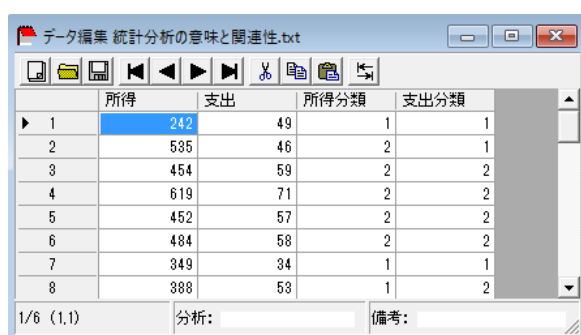


相関係数の検定、t検定、 χ^2 検定の関係について

相関係数の検定、t検定、 χ^2 検定は別々に考えられることが多いが、これらはお互い繋がっている。ここではこの関係を実際のデータを用いて調べてみる。ここでの分析は、メニュー「分析－基本統計－」の中の量的データの集計、量的データの検定、質的データの検定を利用している。

2つの量的変数が与えられたとき、それらの間の関係の有無は一般に相関係数の検定で調べられる。例えばこれを所得と支出とし、図1のようなデータが与えられているものとする。またこれらの変数の基本統計量を図2に示す。



	所得	支出	所得分類	支出分類
1	242	49	1	1
2	535	46	2	1
3	454	59	2	2
4	619	71	2	2
5	452	57	2	2
6	484	58	2	2
7	349	34	1	1
8	388	53	1	2

図1 所得と支出のデータ



	所得	支出
データ数	40	40
最小値	192.000	32.000
最大値	619.000	71.000
平均値	406.650	51.100
中央値	422.000	51.500
レンジ	427.000	39.000
分散	8088.978	74.740
標準偏差	89.939	8.645
不偏分散	8296.387	76.656
標準偏差	91.085	8.755
標準誤差	14.221	1.367
歪度	-0.097	-0.499
尖度	0.079	-0.053

図2 基本統計量

所得と支出の関係は図3のような散布図によって与えられ、関係の有無は相関係数によって計算される。

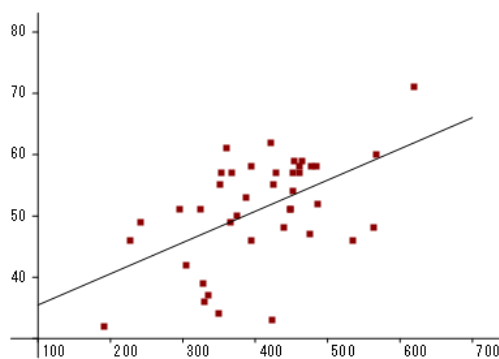


図3 所得と支出の散布図

ここで所得を400(万円)で高群と低群に分けてみる。散布図で分けた場合を図4に、支出の度数をヒストグラム(並べて描いてあるので注意)にした場合を図5に示す。

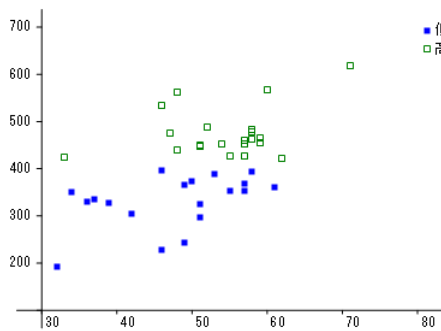


図4 所得で分けた散布図

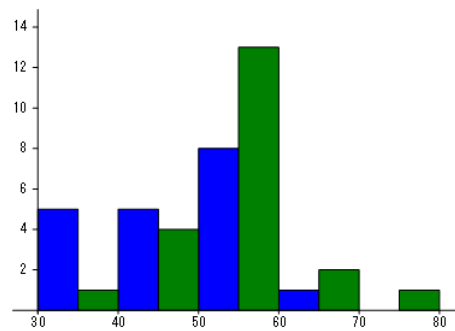


図5 所得で分けた支出のヒストグラム

但し、散布図はヒストグラムとの対応が見易いように、縦軸と横軸を図3とは逆に描いている。このように見ると所得で分けた2群間の支出の差の検定が可能のように思われる。

さて、この2つの群の差はどこから生じたものであろうか。これは2変数の相関（または、回帰直線が引けるかどうか）から生じる差である。もし、相関がなく、散布図が丸くまとまっていたらこの差は表れない。そこで、我々は相関係数の検定と2群のt検定を両方行って結果を比較することにする。

相関係数の検定	相関係数	0.530	2群間のt検定
t統計値	3.8533		t統計値
両側確率P	0.0004		両側確率P
			0.0121

データを2群に分けて情報量を失わせたので、t検定の方が検定確率は大きくなっている。もちろん、状況によって2つの検定確率があまり変わらない場合や大きく異なる場合がある。

我々は、さらに支出を2群に分類して、2群の差について分割表の検定を用いて調べてみる。分割表を図6に示す。

	支出分類1	支出分類2	合計
所得分類1	10	9	19
所得分類2	5	16	21
合計	15	25	40

図6 所得と支出の分割による2次元分割表

所得による支出に差があるかどうかの χ^2 検定結果は以下となる。

χ^2 統計値	2.4127
片側確率P	0.1204

2つの量的データを分割すると失われる情報は多く、確率値はかなり大きくなっていることが分かる。ここではこれらの検定の関係と検定確率の違いについて大体どの程度であるか感じてもらえれば十分である。