

## 2 変量記述統計

2 変量の関係を表す散布図の描画と相関係数・回帰直線を求めるプログラム PAnal2Vars.dpr、および画面のクリックによって設定した散布図に対して相関係数の算出と回帰曲線の描画を行うプログラム PCheckCor.dpr について説明する。

### PAnal2Vars. dpr

このプログラムを実行すると図 1.1 のフォームが表示される。

The screenshot shows a Windows-style dialog box titled "2変量の関係の分析". It contains two text input fields: "Xのラベル" with the value "X変数" and "Yのラベル" with the value "Y変数". Below these is a table with two columns labeled "X =" and "Y =" and one row labeled "1番目". To the right of the table is a vertical column of buttons: "終了", "削除", "追加", "保存", "読出", "計算", "印刷", and "散布図".

	X =	Y =
1番目		

図 1.1 実行開始時のフォーム

StringGrid コンポーネントの行数をデータ数に合わせて増やし、「i 番目」と表示されている行に i 番目のデータの組  $(x_i, y_i)$  を設定する。「X =」の欄に  $x_i$  を、「Y =」の欄に  $y_i$  を設定する。「保存」ボタンのクリックでアクティブなセルの下に新しい行が追加され、「削除」ボタンのクリックでアクティブなセルを含む行が削除される。セルはクリックすると

アクティブになる。

「Xのラベル」および「Yのラベル」の右隣の Edit コンポーネントには、それぞれ第1変量  $x_i$  および第2変量  $y_i$  のラベルを設定する。これらのラベルは散布図において用いられる。

	X =	Y =
1番目	1	1
2番目	1	2
3番目	2	2
4番目	2	3
5番目	3	3
6番目	3	4
7番目	4	5
8番目	5	5
9番目	5	6

図 1.2 データの設定

図 1.2 のように設定したデータは「保存」ボタンのクリックでファイルに保存することができる。「保存」ボタンをクリックすると図 1.3 のダイアログボックスが表示される。

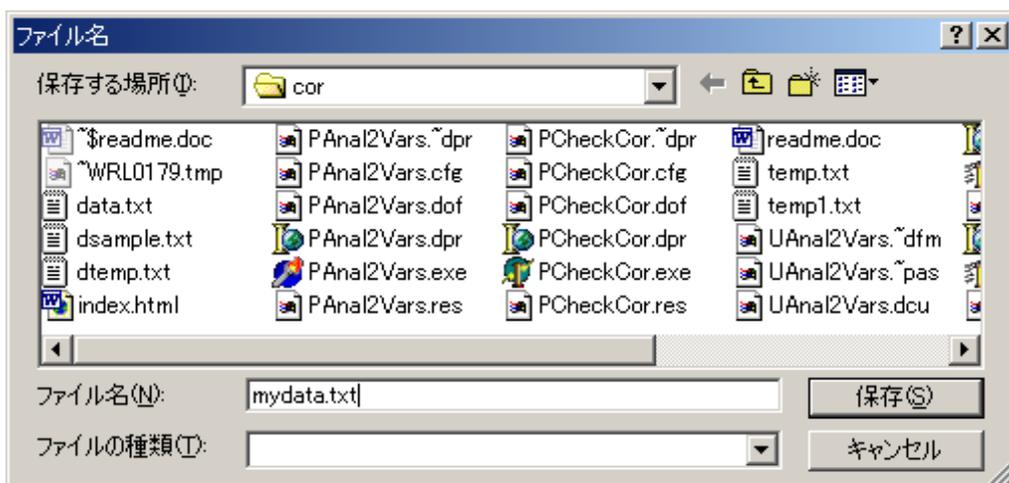


図 1.3 データ保存用ファイル名の設定

図 1.3 ではファイルの拡張子は.txt としているが、これはデータがテキストファイルとして書き出されるのに合わせたものである。拡張子は.txt に限らず何でもよい。ファイル名の設定後、図 1.3 における「保存」ボタンをクリックすると設定したファイル名のファイルにデータが保存される。

