SAMLab Tip Sheet #8

The Linear Regression Equation and the Standard Error of the Estimate

This Tip Sheet provides an overview of how to create a scatterplot for bivariate data and use it to find the equation for the linear regression line based on the least-squares criterion. We will use this equation to compute the standard error of the estimate of y from x ($S_{y.x}$). To begin, input the data below (from Tip Sheet #7) and select "Chart" from the Insert menu as shown.
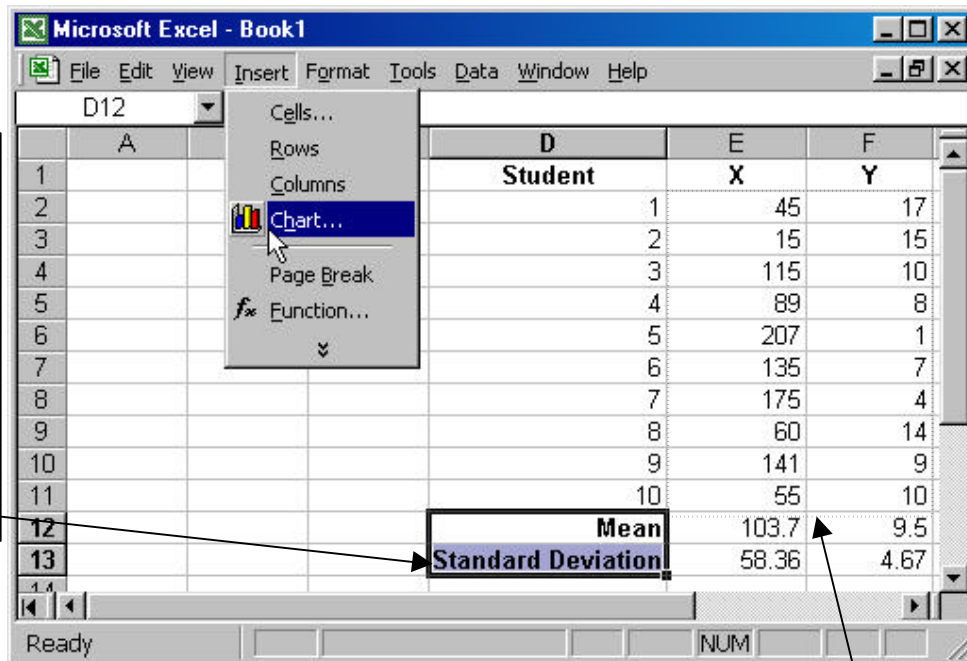
The descriptive standard deviation is used here (**STDEVP**). For more information on standard deviations, see Tip Sheet #6.


The data (fictitious) in this example represent the number of college credits a student has earned (**X**) and the student's score on a metric of test anxiety (**Y**) administered before the final exam period.

