A mathematical procedure for finding the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets ("the residuals") of the points from the curve. The sum of the squares of the offsets is used instead of the offset absolute values because this allows the residuals to be treated as a continuous differentiable quantity. However, because squares of the offsets are used, outlying points can have a disproportionate effect on the fit, a property which may or may not be desirable depending on the problem at hand.

In practice, the vertical offsets from a line (polynomial, surface, hyperplane, etc.) are almost always minimized instead of the perpendicular offsets. This provides a fitting function for the independent variable $X$ that estimates $Y$ for a given $X$ (most often what an experimenter wants), allows uncertainties of the data points along the $X$- and $Y$-axes to be incorporated simply, and also provides a much simpler analytic form for the fitting parameters than would be obtained using a fit based on perpendicular offsets. In addition, the fitting technique can be easily generalized from a best-fit line to a best-fit polynomial when sums of vertical distances are used. In any case, for a reasonable number of noisy data points, the difference between vertical and perpendicular fits is quite small.

The linear least squares fitting technique is the simplest and most commonly applied form of linear regression and provides a solution to the problem of finding the best fitting straight line through a set of points. In fact, if the functional relationship between the two quantities being graphed is known to within additive or multiplicative constants, it is common practice to transform the data in such a way that the resulting line is a straight line, say by plotting $T$ vs. $\sqrt{T}$ instead of $T$ vs. $t$ in the case of analyzing the period $T$ of a pendulum as a function of its length $l$. For this reason, standard forms for exponential, logarithmic, and power laws are often explicitly computed. The formulas for linear least squares fitting were independently derived by Gauss and Legendre.
For nonlinear least squares fitting to a number of unknown parameters, linear least squares fitting may be applied iteratively to a linearized form of the function until convergence is achieved. However, it is often also possible to linearize a nonlinear function at the outset and still use linear methods for determining fit parameters without resorting to iterative procedures. This approach does commonly violate the implicit assumption that the distribution of errors is normal, but often still gives acceptable results using normal equations, a pseudoinverse, etc. Depending on the type of fit and initial parameters chosen, the nonlinear fit may have good or poor convergence properties. If uncertainties (in the most general case, error ellipses) are given for the points, points can be weighted differently in order to give the high-quality points more weight.

Vertical least squares fitting proceeds by finding the sum of the squares of the vertical deviations \( R^2 \) of a set of \( n \) data points

\[
R^2 = \sum [y_i - f(x_i, \alpha_1, \alpha_2, \ldots, \alpha_n)]^2
\]  

(1)

from a function \( f \). Note that this procedure does not minimize the actual deviations from the line (which would be measured perpendicular to the given function). In addition, although the unsquared sum of distances might seem a more appropriate quantity to minimize, use of the absolute value results in discontinuous derivatives which cannot be treated analytically. The square deviations from each point are therefore summed, and the resulting residual is then minimized to find the best fit line. This procedure results in cutlying points being given disproportionately large weighting.

The condition for \( R^2 \) to be a minimum is that

\[
\frac{\partial (R^2)}{\partial \alpha_i} = 0
\]  

(2)

for \( i = 1, \ldots, n \). For a linear fit,

\[
f(a, b) = a + b x,
\]  

(3)

so

\[
R^2 (a, b) = \sum_{i=1}^{n} [y_i - (a + b x_i)]^2
\]  

(4)

\[
\frac{\partial (R^2)}{\partial a} = -2 \sum_{i=1}^{n} [y_i - (a + b x_i)] = 0
\]  

(5)

\[
\frac{\partial (R^2)}{\partial b} = -2 \sum_{i=1}^{n} [y_i - (a + b x_i)] x_i = 0.
\]  

(6)

These lead to the equations

\[
n a + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i
\]  

(7)

\[
a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i
\]  

(8)

In matrix form,

\[
\begin{bmatrix}
\sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\
\sum_{i=1}^{n} x_i & n
\end{bmatrix}
\begin{bmatrix}
\alpha \\
b
\end{bmatrix}
=
\begin{bmatrix}
\sum_{i=1}^{n} y_i \\
\sum_{i=1}^{n} x_i y_i
\end{bmatrix}
\]  

(9)
so
\[
\begin{bmatrix}
a \\
b \\
\end{bmatrix} = \left[ \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} \right]^{-1} \left[ \begin{bmatrix}
\sum_{i=1}^{n} y_i \\
\sum_{i=1}^{n} x_i y_i \\
\end{bmatrix} \right]
\]
\[\text{(10)}\]

