Drawing the Best Fit Line

Showing linear dependence is a useful trick. How is it actually done, though? The best possible linear fit line to a data set would go through the data points and have as many points above it as below it and would also take into account the error bars. There is a parallel in methodology to finding standard deviation. A small standard deviation indicates on average small differences (deviations) from the mean.

Consider the distance of each point from a linear fit. For a "best fit" line those deviations should be as small as possible. If the data actually fell smack dab on a straight line, then the best fit line would be the one that makes the deviations zero! Finding the best fit line involves "eye-balling" where the line should fall in the data. It is a skill that improves with practice (like so many!).

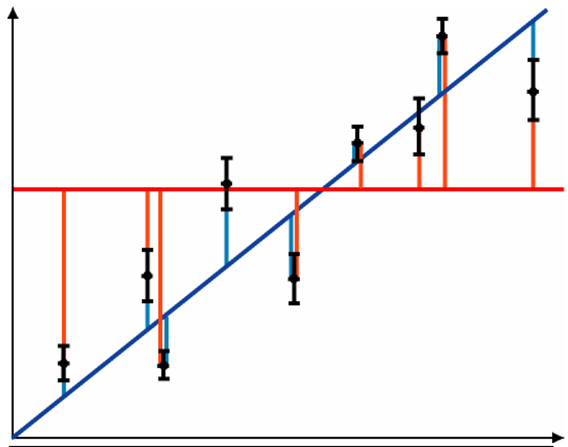

Figure 1: Example of deviations from the best fit line (blue) and from the mean (red).

Look at **Figure 1**. The horizontal red line represents the mean of the dependent variable. The diagonal blue line represents the best fit line of the formula $y = A + Bx$. The vertical red lines represent the distance from the data to the mean, while the vertical blue lines represent the distance to the best fit line. Notice that the distances to the best fit line are smaller than the distances to the mean. In this case the best fit line is a better representation of the relation of the independent and dependent variables than the mean of the dependent variable.

Uncertainty in the Slope of the Best Fit Line

What about finding the uncertainty in the best fit line? Two lines are used to find the least possible slope of the best fit line within error and the greatest possible slope of the best fit line within error.

How should the lines be drawn?

- **Lower Bound of the Slope (green line)**: The first line should go through the bottoms of the error bars on the right half of the graph and through the tops of the error bars on the left half of the graph. Notice that the slope of the drawn line is less steep than the slope of the best fit line. Is an estimate of the lowest (minimum) slope allowed by the error in the data.

- **Upper Bound of the Slope (purple line)**: The second line should go through the tops of the error bars on the right half of the graph through the bottoms of the error bars on the left half of the graph. Notice that this time the slope of this drawn line is steeper than the slope of the best fit line**.** It is an estimate of the highest (maximum) slope consistent with the error in the data.