In the Chart Wizard window shown below left select "XY (Scatter)" in the Chart type field. Next click on the Chart sub-type that has no lines connecting the data points. Click Next > and highlight the Data range field as shown on the lower right, making sure the Data Range tab (circled) is selected. Either select the cells containing your data (to which the arrow points) or manually type in the cell references (Tip Sheet #2). Make sure the radio button next to "Columns" (circled) is selected. Click Next > to continue.
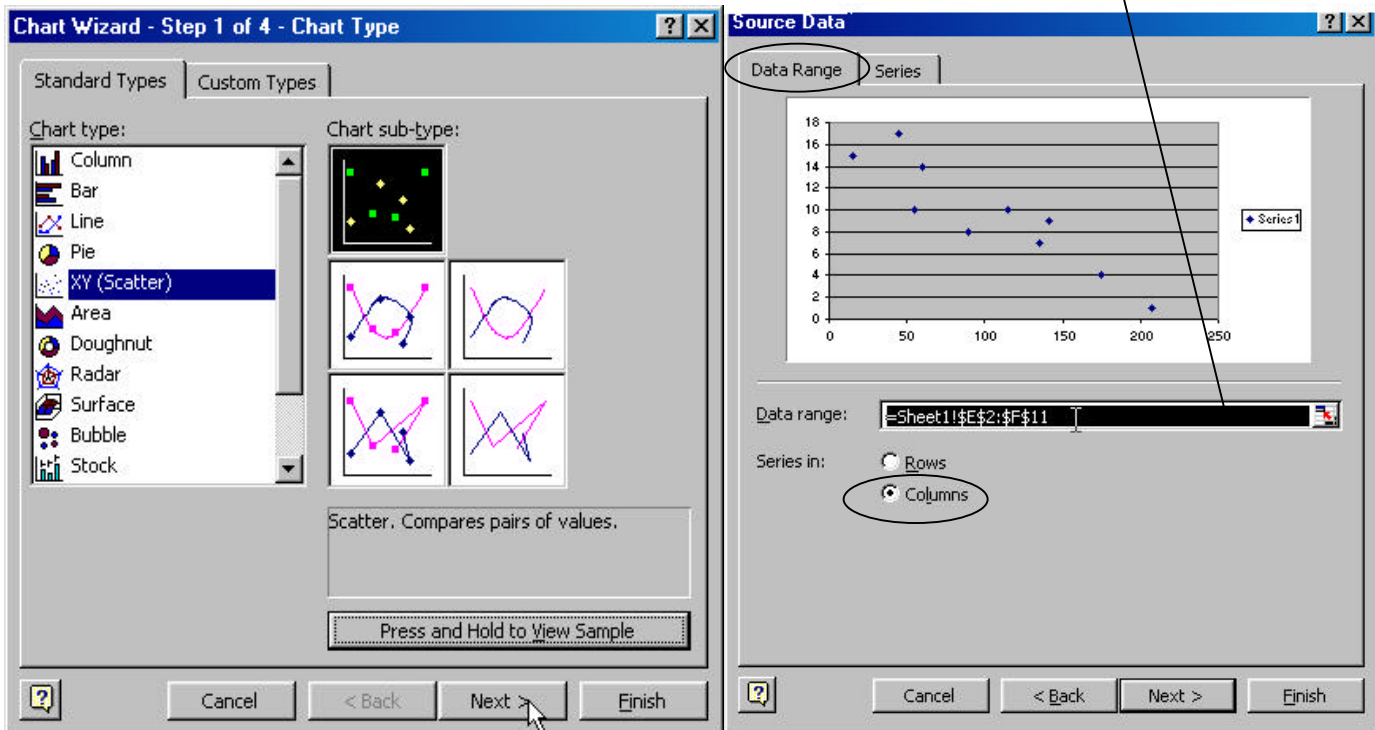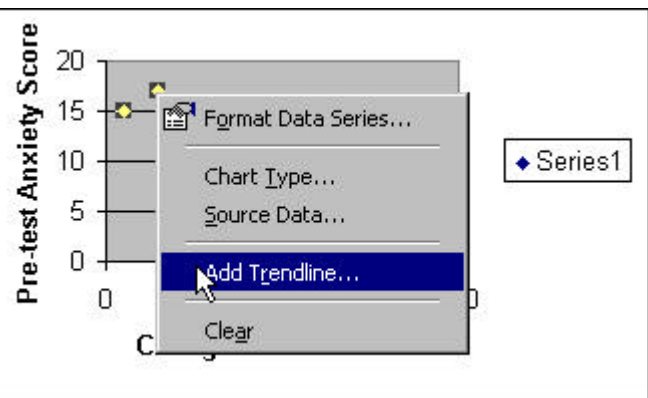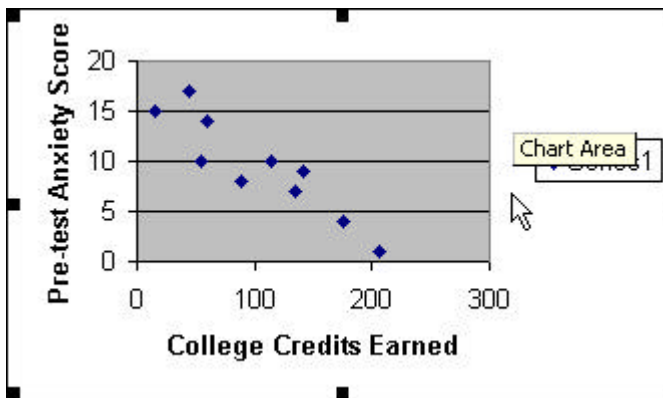
SAMLab Tip Sheet #8