保存したデータは「読出」ボタンのクリックで読み込むことができる。「読出」をクリックすると図 1.4 のようなダイアログボックスが表示される。

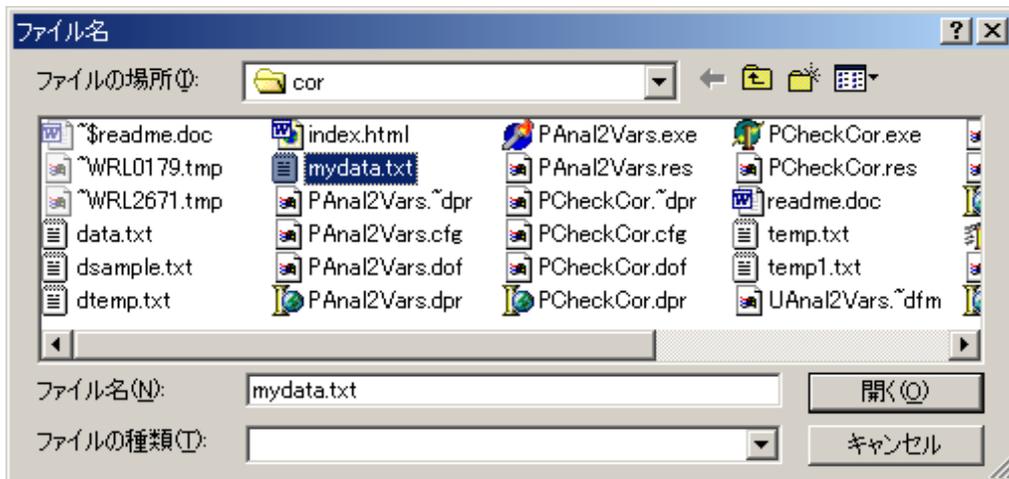


図 1.4 データの読み込み

ファイル名を設定後、「開く」ボタンをクリックするとファイルからデータが読み込まれる。

データの設定後、「散布図」ボタンをクリックすると散布図が描画される。「散布図」ボタンをクリックすると、まず図 1.5 のフォームが表示される。

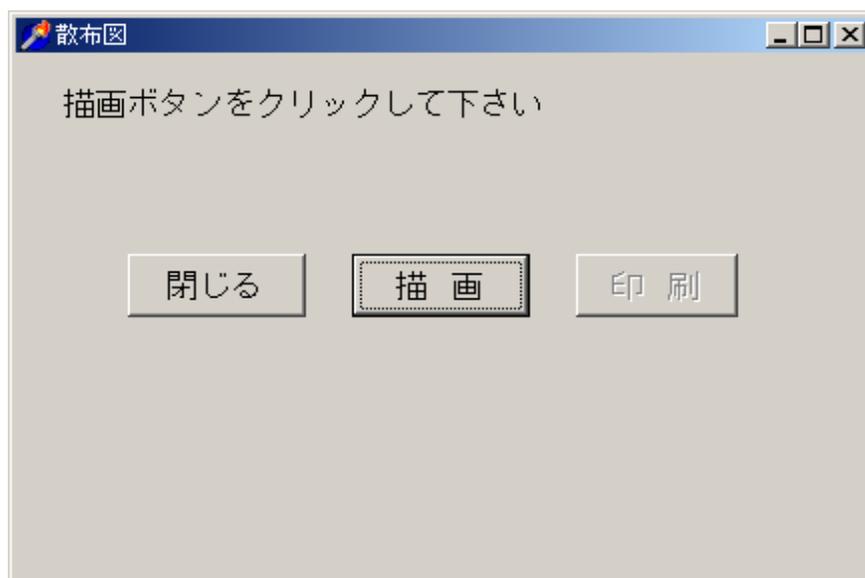


図 1.5 「散布図」 ボタンのクリックで表示される

「描画」 ボタンをクリックすると図 1.6 のようなダイアログボックスが表示される。

