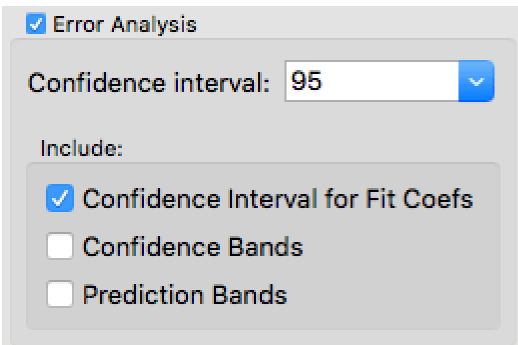
Confidence Bands and Coefficient Confidence Intervals

You can graphically display the uncertainty of a model fit to data by adding confidence bands or prediction bands to your graph. These are curves that show the region within which your model or measured data are expected to fall with a certain level of probability. A confidence band shows the region within which the model is expected to fall while a prediction band shows the region within which random samples from that model plus random errors are expected to fall.

You can also calculate a confidence interval for the fit coefficients. A confidence interval estimates the interval within which the real coefficient will fall with a certain probability.

Note: Confidence and prediction bands are not available for multivariate curve fits.

You control the display of confidence and prediction bands and the calculation of coefficient confidence intervals using the Error Analysis section of the Output Options tab of the Curve Fitting dialog:



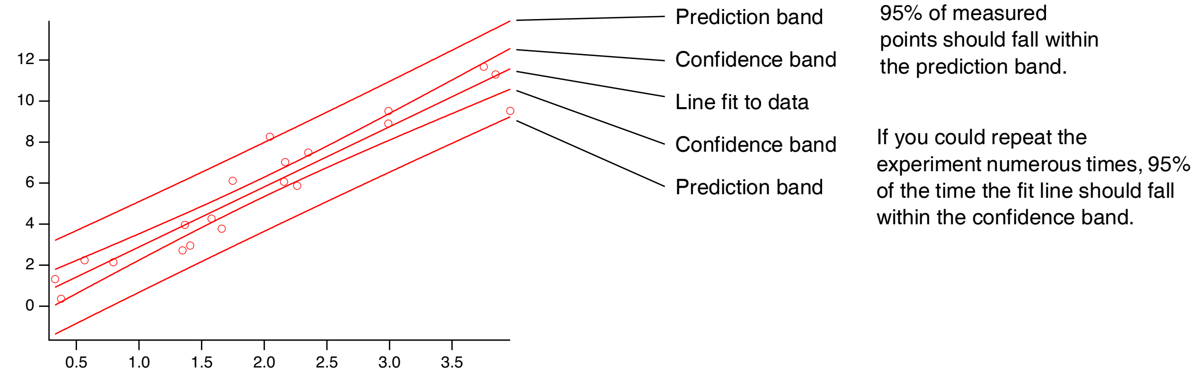
Using the line fit example at the beginning of this chapter (see **A Simple Case — Fitting to a Built-In Function: Line Fit** on page III-182), we set the confidence level to 95% and selected all three error analysis options to generate this output and graph:

Dialog has added the /F with parameters to select error analysis options.

```
•CurveFit line LineYData /X=LineXData /D /F={0.950000, 7}
   fit_LineYData= W_coef[0]+W_coef[1]*x
   W_coef={-0.037971,2.9298}
   V_chisq= 18.25; V_npnts= 20; V_numNaNs= 0; V_numINFs= 0;
   V_startRow= 0; V_endRow= 19; V_q= 1; V_Rab= -0.879789;
   V_Pr= 0.956769;V_r2= 0.915408;
   W_sigma={0.474,0.21}
   Fit coefficient confidence intervals at 95.00% confidence level:
   W_ParamConfidenceInterval={0.995,0.441,0.95}
   Coefficient values ± 95.00% Confidence Interval
      a   =   -0.037971 ± 0.995
      b   =   2.9298 ± 0.441
```

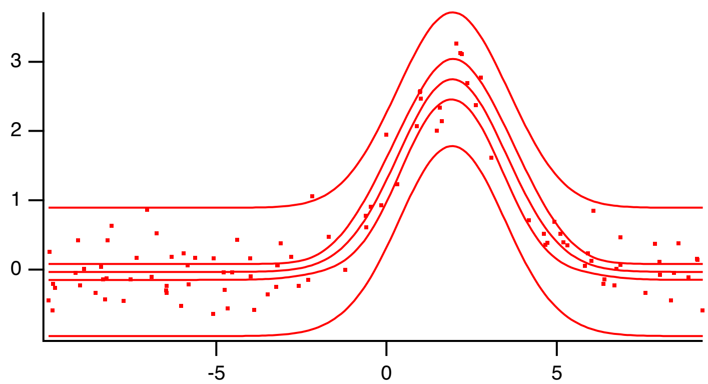
When confidence intervals are available they are listed here instead of the standard deviation.