Give names to your axes as shown on the left below and click Next >. Click the radio button next to "As object in," select "Sheet 1" or wherever you want your chart, and click Finish.
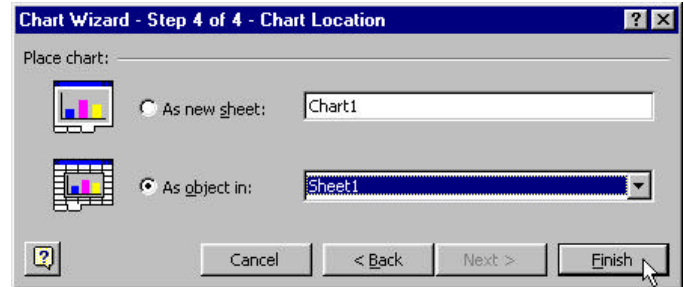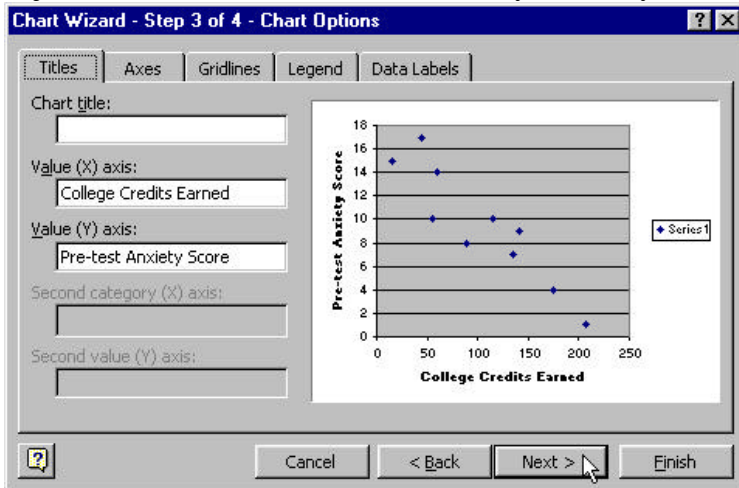
You should now have a graph like the one on the left below. Right click on one of the data points as shown on the right and select "Add Trendline…"