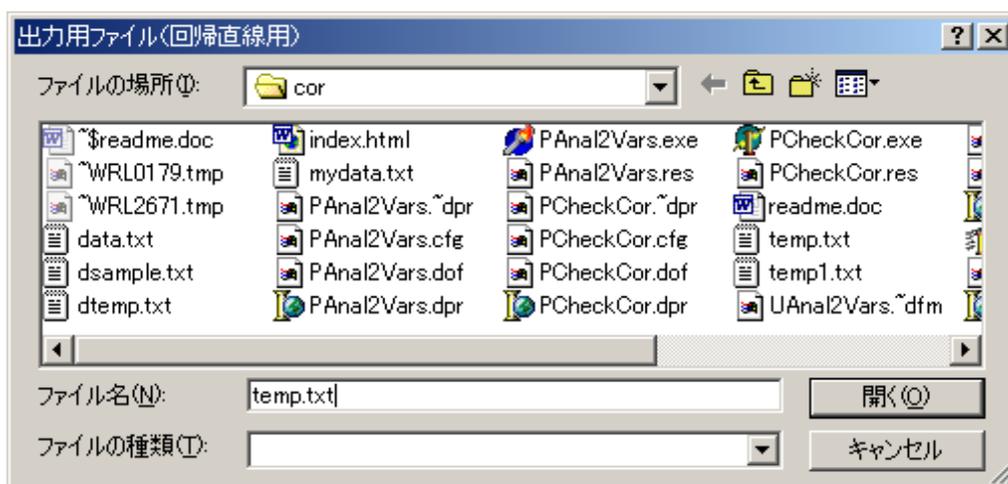


図 1.6 出力用ファイル名の設定 (回帰直線)

図 1.6 のダイアログボックスで設定したファイル名のテキストファイルに回帰分析の結果が書き出される。ファイル名設定後、「開く」ボタンをクリックすると散布図が図 1.7 のように描かれる。青色の直線は回帰直線を表す。

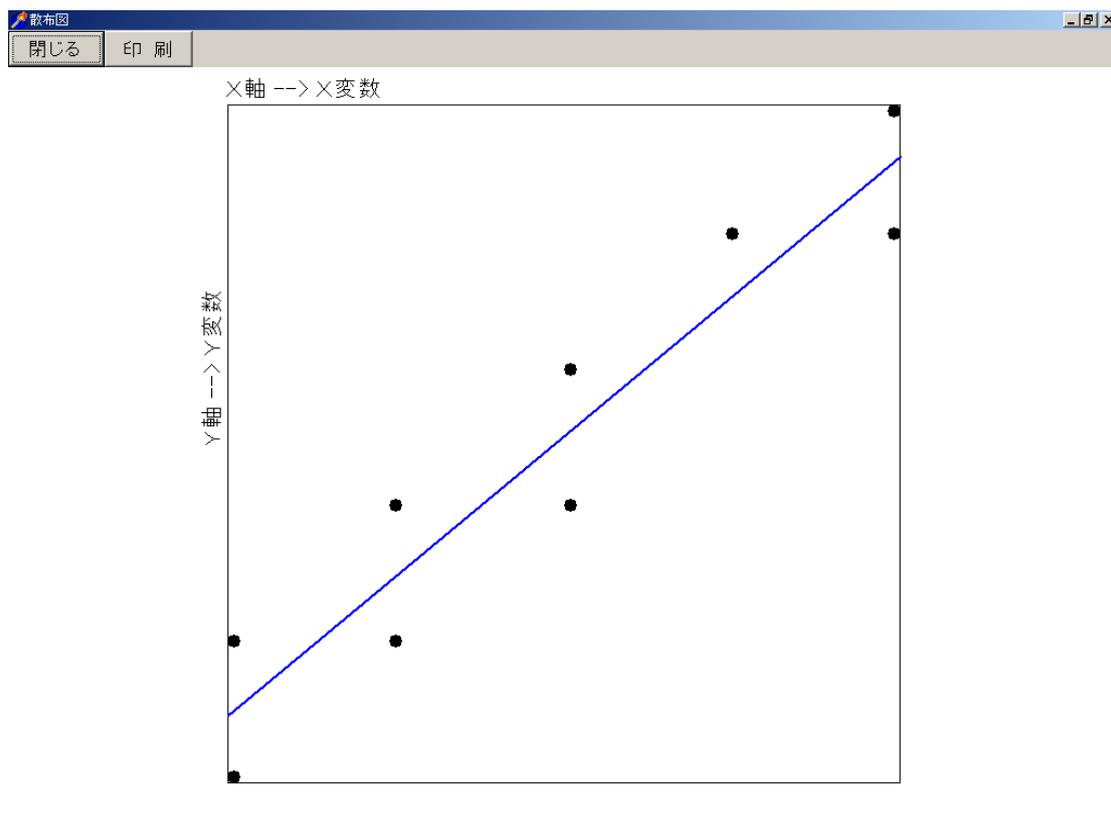


図 1.7 散布図

「印刷」ボタンをクリックすると、散布図がプリンタに出力される。プリンタ出力では散布図におけるドットの大きさが表す度数も出力される。これは、例えば図 1.8 のように同じデータが複数ある場合、その度数に対応した大きさのドットで散布図が描かれるので、ドットの大きさが表す度数を示すものである。



