Scatter Plot, Correlation, and Regression on the TI-89

Summary: When you have a set of (x,y) data points and want to find the best equation to describe them, you are performing a regression. This page shows you how to determine the strength of the association between your two variables (correlation coefficient), and how to find the line of best fit (least squares regression line).

For an illustration of linear regression, we’ll use the data given below in a table. The explanatory variable x is dial settings on a freezer, and the response variable y is temperature of the freezer.

Contents:
- Step 0. Setup
- Step 1. Make the Scatter Plot
- Step 2. Perform the Regression
- Step 3. Display the Regression Line
- Step 4 (optional). Display the Residuals

Step 0. Setup

Set floating point mode, if you haven’t already.

```
[MODE] [▼] [▼] [►] [ALPHA ÷ makes E] [ENTER]
```

The calculator will remember this setting when you turn it off: next time you can start with Step 1.

Step 1. Make the Scatter Plot

Before you even run a regression, you should first plot the points and see whether they seem to lie along a straight line. If the distribution is obviously not a straight line, don’t do a linear regression. (Some other form of regression might still be appropriate, but that is outside the scope of this course.)
Turn off other plots.

Enter the numbers.

<table>
<thead>
<tr>
<th>Dial (x)</th>
<th>0</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp, °F (y)</td>
<td>6</td>
<td>−1</td>
<td>−3</td>
<td>−10</td>
<td>−16</td>
</tr>
</tbody>
</table>

[•] [APPS] and select Stats/List Editor.

[F2] [3] [F2] [4] turns off all plots and functions.

You will use two named lists for the x’s and y’s. Any names are possible, but I’ll use \(lx\) and \(ly\) because they’re short. If those lists already exist, highlight the \(lx\) name and press [CLEAR] [ENTER] to erase previous entries. If \(lx\) isn’t there yet, move to an empty list heading and press [L] [X]. (L is above the 4 key. When you press 4 while naming a list, it will change to L automatically.)

Enter the x numbers, then clear list \(ly\) (or create it) and enter the y numbers.

Note: You can hide an unwanted list by cursoring to the list name and pressing [• ← makes DEL]. The list remains in memory until you use [2nd − makes VARLINK] to delete it.

Set up the scatter plot.

[F2] [1] [F1] opens a dialog box. You want these settings:

- Plot type: Scatter
- Mark: anything except dot (because a data dot looks just like a dot on the grid)
- X: [alpha] [L] [X]
- Y: [alpha] [L] [Y]
- Use Freq and categories: NO

Press [ENTER] to complete the definition.

Plot the points.

[F5] automatically adjusts the window frame to fit the data.

(optional) You can adjust the grid to look better.

[• F2 makes WINDOW], set \(Xscl=1\) and \(Yscl=5\), then [• F3 makes GRAPH] to redisplay it.

Appropriate values of \(Xscl\) and \(Yscl\) may be different for other problems. Pick the values that make the graph look best to you.
Step 2. Perform the Regression

Set up to calculate statistics.

Write down a (slope), b (y intercept), r (correlation coefficient; \( r^* \) is our symbol). Round a and b to two more decimal places than your actual y values have; remember that final rounding should be done only at the end of calculations. Round \( r^* \) to two decimal places unless it’s very close to ±1 or to 0.

\[
\begin{align*}
a &= -3.52 \\
b &= 6.46 \\
r^* &= -0.992
\end{align*}
\]

\( R^2 \) is the coefficient of determination. The closer it is to 1, the better a predictor is the regression equation. Another way to look at it is that in this case \( R^2 \) is about 98%, so 98% of the variation in y is associated with the variation in x.
Statisticians say that $R^2$ tells you how much of the variation in $y$ is “explained” by variation in $x$, but if you use that word remember that it means a numerical association, not necessarily a cause-and-effect explanation.

Only linear regression will have a correlation coefficient $r$, but any type of regression will have a coefficient of determination $R^2$ that tells you how well the regression equation predicts $y$ from the independent variable(s). (The calculator uses $r^2$, but most authors use $R^2$.)

**Step 3. Display the Regression Line**

Show line with original data points. [F3] makes GRAPH

**Step 4 (optional). Display the Residuals**

A plot of residuals can be helpful to show whether linear regression was the right choice.

If the residuals are more or less evenly distributed above and below the axis and show no particular trend, you were probably right to choose linear regression. But if there is a trend, you have probably forced a linear regression on non-linear data. If your data points looked like they fit a straight line but the residuals show a trend, it probably means that you took data along a small part of a curve.

The residuals are automatically calculated during the and stored in a resid list in your Stats/List Editor; all you have to do is plot them on the $y$ axis against your existing $x$ data.

Turn off other plots. [F2] [3] [F2] [4]

Set up the plot of residuals against the $x$ data. [F2] [1] [▼] [F1] selects Plot 2 and opens a dialog box. You want these settings:
Don’t worry about the magnitude of the residuals, because \([\text{ZOOM}] [9]\) adjusts the vertical scale so that the points take up the full screen. What you want to look at is **whether there’s a trend in the residuals**. Here there is no trend, so you conclude that a linear regression was the right choice, as opposed to regression against some curve.