The 2x2 matrix inverse is
\[
\begin{bmatrix}
a \\
b \\
\end{bmatrix} = \frac{1}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \left[ \begin{bmatrix}
\sum_{i=1}^{n} y_i & \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \\
\sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i \\
\end{bmatrix} \right]
\]
\[\text{(11)}\]

so
\[
\alpha = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}
\]
\[= \frac{\bar{y} (\sum_{i=1}^{n} x_i^2) - \bar{x} \sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 - n \bar{x}^2}
\]
\[\text{(12)}\]

\[
b = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}
\]
\[= \frac{(\sum_{i=1}^{n} x_i y_i) - n \bar{x} \bar{y}}{\sum_{i=1}^{n} x_i^2 - n \bar{x}^2}
\]
\[\text{(13)}\]

(Kenney and Keeping 1962). These can be rewritten in a simpler form by defining the sums of squares
\[
ss_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2
\]
\[= \left( \sum_{i=1}^{n} x_i^2 \right) - n \bar{x}^2
\]
\[\text{(16)}\]

\[
ss_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2
\]
\[= \left( \sum_{i=1}^{n} y_i^2 \right) - n \bar{y}^2
\]
\[\text{(18)}\]

\[
ss_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})
\]
\[= \left( \sum_{i=1}^{n} x_i y_i \right) - n \bar{x} \bar{y}
\]
\[\text{(20)}\]

which are also written as
\[
\alpha = \frac{ss_{xy}}{ss_{xx}}
\]
\[\text{(22)}\]

\[
\beta = \frac{ss_{xy}}{ss_{yy}}
\]
\[\text{(23)}\]

\[
\text{cov}(x, y) = \frac{ss_{xy}}{n}
\]
\[\text{(24)}\]

Here, \(\text{cov}(x, y)\) is the covariance and \(\alpha^2\) and \(\beta^2\) are variances. Note that the quantities \(\sum_{i=1}^{n} x_i y_i\) and \(\sum_{i=1}^{n} x_i^2\) can also be interpreted as the dot products.
\[
\sum_{i=1}^{n} x_i^2 = \mathbf{x} \cdot \mathbf{x}
\] (25)

\[
\sum_{i=1}^{n} x_i y_i = \mathbf{x} \cdot \mathbf{y}.
\] (26)

In terms of the sums of squares, the regression coefficient \( b \) is given by

\[
b = \frac{\text{CCV}(\mathbf{x}, \mathbf{y})}{\sigma_x^2} = \frac{ss_{xy}}{ss_{xx}},
\] (27)

and \( \alpha \) is given in terms of \( b \) using (\( \Delta \)) as

\[
\alpha = \bar{y} - b \bar{x}.
\] (28)

The overall quality of the fit is then parameterized in terms of a quantity known as the correlation coefficient, defined by

\[
r^2 = \frac{ss_{xy}^2}{ss_{xx} \cdot ss_{yy}},
\] (29)

which gives the proportion of \( ss_{yy} \) which is accounted for by the regression.

Let \( \hat{y}_i \) be the vertical coordinate of the best-fit line with \( x \)-coordinate \( x_i \), so

\[
\hat{y}_i = \alpha + b \cdot x_i
\] (30)

then the error between the actual vertical point \( y_i \) and the fitted point is given by

\[
e_i = y_i - \hat{y}_i.
\] (31)

Now define \( s^2 \) as an estimator for the variance in \( e_i \),

\[
s^2 = \sum_{i=1}^{n} \frac{e_i^2}{n-2}.
\] (32)

Then \( s \) can be given by

\[
s = \sqrt{\frac{ss_{yy} - \hat{b} \cdot ss_{xy}}{n-2}} = \sqrt{\frac{ss_{yy} - \hat{a}^2 \cdot ss_{xx}}{n-2}}
\] (33)


The standard errors for \( \alpha \) and \( b \) are

\[
\text{SE}(\alpha) = s \sqrt{\frac{1}{n} + \frac{s^2}{ss_{xx}}}
\] (34)

\[
\text{SE}(b) = \frac{s}{\sqrt{ss_{xx}}},
\] (35)

REFERENCES:


Ledvij, N. "Curve Fitting Made Easy." Industrial Physicist 9, 24-27, Apr./May 2003.