図 1.8 同じデータが複数個ある場合

図 1.8 のデータで散布図を描くと図 1.9 のようになる。

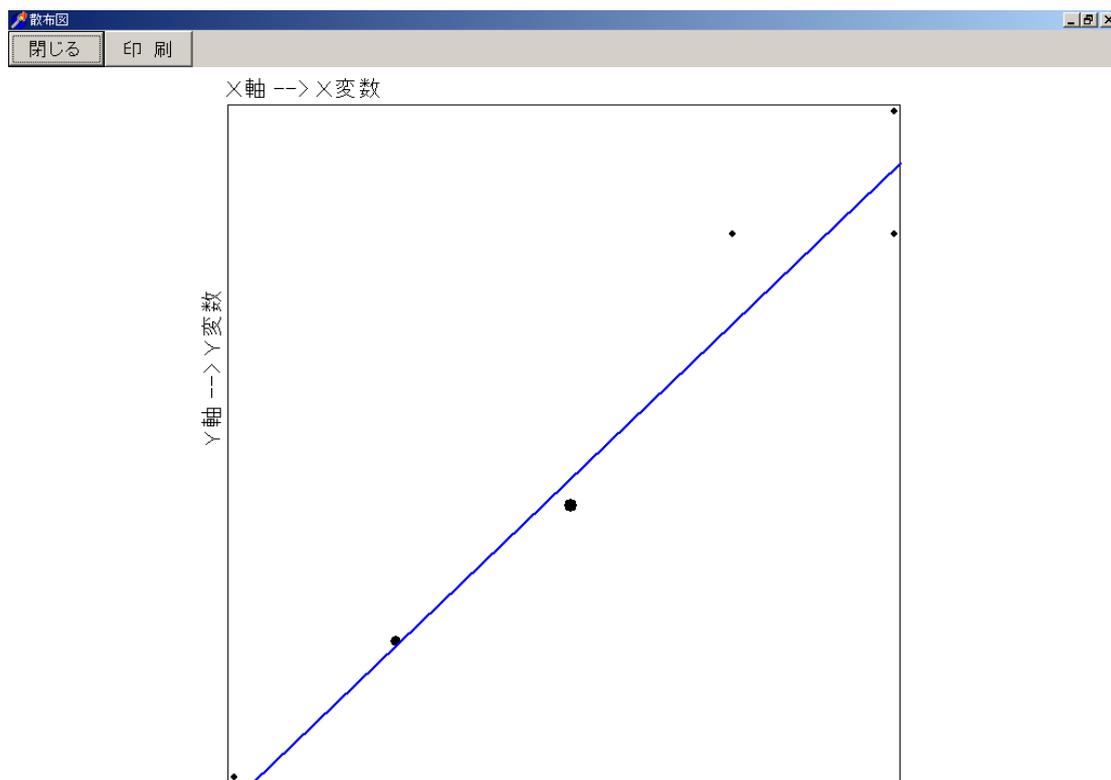


図 1.9 同じデータが複数個ある場合の散布図

データが複数個ある位置のドットの大きさが、データが1個のドットよりも大きいことが分かる。プリンタ出力には、これらのドットのうち最大の大きさと最小の大きさのものが表わしている度数が出力される（図 1.10）。



図 1.10 プリンタ出力の一部（下部）

プリンタ出力には回帰直線の係数も出力されているが、これは図 1.6 で指定したファイル名のテキストファイルにも書き出されている。このファイルはプログラムの実行終了後にエディタなどで見ることができる（リスト 1.1）。

