

比率の差の検定

k 個の条件の下でのそれぞれのデータの総数を N_i 、 $i = 1, \dots, k$ とおき、 N_i 個のデータの内、あるカテゴリに属するものの数が f_i 個であったとする（表 1）。

表 1 観察度数と総度数データ

条 件	条件 1	・ ・ ・	条件 k
該当度数	f_1	・ ・ ・	f_k
総度数	N_1	・ ・ ・	N_k

このとき、条件により比率 f_i/N_i に差が認められるかどうかの検定を AIC 基準で行う。表 1 のデータに対する AIC の値を求めるため尤度を算出する。尤度は次式で与えられる。

$$L(p_1, \dots, p_k) = \prod_{i=1}^k \binom{N_i}{f_i} p_i^{f_i} (1-p_i)^{N_i-f_i}$$

ここで、 p_i は条件 i における当該カテゴリの観測確率である。対数尤度は次式で与えられる。

$$\begin{aligned} l(p_1, \dots, p_k) &= \log L(p_1, \dots, p_k) \\ &= \sum_{i=1}^k \left\{ \log \binom{N_i}{f_i} + f_i \log p_i + (N_i - f_i) \log(1-p_i) \right\} \end{aligned}$$

最大尤度を与える p_i の値を求めるため対数尤度の偏導関数を求める。

$$\frac{\partial}{\partial p_i} l(p_1, \dots, p_k) = \frac{f_i}{p_i} - \frac{N_i - f_i}{1-p_i}$$

偏導関数を 0 とおいて次式を得る。

$$\begin{aligned} \frac{f_i}{p_i} - \frac{N_i - f_i}{1-p_i} &= 0 \\ f_i(1-p_i) &= (N_i - f_i)p_i \\ p_i &= \frac{f_i}{N_i} \end{aligned}$$

したがって、

$$\hat{p}_i = \frac{f_i}{N_i} \quad (1)$$

と置くと、AIC の値は次式で与えられる。

$$AIC = -2 \times \left[\sum_{i=1}^k \left\{ \log \binom{N_i}{f_i} + f_i \log \hat{p}_i + (N_i - f_i) \log(1 - \hat{p}_i) \right\} \right] + 2 \times k$$

いま、例えば

$$p_1 = p_2 = p_{12}$$

という制約条件を置く。このときの対数尤度は次式で与えられる。

$$\begin{aligned} & l(p_{12}, p_{12}, p_3, \dots, p_k) \\ &= \log \binom{N_1}{f_1} + f_1 \log p_{12} + (N_1 - f_1) \log(1 - p_{12}) \\ &+ \log \binom{N_2}{f_2} + f_2 \log p_{12} + (N_2 - f_2) \log(1 - p_{12}) \\ &+ \sum_{i=3}^k \left\{ \log \binom{N_i}{f_i} + f_i \log p_i + (N_i - f_i) \log(1 - p_i) \right\} \\ &= \log \binom{N_1}{f_1} + \log \binom{N_2}{f_2} + (f_1 + f_2) \log p_{12} + \{(N_1 + N_2) - (f_1 + f_2)\} \log(1 - p_{12}) \\ &+ \sum_{i=3}^k \left\{ \log \binom{N_i}{f_i} + f_i \log p_i + (N_i - f_i) \log(1 - p_i) \right\} \end{aligned} \quad (2)$$

したがって、(2) 式を最大にするパラメタ値は、(1) 式を導いたのと同様の方法で次式で与えられる。

$$\hat{p}_{12} = \frac{f_1 + f_2}{N_1 + N_2} \quad (3)$$

$$\hat{p}_i = \frac{f_i}{N_i}, \quad i = 3, \dots, k \quad (4)$$

上式のパラメタ値に対する AIC の値 AIC_{12} は次式で与えられる。

$$\begin{aligned} AIC_{12} &= -2 \times \left[\log \binom{N_1}{f_1} + \log \binom{N_2}{f_2} + (f_1 + f_2) \log \hat{p}_{12} \right. \\ &\quad \left. + \{(N_1 + N_2) - (f_1 + f_2)\} \log(1 - \hat{p}_{12}) \right. \\ &\quad \left. + \sum_{i=3}^k \left\{ \log \binom{N_i}{f_i} + f_i \log \hat{p}_i + (N_i - f_i) \log(1 - \hat{p}_i) \right\} \right] + 2 \times (k - 1) \end{aligned} \quad (5)$$

すなわち、(3)～(5)式より、2つの条件におけるカテゴリの出現確率を等しいと置いたときの AIC の値は、その 2 条件のデータを併合したときの AIC の値に等しいことがわかる。

ただし、 $\log\left(\frac{N_i}{f_i}\right)$ の項は無視するものとする。

プログラム PAIC1S.dpr は上記の分析を行うための AIC の値を求めるものである。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。

