

カイ 2 乗検定 (連関表の分析)

2つの事象の関係、例えば、2つのサイコロを投げたとき、それぞれの目の出方が他方の目の出方と無関係（独立）であるかどうか、ということを検定することを考える。

いま、サイコロAとBを2つ一緒に投げるということを、1000回繰り返したときの結果をまとめると、表1のようになったとする。

表 1 サイコロAとBを2つ同時に1000回投げたときのデータ（シミュレーションによる）

		Bの目					
		1	2	3	4	5	6
A の 目	1	28	32	24	30	32	33
	2	26	29	26	26	22	28
	3	18	19	32	33	26	32
	4	29	27	25	29	32	25
	5	30	37	26	20	31	27
	6	32	22	25	29	30	28

表1は、2つ一緒に投げたときのAの目の数とBの目の数によって、結果を分類したものである。この表によれば、A、Bともに1の目が出た場合が28回、Aの目が1でBの目が2であった場合が32回であったことが分かる。このような表は、連関表と呼ばれている。この表のデータによって、Aの目の出方とBの目の出方が独立であるかどうかということ、カイ2乗検定と呼ばれている方法によって調べてみる。この方法は、1変量の度数分布に対するカイ2乗検定の特別な場合と考えられる（Gibbons,1971）。

一般に、表2のような連関表について説明する。

表2 連関表

		変数 B			
		B ₁	・ ・ ・	B _c	
変 数 A	A ₁	f ₁₁	・ ・ ・	f _{1c}	f _{1・}
	・	・	・	・	・
	・	・	・	・	・
	・	・	・	・	・
	A _r	f _{r1}	・ ・ ・	f _{rc}	f _{r・}
		f _{・1}	・ ・ ・	f _{・c}	f

表2では、観測データが、2つの変数AとBのカテゴリーによって分類されている。変数AについてのカテゴリーがA_i、変数BについてのカテゴリーがB_jであったものの観測度数が、f_{ij}で表わされている。

いま、1つのデータが、変数AについてカテゴリーA_iに分類される確率をP(A_i)、変数BについてカテゴリーB_jに分類される確率をP(B_j)で表わすことにする。データが同時にカテゴリーA_iとB_jに分類される確率は、P(A_i ∩ B_j)で表わす。

変数Aでの分類と変数Bでの分類が独立であれば、

$$P(A_i \cap B_j) = P(A_i) \cdot P(B_j)$$

の関係が成り立つ。従って、表2において、独立性の仮定の下で、カテゴリーがA_iかつB_jである場合の期待度数e_{ij}は、次のようになる。

$$\begin{aligned}
 e_{ij} &= f \cdot P(A_i \cap B_j) \\
 &= f \cdot P(A_i) \cdot P(B_j) \\
 &= f \cdot \frac{f_{i\cdot}}{f} \cdot \frac{f_{\cdot j}}{f} \\
 &= \frac{f_{i\cdot} \cdot f_{\cdot j}}{f}
 \end{aligned}$$

ここで、

$$f_{i\cdot} = \sum_{j=1}^c f_{ij}$$

$$f_{\cdot j} = \sum_{i=1}^r f_{ij}$$

$$f = \sum_{i=1}^r \sum_{j=1}^c f_{ij}$$

である。

1変量の度数分布の検定の場合のカイ2乗検定の式に、上の結果をあてはめると、

$$c^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

$$= \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - f_{i\cdot} \cdot f_{\cdot j} / f)^2}{f_{i\cdot} \cdot f_{\cdot j} / f} \quad (1)$$