リスト 1.1 出力ファイル（散布図に対するもの）

変数 X のラベル --> X 変数		変数 Y のラベル --> Y 変数	
X =	Y =	a*X+b =	誤差
1.000	1.000	1.500	-0.500
1.000	2.000	1.500	0.500
2.000	2.000	2.529	-0.529
2.000	3.000	2.529	0.471
3.000	3.000	3.559	-0.559
3.000	4.000	3.559	0.441
4.000	5.000	4.588	0.412
5.000	5.000	5.618	-0.618
5.000	6.000	5.618	0.382
a =		1.02941	
b =		0.47059	
r =		0.94907	
決定係数 =		0.90074	

回帰直線  $y = ax + b$  の係数  $a$  および  $b$  の値、相関係数  $r$ 、決定係数  $r^2$  が書き出されている。これらの値は図 1.2 のデータに対するものである。

図 1.2 において「計算」ボタンをクリックすると、統計量の計算が始まる。「計算」ボタンのクリックでまず計算結果を出力するファイルの名前の設定を求めるダイアログボックスが図 1.11 のように表示される。

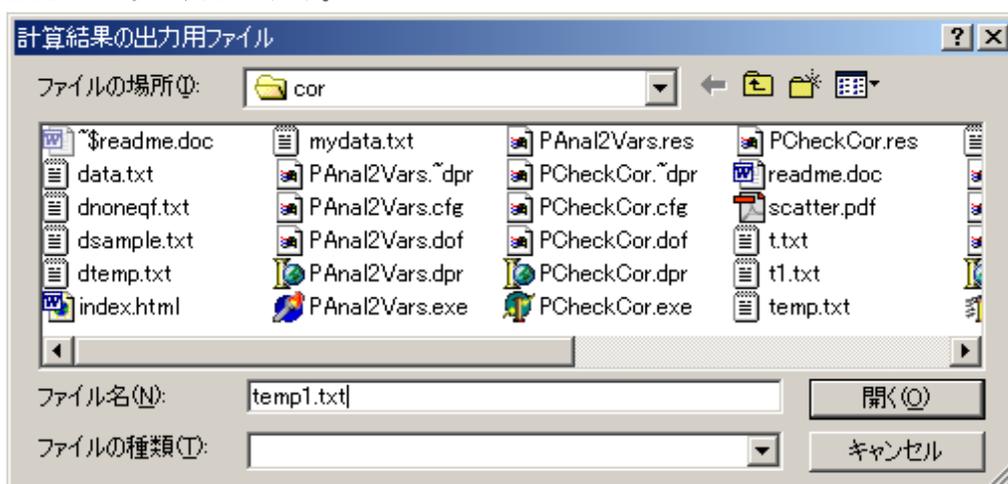


図 1.11 計算結果出力用ファイル名の設定

ファイル名の設定後、「開く」ボタンのクリックで計算が行われ、計算が終了すると図 1.2

のフォームの「終了」ボタンがアクティブになる。出力ファイルにはテキストファイルとして出力されているので、プログラムの実行終了後エディタなどで見ることができる。図 1.2 のデータの場合の出力はリスト 1.2 のようになっている。

リスト 1.2 計算結果の出力例

```

第 1 変数 ==> X 変数
第 2 変数 ==> Y 変数

データ =
  x[1] =      1.00    y[1] =      1.00
  x[2] =      1.00    y[2] =      2.00
  x[3] =      2.00    y[3] =      2.00
  x[4] =      2.00    y[4] =      3.00
  x[5] =      3.00    y[5] =      3.00
  x[6] =      3.00    y[6] =      4.00
  x[7] =      4.00    y[7] =      5.00
  x[8] =      5.00    y[8] =      5.00
  x[9] =      5.00    y[9] =      6.00

X の平均      = 2.88888889
X の分散      = 2.09876543
X の標準偏差  = 1.44871165
X の不偏分散 = 2.36111111
X の不偏分散の平方根 = 1.53659074

Y の平均      = 3.44444444
Y の分散      = 2.4691358
Y の標準偏差  = 1.5713484
Y の不偏分散 = 2.77777778
Y の不偏分散の平方根 = 1.66666667

相関係数 = 0.949070753
T = 7.9698506    df = 7

