$$SS_{res} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$ (4 - residual sum of squares)

$$SS_{reg} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$ (5 - regression sum of squares)

Here, $y_i$ is the i-th Y data point, $\hat{y}_i$ is i-th predicted or model Y value, and $\bar{y}_i$ is the average of the Y data. When you do a line fit, the fit line is guaranteed to pass through the data average, and you can be sure that

SStot = SSres + SSreg (6)

Consequently, equations (1) and (2) are equivalent, with equation (2) emphasizing the interpretation of r-squared as "explained variance". That is, it is the fraction of the data variance that is "explained" or accounted for by the fit line. The r-squared value reported by Igor generally agrees with values from Excel and other applications.

Igor also reports the Pearson correlation coeffient via the automatically-created variable V_Pr. In the case of a line fit, V_Pr$^2$ = V_r2.

### R-Squared and Fits Through the Origin

The situation changes if you fit a line that is constrained to pass through the origin, that is, you set the value of the fit coefficient a to zero and hold it. This is usually done when common sense or theory demands that the quantity of interest must be zero at X=0.

There is no consensus on the correct way to compute something like r-squared for a line fit through the origin. Such a fit does not, in general, pass through the average data value, and equation (6) does not hold, so equations (1) and (2) do not give the same value.

We prefer the interpretation of r-squared as "explained variance" so we use equation (2) for fits through the origin. This has some consequences. One is that Igor's r-squared will not necessarily agree with other applications, notably Excel. Another is that it is guaranteed to be smaller, some would say, "less good", than a line fit to the same data that is not constrained to pass through the origin.

A more surprising consequence is that it is possible for r-squared to be larger than 1. This has a simple interpretation: your fit line has larger variance than the data! Quite possibly a fit through the origin is simply not justified for your data.

Given the lack of consensus and uncertain interpretation of r-squared for a fit through the origin, we cannot recommend citing it, or using it as any sort of indication of goodness of fit, for such fits.

## Estimates of Error

Igor automatically calculates the estimated error (standard deviation) for each of the coefficients in a curve fit. When you perform a curve fit, it creates a wave called W_sigma. Each point of W_sigma is set to the estimated error of the corresponding coefficients in the fit. The estimated errors are also indicated in the history area, along with the other results from the fit. If you don’t provide a weighting wave, the sigma values are estimated from the residuals. This implicitly assumes that the errors are normally distributed with zero mean and constant variance and that the fit function is a good description of the data.

The coefficients and their sigma values are estimates (usually remarkably good estimates) of what you would get if you performed the same fit an infinite number of times on the same underlying data (but with different noise each time) and then calculated the mean and standard deviation for each coefficient.