となる。式(1)での自由度dfは、1変量の度数分布の場合はrc-1であるが、連関表の場合は独立性の仮定による自由度の調節が必要である。

e_{ij} を求めるために、 $P(A_i)$ と $P(B_j)$ の値が使われている。 $P(A_i)$ および $P(B_j)$ には、それぞれ和が1になるという制約があるので、自由に動けるパラメータとしての $P(A_i)$ と $P(B_j)$ の数は、r-1とc-1となる。式(1)に対する連関表の場合の自由度dfは、これらの数を1変量の場合の自由度から減じた次の値になる。

$$df = (rc - 1) - (r - 1) - (c - 1)$$

$$= (r - 1)(c - 1)$$

独立性の検定は、式(1)で算出される値 c^2 を、自由度 $df = (r - 1)(c - 1)$ のカイ2乗分布と比較して行う。

自由度 $(r-1)(c-1)$ のカイ2乗分布に従う確率変数Xに対して、次式

$$P(X \geq c_a^2) = a$$

を満たす値 c_a^2 を基準値として求める。式(1)により算出される値 c^2 が c_a^2 より大きいとき、有意水準 α で、帰無仮説「カテゴリー A_i と B_j は独立である」は棄却される。

連関表のカイ 2 乗検定の場合も、1次元の場合の度数分布のカイ 2 乗検定におけるように、小さい期待度数 e_{ij} に対する注意が必要である。 $r \times c$ 個の期待値 e_{ij} の 20% を越えるものが 5 より小さいということがないように、さらに、どの e_{ij} も 1 より小さいということがないように注意する必要がある (Siegel & Castellan, Jr., 1988)。

プログラム PChiSqrXTable.dpr は、連関表のカイ 2 乗検定を行うものである。このプログラムを実行すると図 1 のフォームが表示される。

	列カテゴリー 1
行カテゴリー 1	

削除(行) 削除(列) 保存 計算

追加(行) 追加(列) 読出 印刷

終了

図 1 起動時のフォーム

行カテゴリーと列カテゴリーの数に合わせて、StringGrid 内の列数と行数を調整する。「追加(行)」あるいは「追加(条件)」ボタンのクリックでアクティブなセルを含む行あるいは

列の後に新しく行あるいは列が追加・挿入される。セルはクリックによりアクティブになる。「削除(行)」あるいは「削除(条件)」ボタンをクリックすると、アクティブなセルを含む行あるいは列が削除される。

列数と行数をそれぞれのカテゴリ数に合わせた後、データを設定する。表1のデータを設定すると、図2のようになる。

The screenshot shows a window titled 'Form1' containing a table with 6 rows and 3 columns. The first column lists row categories from 1 to 6. The second and third columns are labeled '列カテゴリ 5' and '列カテゴリ 6' respectively. The table contains numerical data. Below the table is a set of control buttons: '削除(行)', '削除(列)', '保存', '計算', '追加(行)', '追加(列)', '読出', '印刷', and '終了'.

	列カテゴリ 5	列カテゴリ 6
行カテゴリ 1	32	33
行カテゴリ 2	22	28
行カテゴリ 3	26	32
行カテゴリ 4	32	25
行カテゴリ 5	31	27
行カテゴリ 6	30	28

