$	Σ v^2 /Σ uv

Table 1

Comparison of both methods

The author has applied both methods to several sets of points, and compared the results predicted by each of them. Although he grants from the start that the sets used are very unlikely to appear in any real measurement, the two methods do not have any in-built supposition that prevent us from applying them to such point distribution.

Four distributions will be studied, composed of the following points:

Case A (1,1) (–1,1)
Case B (1,–1) (1,1)
Case C (1,1) (–1,1) (–1,–1) (1,–1)
Case D (1,2) (1,–2) (–1,–2) (–1,2)

The distributions are presented in Table 2 showing the intermediate values necessary to calculate tan 2φ, φ of the LNS, as well as the value of the slope m obtained by the OLS.

Case	Σu^2	Σv^2	Σuv	φ	m
A	2	0	0	0	0
B	0	2	0	π/2	0/0
C	4	4	0	0/0	0
D	16	4	0	π/2	0

Table 2

It can be seen that for Case A both methods give the expected null slope of a horizontal line.

Should we deal with a vertical line (Case B) the ordinary method gives an undefined value, rather than the expected ∞, obtained by the LNS.

For the four corners of a square (Case C) the ordinary method gives a null slope, not recognizing the symmetry of the distribution, put in evidence by the LNS result of 0 / 0.

If the square is deformed to a rectangle with its larger side on the horizontal (not shown), both methods give a null slope, as expected. But if its larger side is vertical (Case D) this difference is not accounted for by the ordinary method ($m = 0$ again) as it is by the LNS ($\varphi = \pi/2$).

Conclusion and proposal

Due to the fact that the structural equation seems to give more reasonable results than the OLS method commonly used, the author would like to propose that the LNS method be considered as an alternative method to carry out line fitting. To further support this claim, he would like to draw attention to the odd situation that occurs when the OLS is used: mainly, that two different results for the slope are obtained when the OLS method is used (see for example [7]), which can be avoided if the LNS method is used from the start, for it intrinsically allows for inversion.

Although there exists some contradiction between studies done by Lakshminarayanan and Gunst [8] who suggest that some 200r data points are required to obtain slope values within 1 % (r = number of times every point is replicated), and Robertson [5] who writes "We see from these results that for sensible parameter values n does not need to be large to make the first-order approximations good enough for practical purposes", the fact is that today it is very common to take the data points digitally with the help of a given (and probably not very expensive) interface with a PC. Hence the number of data points may not now be a problem even in small laboratories in developing countries. ■

References

[1] Kendall, M.G. and Stuart,A. The advanced theory of Statistics, 2, Chap 29, London: Griffin.

[2] Mandansky, A. The fitting of straight lines when both variables are subject to error, Am. Statis. Ass. J. 54 (1959) 173–205

[3] Lindley, D.V. and El-Syyad, G.M. The Bayesian estimation of a linear functional relationship, J. Roy. Statis.Soc 30 (1968) 190–202

[4] Solari, M. The Maximum likelihood solution of the problem of estimating a linear functional relationship, J. Roy Statis. Soc 31 (1969) 372–375.

[5] Robertson, C. Large-sample theory for the linear structural relation, Biometrika (1974), 61, 2, pp. 353–359

[6] Birch, M.W. A note on the maximum likelihood estimation of a linear structural relationship, Am. Statis.Ass. J. 59, (1964) 1175–1178

[7] Berkson, J. Are there two regressions?, J. Am. Statis. Ass. 45 (1950) 164–180.

[8] Lakshminarayanan, M and Gunst, R. Estimation of parameters in linear structural relationships: Sensitivity to the choice of the ratio of error variances, Biometrika (1984) 71, 3, pp. 569–573

The author may be contacted by e-mail:
jbrenes@cariari.ucr.ac.cr