

# A First Course on Kinetics and Reaction Engineering

## Example S3.1

### Problem Purpose

This example shows how to use the MATLAB script file, FitLinSR.m, to fit a linear model to experimental data.

### Problem Statement

Assume that in the course of solving a kinetics problem, you have reached the point where you have data from eleven experiments, and that the data have been worked up to give the eleven data points listed in Table 1. The table shows that there are three set variables,  $x_1$  through  $x_3$  and one response variable,  $\hat{y}$ . Further assume that in order to solve this problem, you need to determine whether the linear model given in equation (1) offers a good representation of these data. That is, you need to fit equation (1) to these data, decide whether the fit is acceptable, and if it is acceptable, report the best values for the parameters  $m_1$ ,  $m_2$ ,  $m_3$  and  $b$ .

Table 1

$x_1$	$x_2$	$x_3$	$\hat{y}$
0	5	20	12.27
1	4	19	11.15
2	3	18	9.77
3	2	17	8.88
4	1	16	8.21
5	0	15	7.05
6	1	14	15.46
7	2	12	26.0
8	3	10	36.9
9	4	6	49.4
10	5	4	58.1

$$y = m_1x_1 + m_2x_2 + m_3x_3 + b \quad (1)$$

### Problem Analysis

The model is a general linear equation, and it does have an intercept, so it can be fit to the data in its present form using the MATLAB script file FitLinSR.m that is provided with this supplemental unit.

### Problem Solution

The first step is to import or enter the data and store them in the MATLAB workspace in the form of a matrix,  $x$ , and a column vector,  $y\_hat$ . The first column in the matrix  $x$  should correspond to set variable  $x_1$ , the second column should correspond to set variable  $x_2$  and the third column to set variable  $x_3$ . There should be a fourth column in the matrix,  $x$ , that contains the value 1.0 in every row. The column vector,  $y\_hat$ , should contain the measured values of the response variables. The matrix,  $x$ , and the column vector,  $y\_hat$ , should each have 11 rows, with one row corresponding to each of the experimental data points, as shown in Listings 1 and 2.

$x =$

```

0  5  20  1
1  4  19  1
2  3  18  1
3  2  17  1
4  1  16  1
5  0  15  1
6  1  14  1
7  2  12  1
8  3  10  1
9  4   6  1
10 5   4  1

```

*Listing 1.*

$y\_hat =$

```

12.2700
11.1500
 9.7700
 8.8800
 8.2100
 7.0500
15.4600
26.0000
36.9000
49.4000
58.1000

```

*Listing 2.*

The MATLAB script file, FitLinSR.m, does not require any modification before it can be used. All that is required is that FitLinSR.m is stored somewhere in the MATLAB search path. If it is, then simply typing "FitLinSR" at the MATLAB command prompt will execute the script, producing the output shown in Listing 3.

From this you can see that the correlation coefficient,  $r^2$ , is equal to 0.9994 and the parameters have the following values and 95% confidence limits:  $\theta_1 = 2.88 \pm 0.99$ ,  $\theta_2 = 5.04 \pm 0.51$ ,  $\theta_3 = -1.12 \pm 0.64$  and  $\theta_4 = 9.29 \pm 15.0$ . The order that the parameters and uncertainties are listed is the same as the order of the columns in the matrix  $x$  that was created earlier.

In addition to this output, MATLAB also creates and displays four plots (they may be positioned one on top of the other, requiring you to drag them to new positions in order to see all of them). One is a parity plot and the other three are residuals plots in which the residuals are plotted separately against each of the set variables.

```
>> FitLinSR  
  
r_squared =  
  9.9942e-01  
  
m =  
  2.8836e+00  
  5.0418e+00  
 -1.1181e+00  
  
m_u =  
  9.9147e-01  
  5.0940e-01  
  6.4459e-01  
  
b =  
  9.2882e+00  
  
b_u =
```