図2 データの設定

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3のダイアログボックスが表示される。

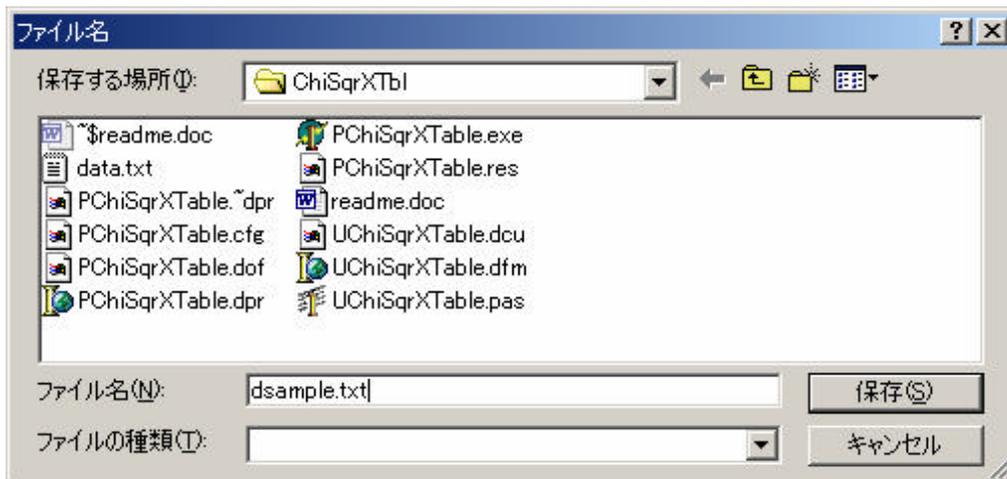


図3 データ保存用ファイル名の設定

データを保存するファイル名の設定後、図3の「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。

ファイルに保存したデータは、「読出」ボタンのクリックで読み込むことができる。「読出」ボタンをクリックすると図4のようなダイアログボックスが表示される。

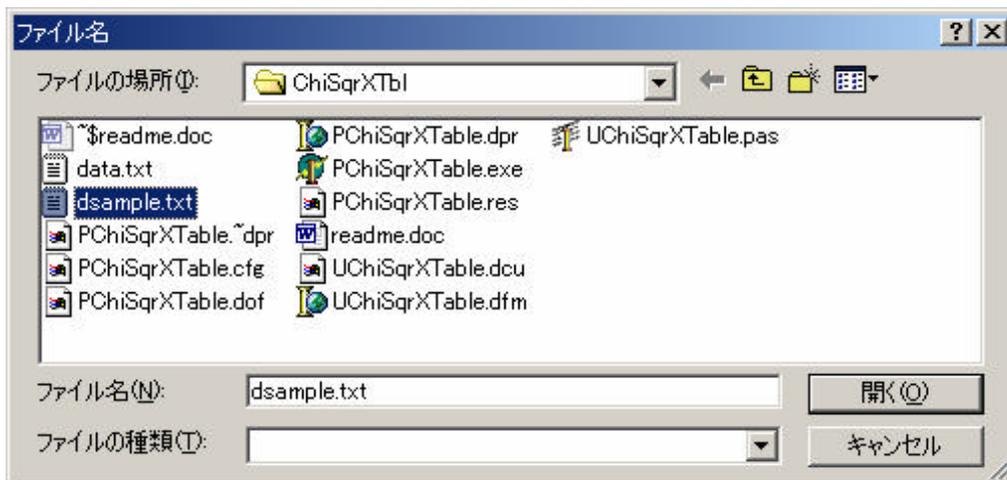


図4 データ読み込みファイル名の設定

ファイル名の設定後、「開く」ボタンをクリックすると設定した名前のファイルからデータが読み込まれる。

図2のようにデータを設定した後、「計算」ボタンをクリックすると計算が始る。「計算」ボタンをクリックすると図5のダイアログボックスが表示される。

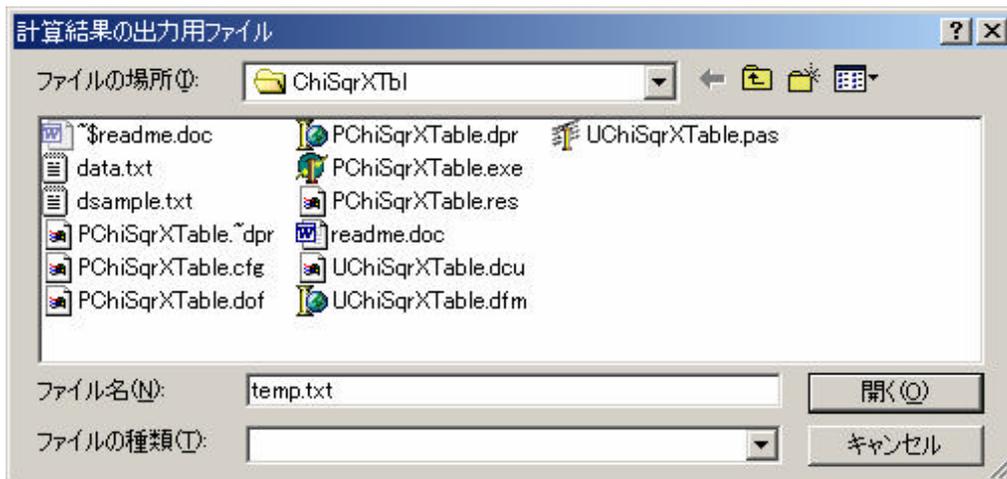


図5 計算結果の出力用ファイル名の設定

図5で設定した名前のファイルに計算結果が書き出される。このファイルはテキストファイルなので、プログラムの実行終了後エディタなどで開いて見ることができる。ファイル名の設定後「開く」ボタンをクリックすると計算が始る。計算が終了すると図6のように「終了」ボタンがアクティブになる（フォーカスをもつ）。

