**How to Fit the Data to a Least Squares Regression Line**

Without getting into too much technical detail, the least squares regression line is the straight line that "best fits" the data. The metric for determining best fit is to measure the square of the distance from each point to the best fit line, sum those squares, and then select the line that minimizes that sum. Fortunately, this has been studied and converted to formulas.

The best fit line is expressed in the form $y = B_1x + B_0$, where formulas exist for determining $B_1$ and $B_0$.

$$B_1 = \frac{SS_{xy}}{SS_{xx}} \quad \text{where} \quad SS_{xy} = \Sigma xy - (1/n)(\Sigma x)(\Sigma y) \quad \text{and} \quad SS_{xx} = \Sigma x^2 - (1/n)(\Sigma x)^2$$

The summation sign Σ refers to summing up each value of x or y, as appropriate. In our case we have n = 3 points with (x,y) values (2, 469), (3, 615), and (4, 1,058). The math is shown below:

$SS_{xy}$ = (2)(469) + (3)(615) + (4)(1058) - (1/3)(2 + 3 + 4)(469 + 615 + 1,058) =
= 938 + 1,845 + 4,232 - (1/3)(9)(2,142) = 7,015 - 6,426 = 589

$SS_{xx} = 2^2 + 3^2 + 4^2 - (1/3)(2 + 3 + 4)^2 = 4 + 9 + 16 - (1/3)(9^2) = 29 - 27 = 2$

Thus $B_1$ = 589 / 2 = 294.5

The formula for $B_0$ is: $B_0 = (1/n)(\Sigma y) - B_1(1/n)(\Sigma x)$

Thus $B_0$ = (1/3)(469 + 615 + 1,058) - (294.5)(1/3)(2 + 3 + 4) =
= 714 - (294.5)(3) = -169.5

Thus the least squares regression line is: y = 294.5x - 169.5

The $B_1$ and $B_0$ parameters can also be determined in Excel using the SLOPE and INTERCEPT functions. In each case the functions are applied to a data range of y values, followed by a comma and the data range of x values.

Press the back button to return to the blog entry. If you press the x, you'll have to start over.