Plot – One or Two Numeric Variables

This procedure creates a scatterplot of the data in up to two numeric columns. It also calculates the correlation coefficient for the variables.

The data for this analysis consist of *n* values of two numeric variables. Let

y_i = *i*-th value of variable 1.

x_i = *i*-th value of variable 2. If only one variable is specified, row numbers will be used for x.

Access

Highlight: two numeric columns. Any column other than a *Character* type column may be selected.

Select: *Describe* from the main menu.

Output Page 1: A scatterplot of the data using point symbols.

Output Page 2: A line plot of the data without point symbols.

Output Page 3: A connected scatterplot of the data using both points and lines

Note: if both variables are *Response* variables, the scatterplot is included as part of a more extensive analysis as described in the document titled *Describe – Multiple Response Variables.*

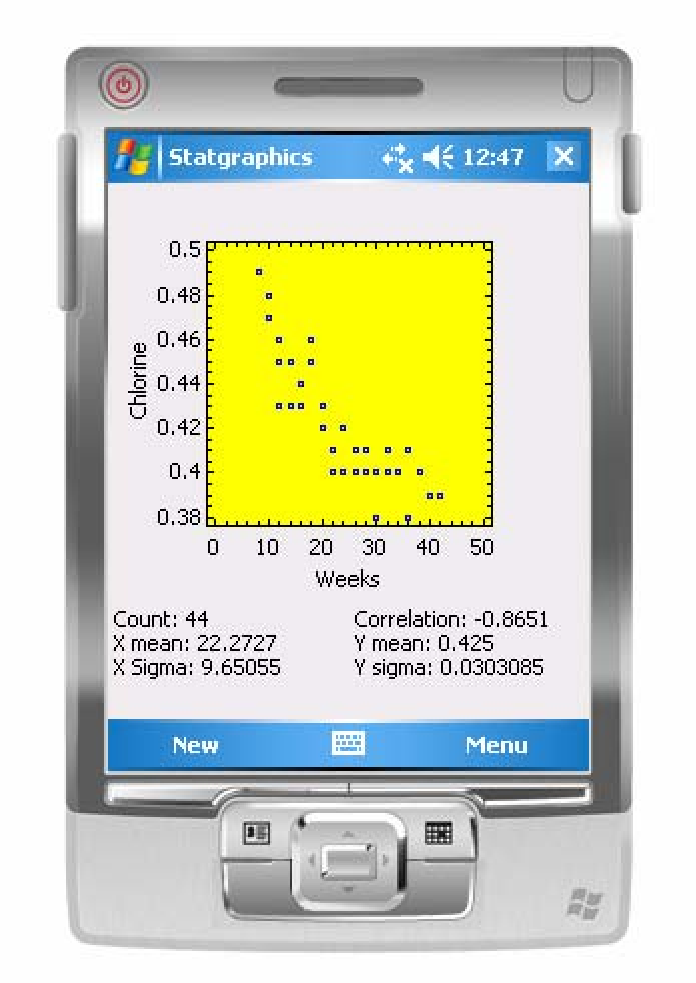
Sample Data

The text entitled Applied Regression Analysis, third edition by Draper and Smith (Wiley, 1998) contains a sample of *n* = 44 measurements of the age and amount of chlorine in samples of a product. The data is contained in the file *chlorine.sgm*. The first several rows of the file are shown below:

Row	Chlorine	Weeks
1	8	0.49
2	8	0.49
3	10	0.48
4	10	0.47
5	10	0.48
6	10	0.47
7	12	0.46
8	12	0.46
9	12	0.45
10	12	0.43

Scatter Plot

The *Scatter Plot* plots all pairs of values in the two variables.



It also displays the sample mean for each variable, the sample standard deviation, and the correlation coefficient. The sample mean of a variable is calculated by

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{1}$$

The sample standard deviation is calculated by

$$s = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}} \tag{2}$$

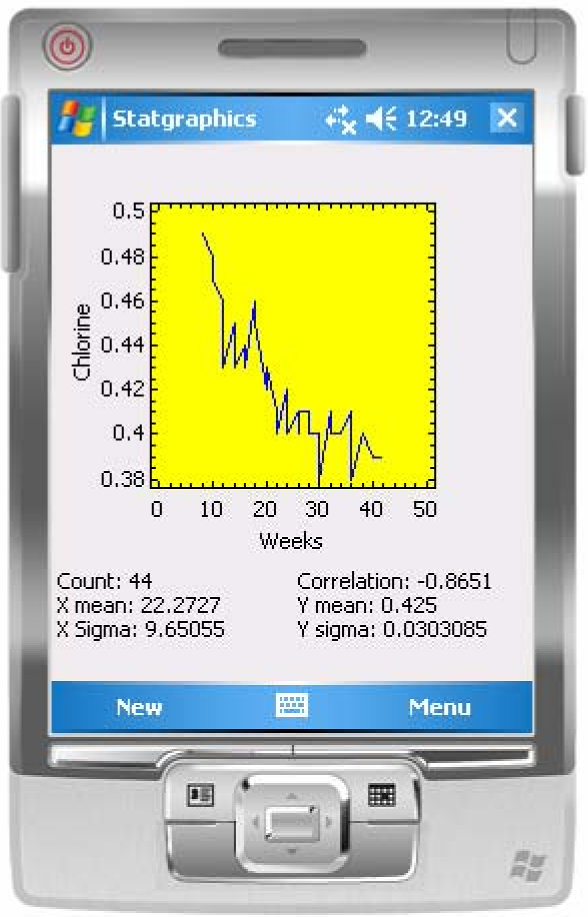
The correlation coefficient *r* ranges from -1 to +1 and measures the strength of the linear correlation between the variables. It is calculated from

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (3)$$

The closer the values fall to a straight line with positive slope, the closer *r* is to 1. Points that lie close to a line with negative slope will yield an *r* close to -1. A value of *r* close to 0 indicates little if any correlation between the variables.

Line Plot

The *Line Plot* connects the data values in row order without displaying point symbols.



Connected Plot

The *Connected Plot* connects the data values in row order and also displays point symbols.