When you see the screen below left, select "Linear" for Trend/Regression type. Next, click the "Options" tab and in the window shown below right check the box next to "Display equation on chart." Click OK.
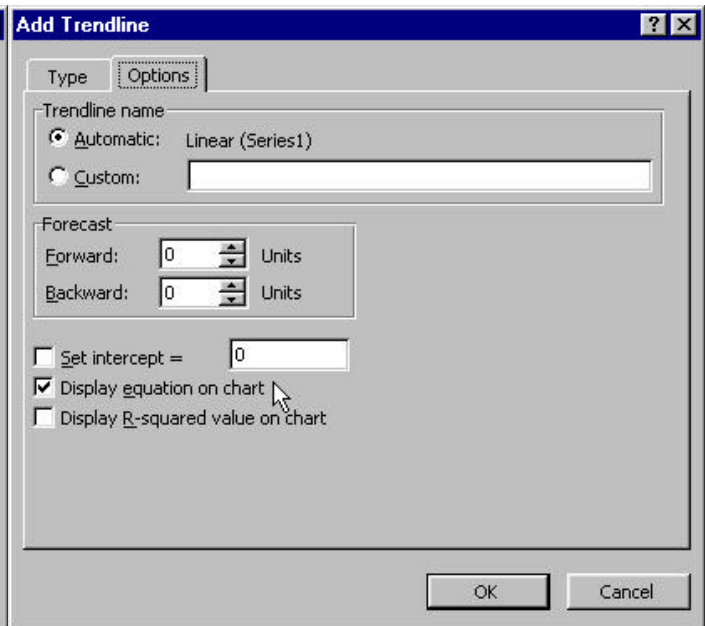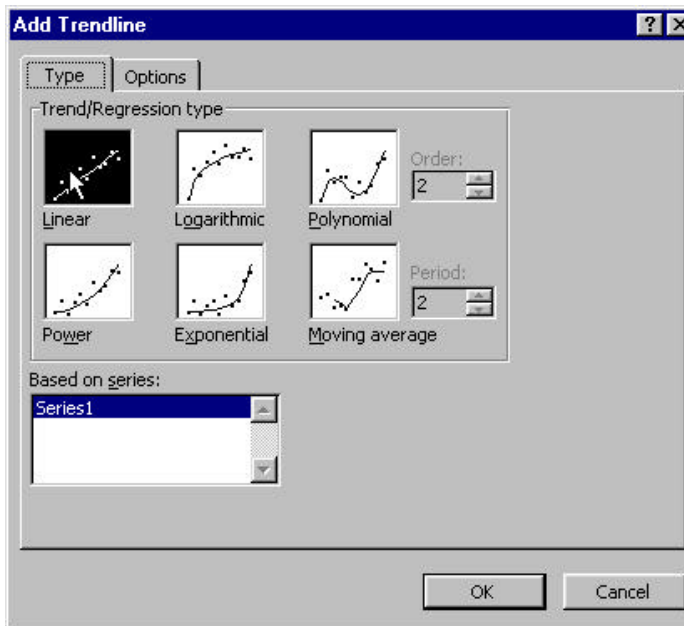
## Calculating the Standard Error of the Estimate of Y from X

You should now see a regression equation on the your graph. If you cannot read it easily, click and drag it to a better place. Label the column next to your "**Y**" scores "**Y est.**" and use the regression equation from your graph to compute the estimated y-values from the x-values. The equation is shown in the formula bar.
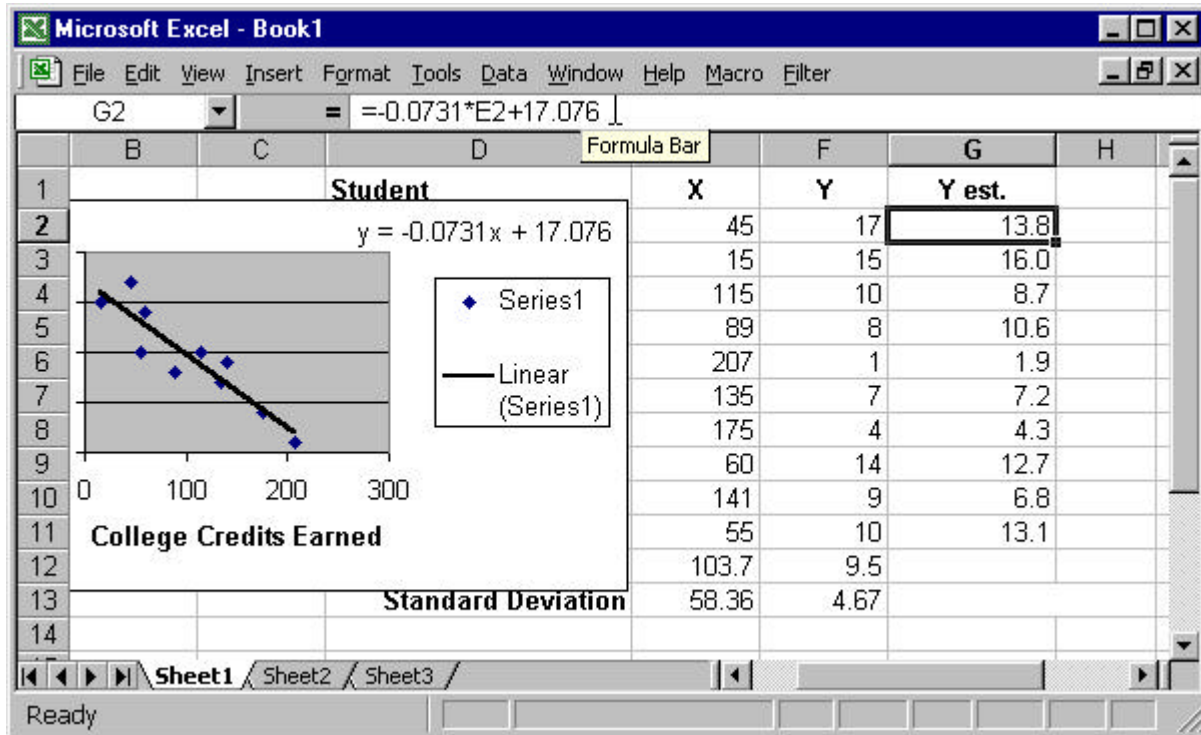
Next, create a column for difference scores, subtracting estimated y-values from actual y-values. Create another column containing the squares of these difference scores and find their sum. Divide the sum by N (10 in this case) and take the square root to obtain $S_{y \cdot x}$.