```

リスト 1.2 の最後に出力されている T の値は次式で与えられるものである。

$$T = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}$$

母集団での相関係数が 0 であるとき、T は自由度  $df = n - 2$  の t 分布に従う<sup>1</sup>。

<sup>1</sup> 岡本安晴 「Delphi で学ぶデータ分析法」、CQ 出版社、1998、p.83 などを参照

## PCheckCor.dpr

このプログラムは、散布図と相関係数の関係を視覚的に体験するためのものである。プログラムを実行すると図 2.1 のフォームが表示される。



図 2.1 実行開始時のフォーム

「GO」 ボタンのクリックで図 2.2 の画面になる。

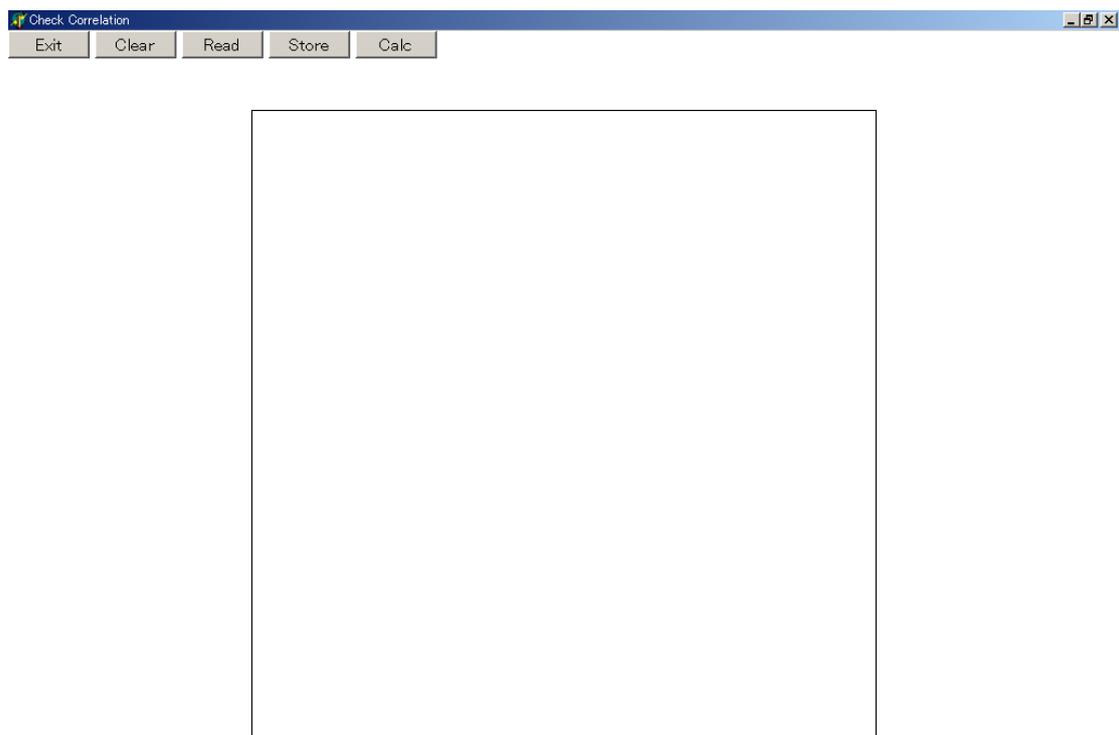


図 2.2 散布図用の画面

画面内の正方形の中をマウスの左ボタンでクリックすると青色のドットが描かれる。青色のドットを右ボタンでクリックするとそのドットは消える。正方形内のいろいろな位置を左ボタンでクリックする、あるいは消したいドットは右ボタンクリックで消すということを実行を繰り返すことにより、ドットのパターン（散布図）を描くことができる（図 2.3）。

