

## Least-Squares Fitting – Doing it the "Hard Way" With Excel

*Purpose: To ensure that you know what is involved in fitting a simple function to a set of data. To master the basic data-fitting and plotting functions available with Excel.*

Attached to the end of this assignment is a series of data (X v. Y) with realistic scatter (noise). We will test the hypothesis that they are directly correlated, that is that a simple linear function ( $y=mx+b$ ) fits the data and can be used to interpolate between the data points.

Follow the basic instructions below. As always, you are free (encouraged) to go further than strictly required. Print out your spreadsheet(s) and plot(s) and leave them in my mailbox if I'm not around. Email your Excel workbook to me. This is due before 7 November, but don't put it off. If you have questions, email me.

- 1) Read these data into an Excel spreadsheet. Make sure it gets saved as an excel workbook. Your first 2 columns should be "x\_data" and "y\_data".
- 2) Sort your rows in order of increasing X. Use the Excel charting function to plot your x v. y data points. This should be a "scatter" plot. Do not connect the points with any kind of line.
- 3) Generate an initial guess for a linear fit. Use  $m_0 = (y_{\max} - y_{\min}) / (x_{\max} - x_{\min})$  and  $b_0 = x_{\min} - b_0 * x_{\min}$  as your initial guesses. Enter these values somewhere on the spreadsheet.
- 4) Generate a new "y\_fit" column using your x\_data values and y calculated from a linear fit.
- 5) Add a line (no data points, just a line) to your chart by plotting x\_data v. y\_fit.
- 6) Generate 2 more columns. The first is  $y_{\text{fit}} - y_{\text{data}}$ , and the second is  $(y_{\text{fit}} - y_{\text{data}})^2$ . At the bottom of each, calculate the sum for each column.
- 7) Now the fitting begins. You need to change your initial guesses, generate a new fit, and keep going until you have a "best fit", which is usually described by the fit that minimizes the sum of the square of the residuals.

- 8) In another spreadsheet, or somewhere else on this spreadsheet, record your initial guesses for  $m$  and  $b$ , the sum of the residuals, and the sum of the squares of the residuals. Repeat this procedure 7 to 10 times. Try to select your guesses so that you quickly zero in on the "right answer".
- 9) Overplot your final best as a line on the data (if you set your spreadsheet up right, this is already done). Annotate your plot with the equation of your best fit line.
- 10) Right click on a data point and select "fit a trendline". Use a linear function, a thick solid line, and select the option to print the trendline parameters on your plot.

X	Y
1.20	1.42
1.22	1.72
1.23	2.20
2.61	2.50
2.50	2.78
2.34	3.00
2.68	3.02
2.85	3.57
2.19	3.60
2.31	3.79
2.31	3.85
2.86	4.36
3.22	4.95
3.30	5.11
3.92	5.13
2.96	5.17
3.20	5.29
3.18	5.55