CHAPTER 3 RELATIONSHIPS BETWEEN TWO QUANTITATIVE VARIABLES

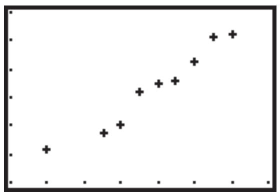
Section 3.1 ▪ Scatterplots

Setup

The TI-83 and TI-83 Plus can display scatterplots of bivariate data contained in any two lists, say, lists L1 and L2. This particular example uses the mean net income for family practitioners from Display 3.20 on page 119 of the student text.

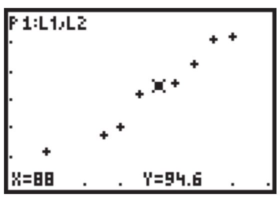
Go to the Stat Plot screen, 2nd [**STAT PLOT**], and define a scatterplot. Scatterplots are the first option under **Type**.

Press ZOOM **9:ZoomStat** to fill the Graph window.



[80, 94, 2, 60, 120, 10]

When you trace the scatterplot, you see the coordinates of the data points.



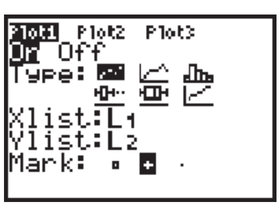
[80, 94, 2, 60, 120, 10]

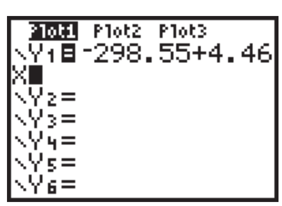
It is important that you keep straight which is the independent variable (**Xlist**) and which is the dependent variable (**Ylist**).

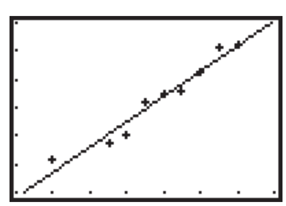
Section 3.2 ▪ Getting a Line on the Pattern

Lines as Summaries

In order to graph a line on a scatterplot, you must separately define the scatterplot and enter the equation of the line into one of the functions on the Y= screen. This particular example corresponds to the family practitioners' income example on pages 119–120 of the student text.







[80, 94, 2, 60, 120, 10]

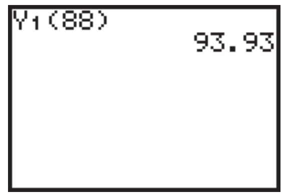
(continued)

Using Lines for Prediction

Interpolation and Extrapolation

Once an equation is entered into the Y= screen, you can evaluate the function at any value of the independent variable on the Home screen. You access the function names by pressing [VAR] **Y-VARS 1:Function.**

For example, with **Y_1= −298.55 + 4.46X** defined, you can predict the *y*-coordinate when $x = 88$, by entering **Y_1(88)**.

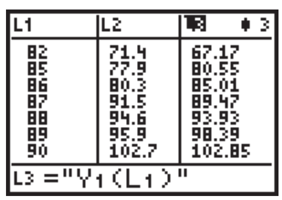


Finding Residuals

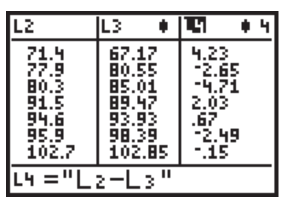
The TI-83 and TI-83 Plus will efficiently calculate residuals for any line of fit using the spreadsheet capabilities of the List Editor.

a. First make sure that your data are entered into lists L_1 and L_2, and that the equation of your fitted line is entered into the Y= screen.

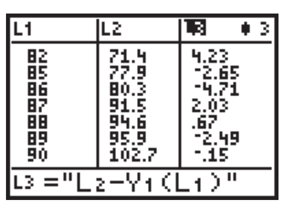
b. Define list L_3 as the predicted *y*-coordinates, $Y_1(L_1)$.



c. Then define list L_4 as the residuals, L_2-L_3.



As an alternative, you could directly define list L_3 as the residuals using the expression $L_2-Y_1(L_1)$.



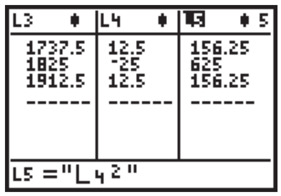
(continued)

Least Squares Regression Lines

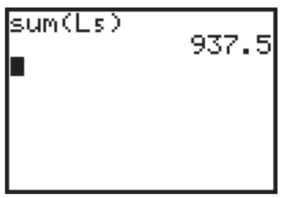
Sum of Squared Errors (SSE)

If the residuals for a fitted line are contained in a list, say, list **L4**, the Sum of Squared Errors (SSE) can be calculated. The examples below uses the DC9 data from pages 123–126 of the student text.

a. Define list **L5** as the squares of the residuals, **L4**^{2}.



b. On the Home screen, calculate the sum of list **L5**. Find the **sum(** command by pressing [2nd] **[LIST] MATH 5:sum(**. The result is the SSE.