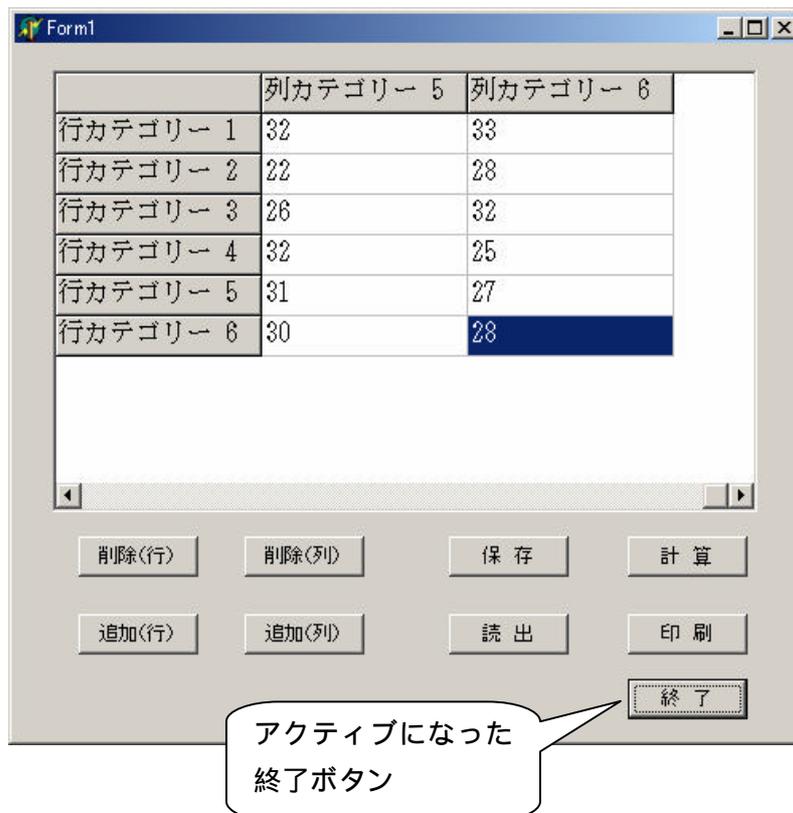


図6 計算終了時のフォーム

「終了」ボタンのクリックでプログラムの実行が終了する。

図2のデータの場合の計算結果の出力ファイルの内容は、リスト1のようになっている。

リスト1 計算結果の出力ファイル

入力データ					
28	32	24	30	32	33
26	29	26	26	22	28
18	19	32	33	26	32
29	27	25	29	32	25
30	37	26	20	31	27
32	22	25	29	30	28
e =					
29.2	29.7	28.3	29.9	31.0	31.0
25.6	26.1	24.8	26.2	27.2	27.2
26.1	26.6	25.3	26.7	27.7	27.7
27.2	27.7	26.4	27.9	28.9	28.9
27.9	28.4	27.0	28.6	29.6	29.6
27.1	27.6	26.2	27.7	28.7	28.7
Chi square = 20.13 df = 25					

リスト1に示されているカイ2乗の値に対するp値は次のようになる。

$$P(\chi_{25}^2 > 20.13) \approx 0.74 = 74\% > 5\%$$

したがって、表1のデータの場合、2つのサイコロAとBの目の出方が独立であるという帰無仮説は有意水準5%で棄却できない。

参考文献

- (1) Gibbons,J.D.(1971). *Nonparametric statistical inference*. McGraw-Hill,Inc.
- (2) 岡本安晴 (1998). *Delphi で学ぶデータ分析法* . CQ 出版社 .
- (3) Siegel,S. & Castellan,N.J.,Jr.(1988). *Nonparametric statistics for the behavioral sciences. 2nd. Ed.* McGraw-Hill,Inc.