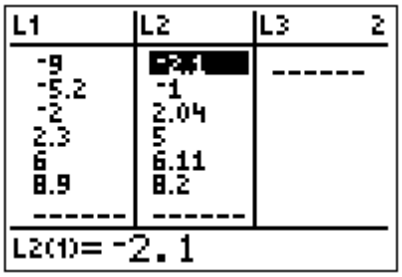
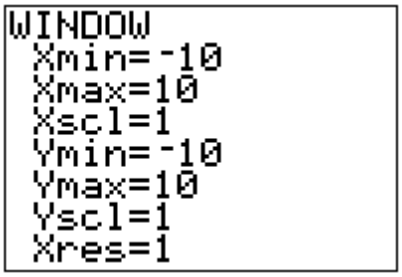
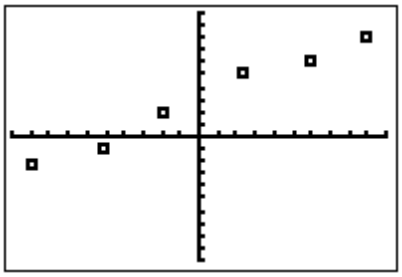
Calculator Appendix I

(Linear Regression)

The process of regression is finding a function that "best-fits" a scatter-plot. In this Appendix a **linear regression** that determines the best-fit of a line will be demonstrated. Shown below is the data and the resulting scatter-plot from **Appendix H** that will be used here. That Appendix shows how to enter the data and produce the scatter-plot.

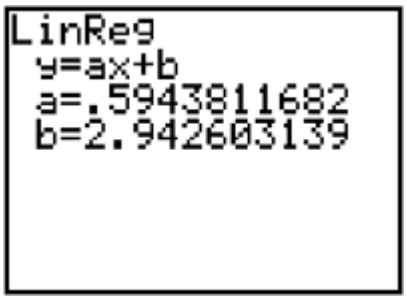






With the above data already in the lists **L1** (the *x* values) and **L2** (the *y* values), press the **STAT** button. With the **RIGHT ARROW** move to **CALC** and then with the **DOWN ARROW** down to **4: LineReg(ax+b)**.

Press **ENTER** twice to produce the *a* and *b* values defining the best-fit line:

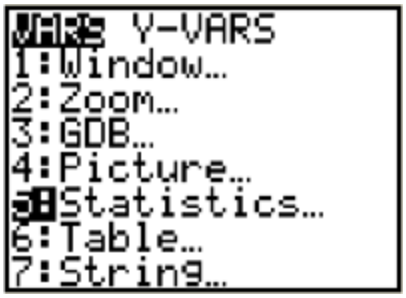


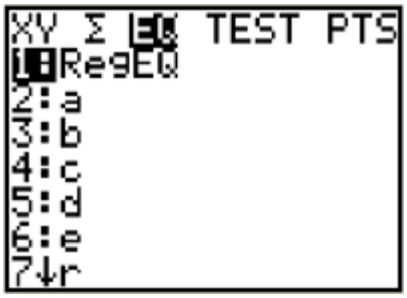
At this point we could press **Y=** and enter:

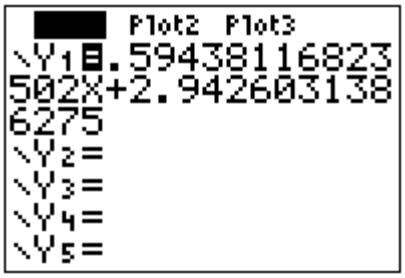
Y1 = .5943811682X+2.942603139

However, there is an easier way. Rather than having to manually enter the *a* and *b* values for **Y1**, it is possible to automatically have them entered.

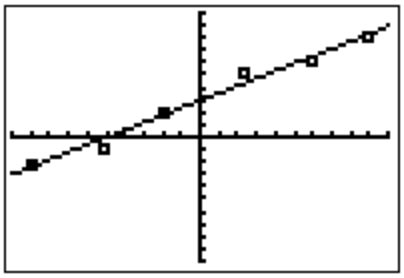
With the cursor on **Y1=** (on the **Y=** screen), press the **VARS** button, choose **5: Statistics**, **RIGHT ARROW** across to **EQ**, and then down to **1: RegEQ**. At this point the regression equation for the best-fit is automatically entered for **Y1**.







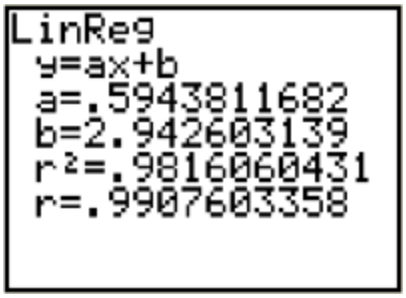
Press **GRAPH** and if **STAT PLOT** is still enabled (see **Appendix H**), the following should display as the best-fit for for the scatter-plot.



Some data does not lend itself to being best-fit with a line. For data that obviously "curves", other type regressions are available (parabolic (QuadReg), cubic, exponential, etc.)

When doing various regressions on a set of data points, it is possible to get "scores" on the quality of each fit.

First, it will be necessary to turn on "diagnostics" so that a score can be displayed. Press **2ND| CATALOG**, **Down ARROW** down to **1: Diagnostics ON**, and press **ENTER** twice to turn on diagnostics.



With diagnostics on, a regression will now display r^2 = "some number less than or equal to one". Shown here is the same linear regression as above except now r^2 = .9816060431 is displayed along with a corresponding *r* value. r^2 is the square of the **correlation factor**, *r*, which can range from -1 to 1.

To eliminate the sometimes confusing negative *r* values, r^2 is often given as a figure of merit, were of, of course, $r^2 = 1$ is a **perfect score** as a result of a perfect fit.

For a linear regression, a negative *r* value indicates that the line of best-fit has a negative slope and the **correlation is said to be negative**. Likewise, a best-fit line with a positive slope is represented with a positive value of *r* and the **correlation is side to be positive**.