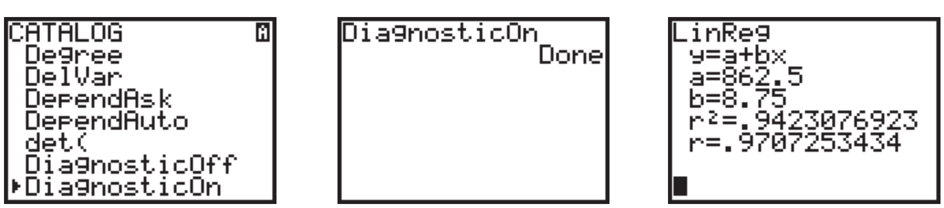


Using the Calculator to Find the Least Squares Regression Line

The TI-83 and TI-83 Plus provide two forms of the least squares regression line: $y = ax + b$ and $y = a + bx$. The **LinReg(ax+b)** and **LinReg(a+bx)** commands, as well as other regressive techniques, are found in the [STAT] **CALC** submenu. As explained on page 86 of the *Statistics in Action Instructor's Guide,* the form $y = a + bx$ has been accepted as the standard notation in statistics.

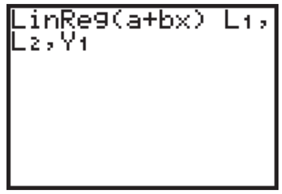
Although the regression commands default to lists **L1** and **L2**, it is strongly recommended that you *always* include the names of the lists—independent then dependent—separated by commas.

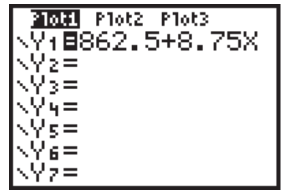
If you'd like the calculator to calculate and display the correlation, r, and coefficient of variation, r^2, you must first turn Diagnostics on. Press [2nd] **[Catalog] [D]**, arrow down to **DiagnosticOn**, and press [ENTER] [ENTER].



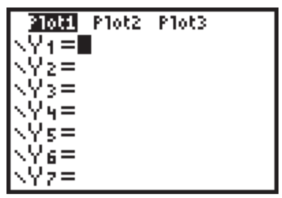
(continued)

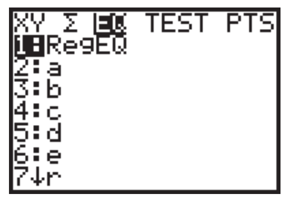
You may paste the equation of the least squares line into the Y= screen in one of two ways. First, when using the regression command, you can specify a function after specifying the lists, for example **LinReg(a+bx)L1,L2,Y1**. The calculator will immediately put the equation into the Y= screen.

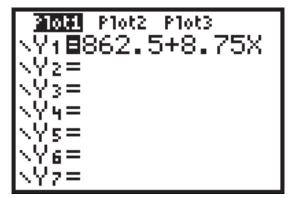




Second, you can paste the equation into the Y= screen after running the regression command. Press [Y=] and position the cursor beside one of the available functions. Press [VARS] **5:Statistics EQ 1:RegEq** to paste the equation in place.







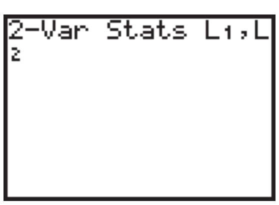
Section 3.3 ▪ Correlation: The Strength of a Linear Trend

See “Using the Calculator to Find the Least Squares Regression Line” above to learn how to get the correlation as part of regression. There is no way to calculate correlation automatically other than by performing a regression.

A Formula for the Correlation, r

If desired, the spreadsheet capabilities of the List Editor can aid the calculation of the correlation. This process resembles Display 3.41 on page 138 of the student text.

a. Enter the bivariate data into two lists, say, lists L1 and L2.



b. Press [STAT] **CALC 2:2-VarStats** and specify the lists. Doing a 2-Var Stats calculates and stores the means, $\bar{x}$ and $\bar{y}$, the standard deviations, **Sx** and **Sy**, and the number of data values, **n**. All of these variables can now be accessed by pressing [VARS] **5:Statistics.**

c. Press [STAT] **1:Edit** to return to the List Editor screen.

d. Define list L3 with the expression **(L1−$\bar{x}$)/Sx.**

e. Define list L4 with the expression **(L2−$\bar{y}$)/Sy.**

f. Define list L5 with the expression **L3*L4.**

(continued)

g. On the Home screen, enter **sum(L5)/(n−1)**. The result is the correlation, *r*. Find **sum(** by pressing 2nd [LIST] MATH 5:sum(.

Alternatively, after running the 2-Var Stats, you can directly calculate *r* on the Home screen by entering **sum(((L1−$\bar{x}$)/Sx)((L2−$\bar{y}$)/Sy))/(n−1)**.

Section 3.4 ▪ Diagnostics: Looking for Features That the Summaries Miss

Which Points Have Influence?