Coefficient confidence intervals are stored in the wave W_ParamConfidenceInterval. Note that the last point in the wave contains the confidence level used in the calculation.



You can do this with nonlinear functions also, but be aware that it is only an approximation for nonlinear functions:

```
Make/O/N=100 GDataX, GDataY                        // waves for data
GDataX = enoise(10)                                // Random X values
GDataY = 3*exp(-((GDataX-2)/2)^2) + gnoise(0.3)    // Gaussian plus noise
Display GDataY vs GDataX                           // graph the data
ModifyGraph mode=2,lsize=2                         // as dots
CurveFit Gauss GDataY /X=GDataX /D/F={.99, 3}
```



The dialog supports only automatically-generated waves for confidence bands. The CurveFit and FuncFit operations support several other options including an error bar-style display. See **CurveFit** on page V-124 for details.

Calculating Confidence Intervals After the Fit

You can use the values from the W_sigma wave to calculate a confidence interval for the fit coefficients after the fit is finished. Use the StudentT function to do this. The following information was printed into the history following a curve fit:

```
Fit converged properly
fit_junk= f(coeffs,x)
coeffs={4.3039,1.9014}
V_chisq= 101695; V_npnts= 128; V_numNaNs= 0; V_numINFs= 0;
W_sigma={4.99,0.0679}
```

To calculate the 95 per cent confidence interval for fit coefficients and deposit the values into another wave, you could execute the following lines:

```
Duplicate W_sigma, ConfInterval
ConfInterval = W_sigma[p]*StudentT(0.95, V_npnts-numpnts(coeffs))
```

Naturally, you could simply type "126" instead of "V_npnts-numpnts(coeffs)", but as written the line will work unaltered for any fit. When we did this following the fit in the example, these were the results:

```
ConfInterval = {9.86734,0.134469}
```

Clearly, coeffs[0] is not significantly different from zero.

Confidence Band Waves

New waves containing values required to display confidence and prediction bands are created by the curve fit if you have selected these options. The number of waves and the names depend on which options are selected and the style of display. For a contour band, such as shown above, there are two waves: one for the upper contour and one for the lower contour. Only one wave is required to display error bars. For details, see the **CurveFit** operation on page V-124.

Some Statistics

Calculation of the confidence and prediction bands involve some statistical assumptions. First, of course, the measurement errors are assumed to be normally distributed. Departures from normality usually have to be fairly substantial to cause real problems.

If you don't supply a weighting wave, the distribution of errors is estimated from the residuals. In making this estimate, the distribution is assumed to be not only normal, but also uniform with mean of zero. That is, the error distribution is centered on the model and the standard deviation of the errors is the same for all values of the independent variable. The assumption of zero mean requires that the model be correct; that is, it assumes that the measured data truly represent the model plus random normal errors.

Some data sets are not well characterized by the assumption that the errors are uniform. In that case, you should specify the error distribution by supplying a weighting wave (see **Weighting** on page III-199). If you do this, your error estimates are used for determining the uncertainties in the fit coefficients, and, therefore, also in calculating the confidence band.

The confidence band relies only on the model coefficients and the estimated uncertainties in the coefficients, and will always be calculated taking into account error estimates provided by a weighting wave. The prediction band, on the other hand, also depends on the distribution of measurement errors at each point. These errors are not taken into account, however, and only the uniform measurement error estimated from the residuals are used.

The calculation of the confidence and prediction bands is based on an estimate of the variance of a predicted model value:

$$V(\hat{Y}) = a^T C a$$
$$a = \delta F / \delta p|_x$$

Here, $\hat{Y}$ is the predicted value of the model at a given value of the independent variable X, $\boldsymbol{a}$ is the vector of partial derivatives of the model with respect to the coefficients evaluated at the given value of the independent variable, and $\boldsymbol{C}$ is the covariance matrix. Often you see the $\boldsymbol{a^T C a}$ term multiplied by σ^2, the sample variance, but this is included in the covariance matrix. The confidence interval and prediction interval are calculated as:

$$CI = t(n-p, 1-\alpha/2)[V(\hat{Y})]^{1/2} \text{ and } PI = t(n-p, 1-\alpha/2)[\sigma^2 + V(\hat{Y})]^{1/2}.$$

The quantities calculated by these equations are the magnitudes of the intervals. These are the values used for error bars. These values are added to the model values ($\hat{Y}$) to generate the waves used to display the bands as contours. The function $t(n-p, 1-\alpha/2)$ is the point on a Student's t distribution having probability $1-\alpha/2$, and σ^2 is the sample variance. In the calculation of the prediction interval, the value used for σ^2 is the uniform value estimated from the residuals. This is not correct if you have supplied a weighting wave with nonuniform values because there is no information on the correct values of the sample variance for arbitrary values of the independent variable. You can calculate the correct prediction interval using the **StudentT** function. You will need a value of the derivatives of your fitting function with respect to the