*Listing 3.*

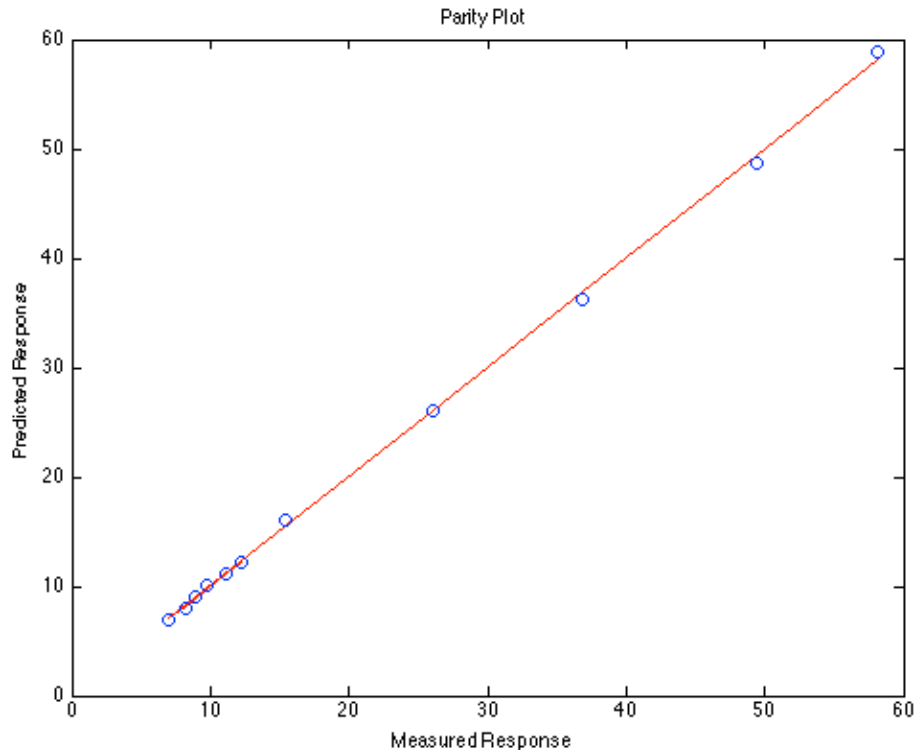


Figure 1. Parity plot for the fit of a linear model to the data in Table 1.

The correlation coefficient is nearly equal to 1.0, so the fit appears to be very good, but we still should examine the parity plot, Figure 1, and the residuals plots, Figures 2, 3 and 4. As expected given the good correlation coefficient, the parity plot shows all of the data points to be very close to the diagonal line. Occasionally a parity plot will show systematic deviations (this one does not), but often it will not, even if there are systematic trends in the errors with respect to one of the set variables. For this reason, the residuals plots should always be checked even if the parity plot and correlation coefficient indicate a good fit.

In this case, the residuals plots on the following pages show random scatter of the residuals with respect to each of the set variables. On first impression, the *magnitude* of the scatter in the residuals plots looks very large, but that is only because the vertical scale being used is very small. The scatter seen in these figures is the same as the scatter seen in Figure 1. The important feature of the residuals plots is that they do not indicate any systematic trend in the residuals as a function of any of the set variables. Overall, the correlation coefficient, parity plot and residuals plots all indicate that the linear model offers a very good representation of the experimental data. Therefore the model is acceptable with the parameter values and uncertainties given earlier.

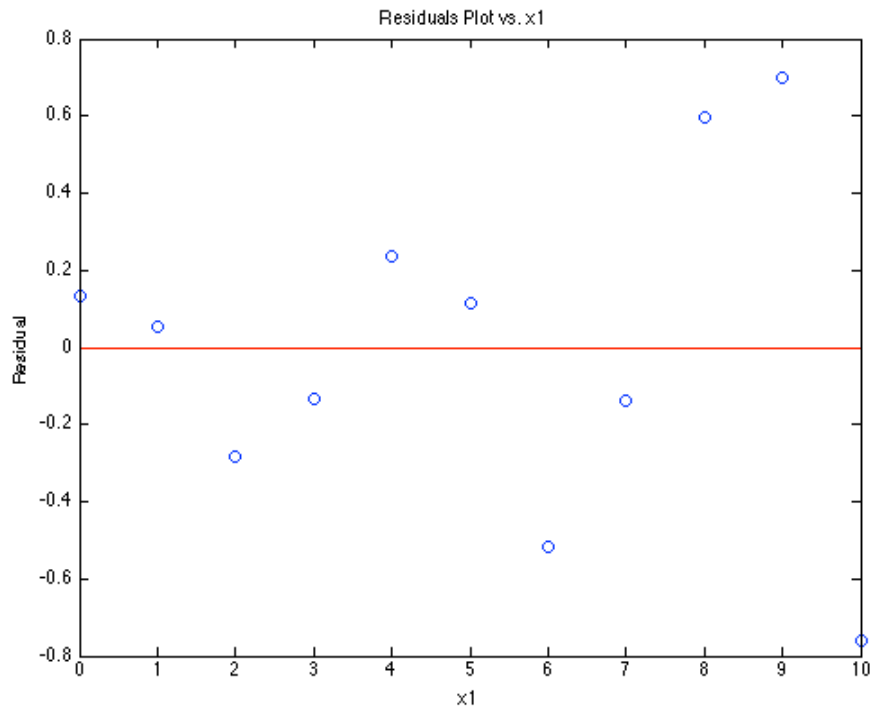


Figure 2. Residuals plotted against set variable  $x_1$ .