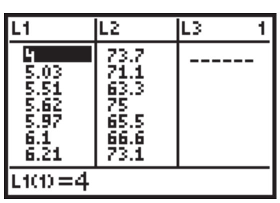
A calculator can support the investigation of potentially influential points. By eliminating questionable data points and rerunning the regression, you can numerically and graphically ascertain the effect of removing points that do not fit the general pattern of the scatterplot. Please note, however, that you must separately delete both the *x*- and *y*-coordinates of the data point in question. Within the List Editor screen, position the cursor on the *x*-coordinate to be deleted and press DEL. Repeat for the corresponding *y*-coordinate.

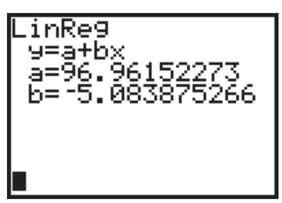
Residual Plots: Putting Your Data Under a Microscope

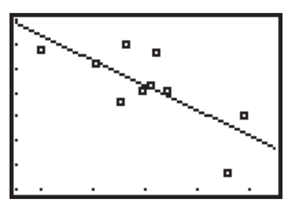
Each time The TI-83 and TI-83 Plus performs a regression, the calculator *automatically* computes the residuals and stores them in a list called **RESID**. Access this list by pressing 2nd [LIST] and going to the **NAMES** submenu.

Note that you can easily verify a necessary condition of residuals—that they sum to zero—by summing the values in **RESID**.

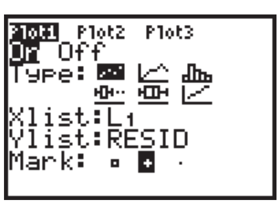
You can easily make a residual plot by creating a scatterplot with **RESID** as the **Ylist**. Here is the original plot and the residual plot for the data on percentage of on-time arrivals versus mishandled baggage from pages 161–162 of the student text.

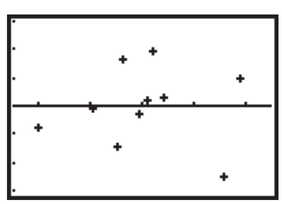






[3.5, 8.5, 1, 45, 80, 5]

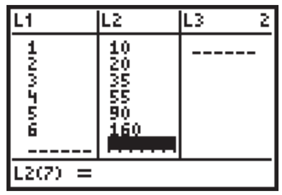


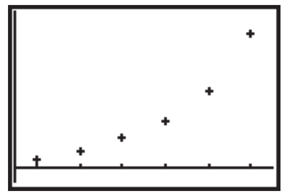


[3.5, 8.5, 1, −12, 12, 4]

Section 3.5 ▪ Shape-Changing Transformations

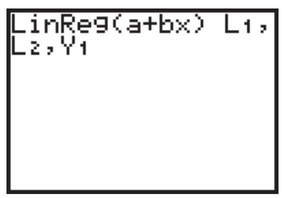
The TI-83 and TI-83 Plus can accomplish all of the transformations that are discussed in this section. However, the calculator always returns the coefficients of an equation in the form $y = a + bx$. It is up to you to note which transformations were used on which variables. For example, consider this small hypothetical data set that appears to exhibit an exponential relationship. Assume this data set relates days and growth, or data points in the form (d, g).

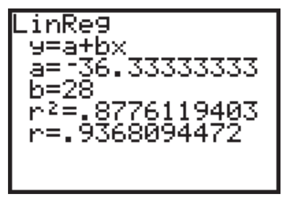




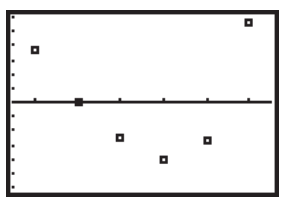
[0.5, 6.5, 1, −15.5, 185.5, 1]

A least squares regression on the original data gives the equation $y = -36.333 + 28x$. You should interpret this as $\hat{g} = -36.333 + 28d$.



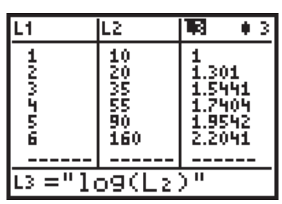


A residual plot confirms that a linear equation is a bad fit.



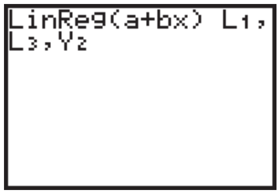
[0.5, 6.5, 1, −30, 30, 5]

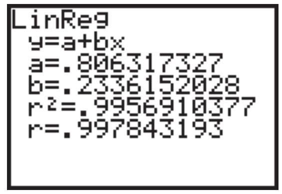
Now use the List Editor screen to take the logarithm (base 10) of each of the growth values (list L2). Store these transformed values in list L3.



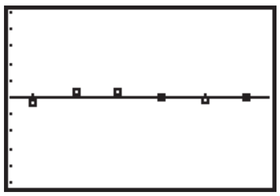
(continued)

Performing a least squares regression on the transformed data, $(d, \log(g))$, gives the equation $y = 0.806 + 0.234x$. You should interpret this as $\widehat{\log(g)} = 0.806 + 0.234d$.





A residual plot shows that the new equation is a better fit.



[0.5, 6.5, 1, −0.5, 0.5, 0.1]