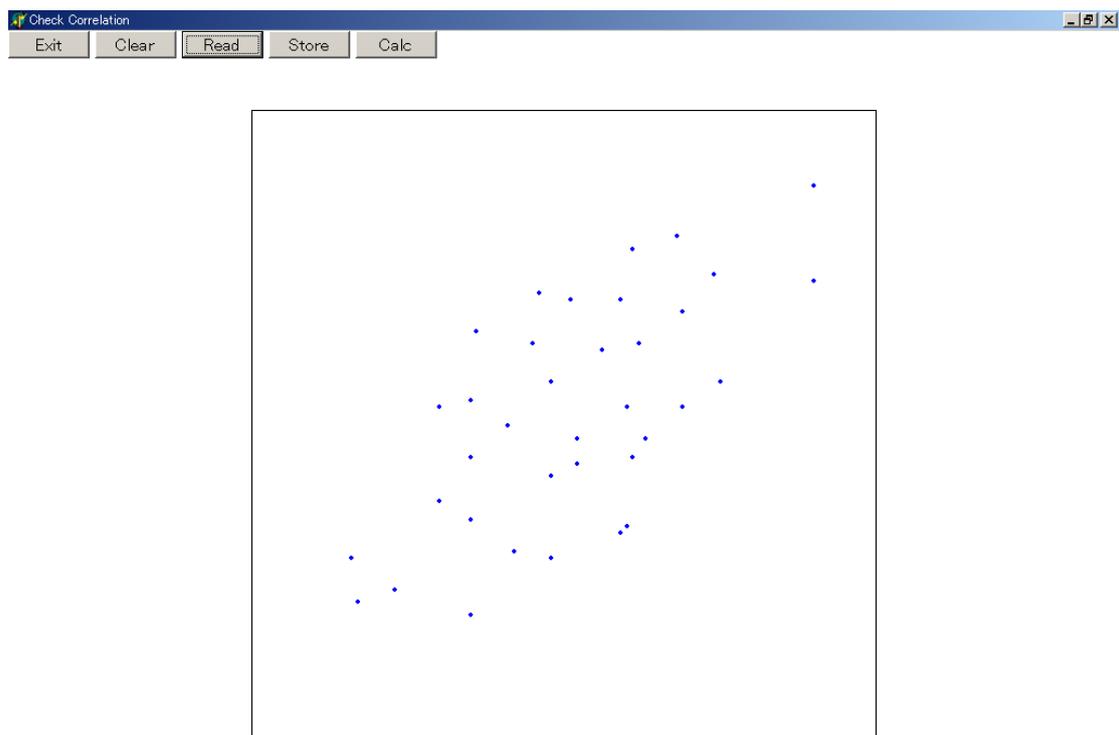


図 2.3 散布図の作成

作成したドットパターンは、「Store」ボタンのクリックによりファイルに保存することができる。「Store」ボタンのクリックで表示されるダイアログボックスにおいてデータを保存するためのファイル名を設定する。ファイル名の拡張子は何でもよいが、ファイルはテキストファイルなので\*.txt としておけばエディタで開くときに便利である。

保存したデータは「Read」ボタンのクリックで読み出すことができる。

図 2.3 のようにドットを配置した状態で「Calc」ボタンをクリックすると、図 2.4 のように画面上の散布図に対する相関係数が算出・表示され、回帰直線が描かれる。

