CM3450
Fall 2008
Drill 13

1. Do example 5.2.
2. Do example 5.2.3, page 73. Use **genfit( )** and also separately use **minerr( )**.

Linear and Nonlinear Regression in MathCad: Scalar Case
(Dr. Tom Co 10/11/2008)

Introduction.

Mathematical models (empirical or theoretical) need to fit with experimental data. However, experimental data often contain measurement noise and error. Thus, the models do not need to pass through all the data points. Instead, the models need to just be close to the data. One criteria is to use minimize the sum of squared residual errors.

Partial Glossary and Notation of Terms

1. **Experimental data**: $(x_1, y_1), \dots, (x_N, y_N)$ often given in tabular form, where x_k is the kth independent variable data and y_k is the kth dependent variable data.
2. **Model**: a function relating the dependent variable (y) with independent variable (x).
3. **Model parameters**: constants in a model that are adjusted to improve the fit of the model to experimental data.
4. **Regressed data:** $\hat{y}_k$ the value of the dependent variable evaluated at $x = x_k$.
   Example:
   a) Linear model: $y = mx + b$
      y is the dependent variable, x is the independent variable, m and b are model parameters.
      $\hat{y}_k = mx_k + b$ is the kth regressed data.
   b) Nonlinear model: $C = k_0 \exp\left(\frac{-E}{RT}\right)$
      C is the dependent variable, T is the independent variable, E and k_0 are model parameters. (Note: R is the universal gas constant, it is not a model parameter)
      $\hat{C}_k = k_0 \exp\left(\frac{-E}{RT_k}\right)$ is the kth regressed data.
5. **Residuals** (or **residual error**): the difference between the kth regressed data and the kth experimental data: $(\hat{y}_k - y_k)$
6. **Coefficient of Determination:** r^2 (closer to 1 means fit is better than simple mean)
$$r^2 = \frac{\left[\sum (y_k - \mu)^2\right] - \left[\sum (y_k - \hat{y}_k)^2\right]}{\left[\sum (y_k - \mu)^2\right]}$$
   where μ is the average of y_k.
7. **Correlation Coefficient:** r

Linear Regression

Model: $y = mx + b$
Model parameters: m is the slope and b is the y-intercept.

MathCad Procedure:

a) $m \coloneqq \mathbf{slope}(\, xdata,\ ydata\,)$
b) $b \coloneqq \mathbf{intercept}(\, xdata,\ ydata\,)$
c) $y(x) \coloneqq m \cdot x + b$

Polynomial Regression

Model: $y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$
Model parameters: $a_0, \ldots, a_n$ are the polynomial coefficients.

MathCad Procedure:

a) $coef \coloneqq \mathbf{regress}(\, xdata,\ ydata,\ order)$
b) $y(x) \coloneqq \mathbf{interp}(\, coef,\ xdata,\ ydata,\ x\,)$

Note:

- the coefficients are $a_0 = coef_3, \ldots, a_{order} = coef_{3+order}$
  (assuming vector $coef$ starts at index 0)
- does not support units

Nonlinear Regression

MathCad Procedure:

a) Define the nonlinear function: $ymodel(x, a, b, c, \ldots) = \cdots$
b) Set up initial guesses for model parameters:

$$InitialGuess \coloneqq \begin{pmatrix} a_{guess} \\ b_{guess} \\ c_{guess} \\ \vdots \end{pmatrix}$$

c) Use **genfit( )** to optimize parameters for best fit:

$$\begin{pmatrix} a \\ b \\ c \\ \vdots \end{pmatrix} = \mathbf{genfit}(xdata, ydata, InitialGuess, ymodel)$$

Correlation Coefficients

MathCad Functions:

$$r \coloneqq \mathbf{corr}(\ ydata,\quad \overrightarrow{f(xdata)}\,)$$

Note: the arrow above $f(xdata)$ is the vectorization symbol. This can be achieved by first selecting the group of terms to be "vectorized", then click **[CTRL minus]**.

Alternative Approach using Minerr( ) Function

$$ymodel(x, a, b) := \frac{a \cdot e^{-x}}{1 + b \cdot x}$$

$a := 1 \qquad b := 1$

Given

$$y = ymodel(x, a, b)$$

$$\begin{pmatrix} a \\ b \end{pmatrix} := minerr(a, b)$$

[Right Click] here,
then select **[Nonlinear]→ [Levenberg-Marquardt]**

$a = 33.226 \qquad b = 2.319$

Remarks:

1. This approach can use units
2. One can include add constraints inside the **Given - Minerr** block.
3. It can handle multivariable problems: multiple dependent and multiple independent variables.