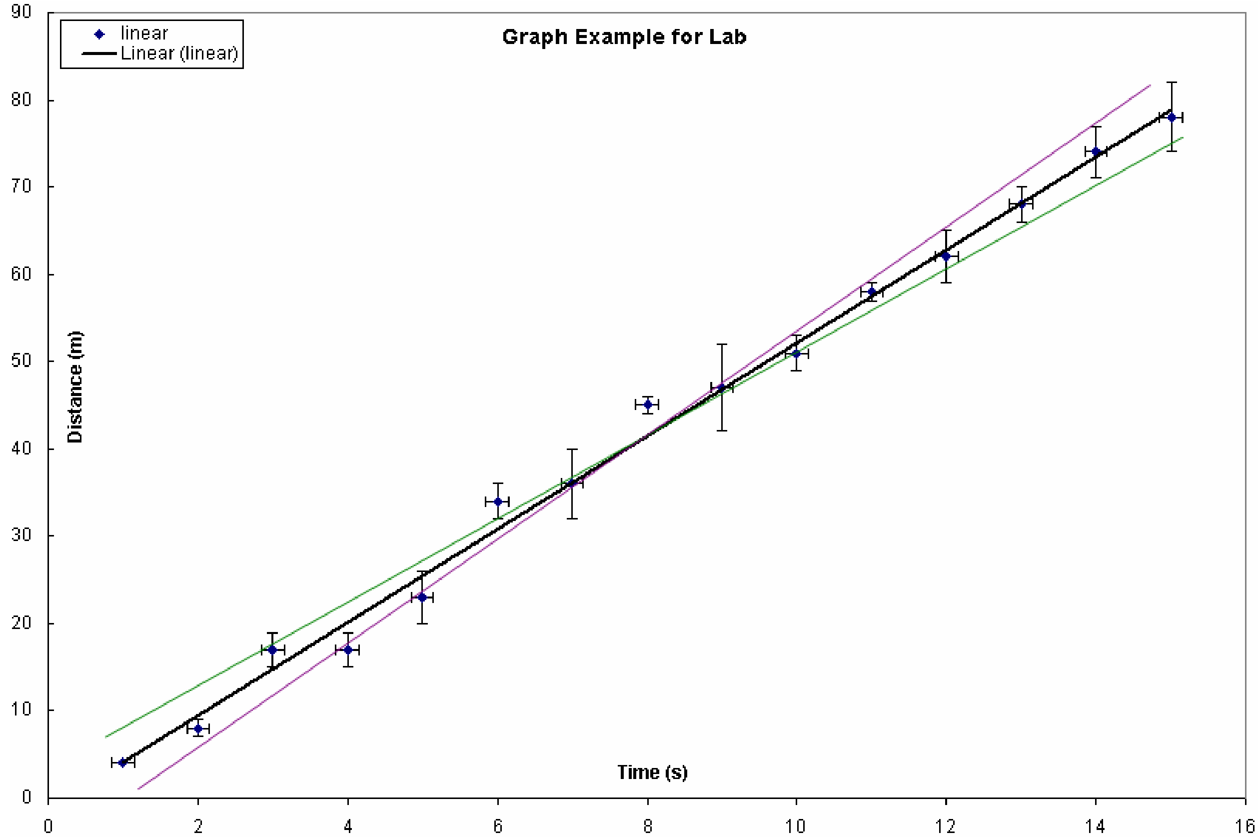

Figure 2: Graph with best fit line (black) and lines estimating lower bound of the slope (green) and the upper bound of the slope (purple) --- both within the given error.

Once the lines are drawn both of their slopes must be estimated. To figure out the slope of a line, get the coordinates of two points on the line and use the definition of the slope as “rise over run” to calculate it:

$$slope = \frac{\Delta y}{\Delta x} = \frac{(y_2 - y_1)}{(x_2 - x_1)}$$

The two slopes from the lines are the upper and lower bounds on the slope of the best fit line consistent with error. The difference of the upper and lower bound slopes is actually twice the error of the slope.

$$slope\ \ error\ \ for\ \ linear\ \ trendline = \frac{upper\ bound\ slope - lower\ bound\ slope}{2}$$

For example, two points on the lower bound of the slope are (1.0, 7.0) and (3.5, 19) and two points on the upper bound of the slope are (6.5, 32) and (9.5, 49).

$$upper\ bound\ slope = \frac{(y_2 - y_1)}{(x_2 - x_1)} = \frac{(19m - 7.0m)}{(3.5s - 1.0s)} = \frac{12m}{2.5s} = 4.8\,{}^{m}\!/_{s}$$

$$lower\ bound\ slope = \frac{(y_2 - y_1)}{(x_2 - x_1)} = \frac{(49m - 32m)}{(9.5s - 6.5s)} = \frac{17m}{3.0s} = 5.6\,{}^{m}\!/_{s}$$

The lower bound of the slope within error is 4.8 m/s and the upper bound of the slope within error is 5.6 m/s.

$$slope\ error\ for\ linear\ trendline = \frac{upper\ bound\ slope - lower\ bound\ slope}{2} = \frac{5.6 - 4.8}{2} = \frac{0.80}{2} = 0.40\,{}^{m}\!/_{s}$$

The slope of the best fit line is 5.1929. Therefore, the slope of the best fit line with error is 5.2 ± 0.40 m/s. Why isn't it 5.1929 ± 0.40 m/s? Follow the rules in the sections at the end of the lab manual for reporting error. If there is an error of 0.40 m/s, it makes 5.1929 m/s more accurate than error allows --- a bit silly!