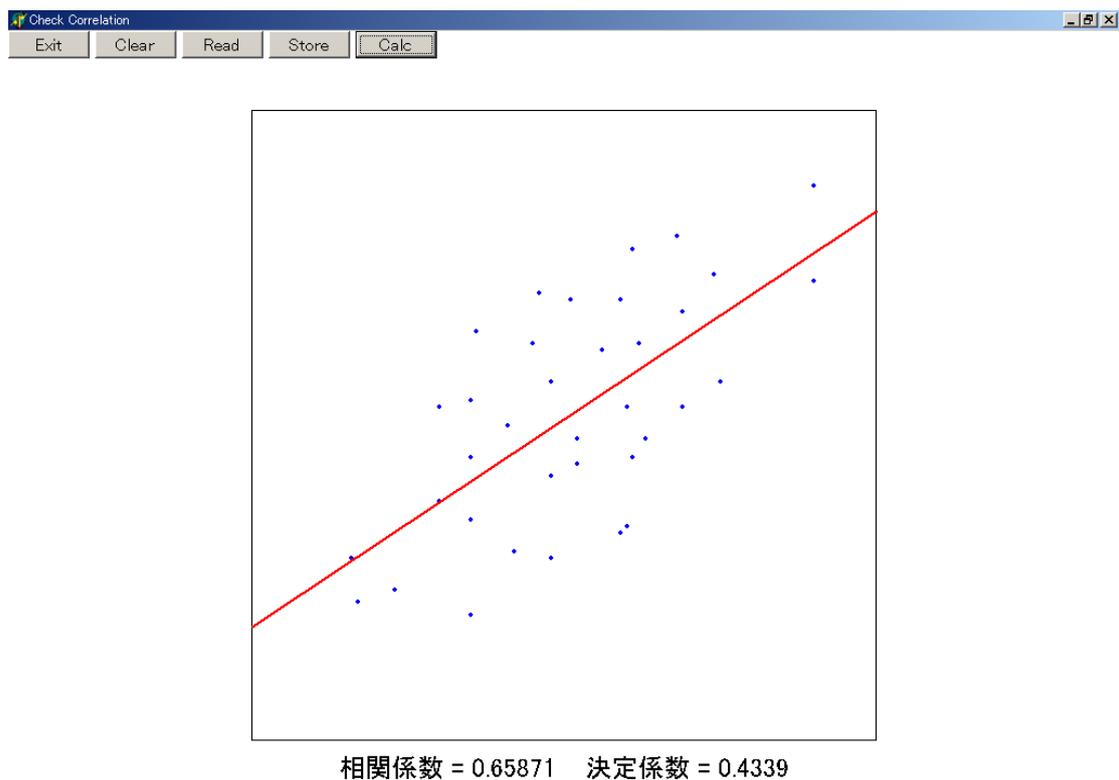


図 2.4 回帰直線の描画

図 2.4 の状態で、さらに左ボタンクリックあるいは右ボタンクリックによって散布図におけるドットパターンを変えた後、「Calc」ボタンをクリックすると、変更した散布図に対する相関係数などが再計算され表示される（図 2.5）。

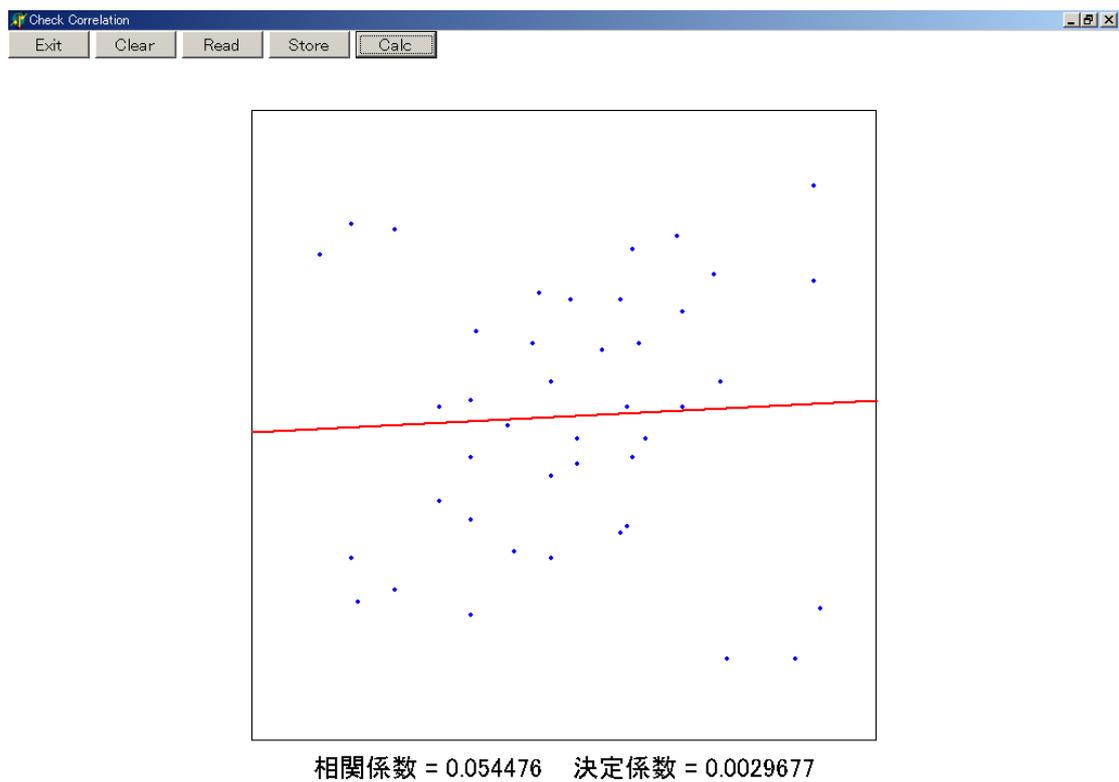


図 2.5 ドットパターンの変更後の再描画