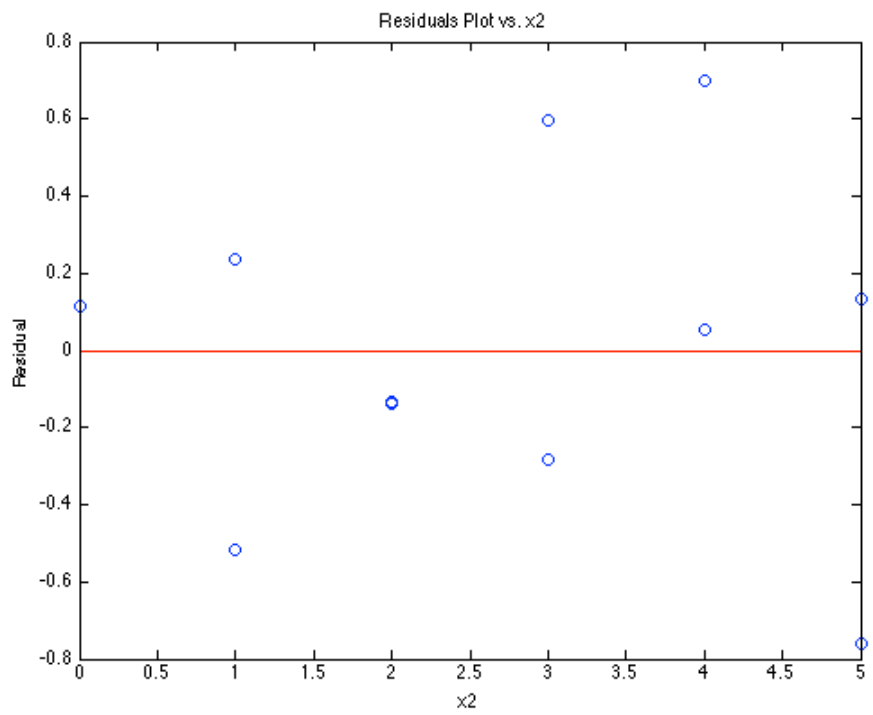


Figure 3. Residuals plotted against set variable  $x_2$ .

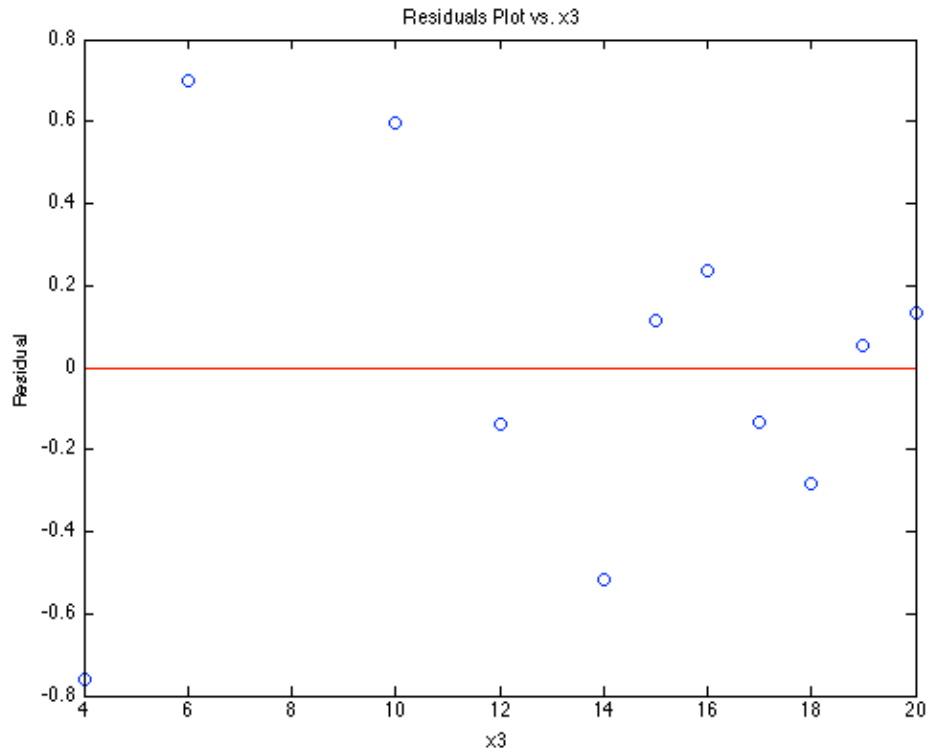


Figure 4. Residuals plotted against set variable  $x_3$ .