he equation for the
andard error of the
timate is:

$$S_{y \cdot x} = \sqrt{\frac{\sum (Y - Y')^2}{N}}$$
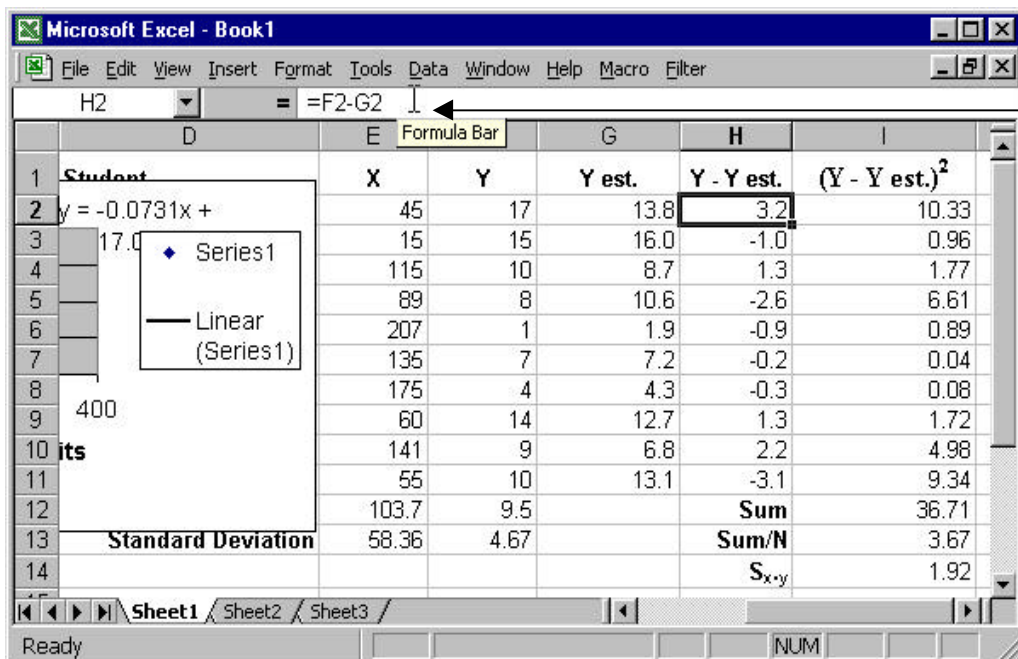
Note that the difference scores here do not require a static cell reference. Also, Excel does have a paste function for $S_{y \cdot x}$, but it is an inferential version, which divides the sum of squares by N-2.