図 1 実行開始時のフォーム

「追加」ボタンのクリックで条件数（列の数）を増やすことができる。不用な列は「削除」ボタンのクリックで削除できる。列の追加・削除はアクティブなセルを含む列を基準に行われるが、セルはクリックによりアクティブになる。

表 2 該当カテゴリの生起度数例

条件	条件 1	条件 2	条件 3
該当度数	15	8	7
総度数	27	23	22

表 2 のデータを分析するときは、まず図 2 のようにデータを設定する。

図 2 3 条件のデータの設定

データの設定後「計算」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。

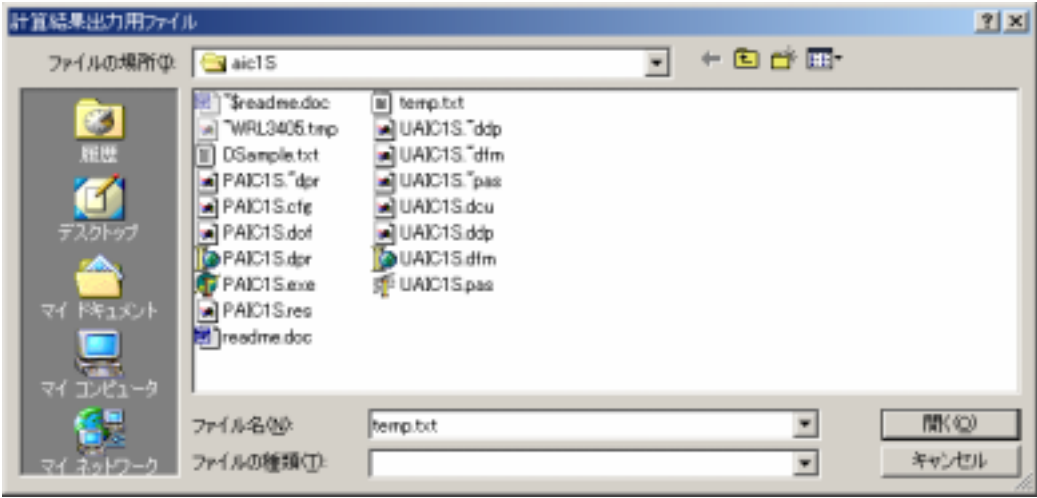


図3 出力ファイル名設定のダイアログボックス

計算結果は図3のダイアログボックスで設定した名前のテキストファイルに書き出される。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まり、計算結果が図4のようにフォームの上部にAICの値として表示される。

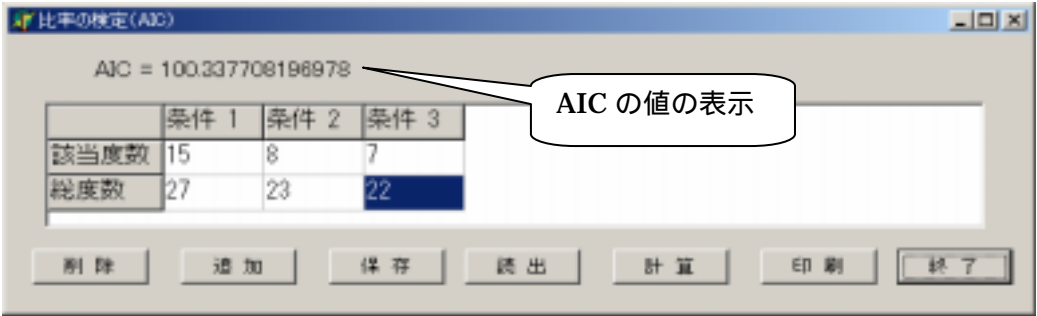


図4 計算結果の表示

図4の結果によると

$$AIC = 100.34$$

であることが判る。

図3で設定した名前の出力ファイルには、リスト1のように結果が書き出されている。

リスト1 出力ファイルの内容			
条件 1	条件 2	条件 3	
15	8	7	
27	23	22	
AIC = 100.337708196978			

まず、設定されているデータが書き出され、続いて図4で表示されている AIC の値が書き出されている。

条件 i における該当カテゴリの観測確率を p_i とおくと、図4の結果は3条件の p_i に制約条件のない場合の AIC の値である。制約条件「 $p_1 = p_2$ 」の下での AIC の値を求めるときは、データを図5のように設定する。

	条件 1	条件 2
該当度数	23	7
総度数	50	22

図5 $p_1 = p_2$ の制約条件での AIC を求めるためのデータの設定

図5の設定では、制約条件「 $p_1 = p_2$ 」に合わせて表2における条件1と条件2のデータが併合されて条件1のデータとして設定され、続いて条件2のデータとして表2における条件3のデータが設定されている。図5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、AIC の値が図6のように表示される。

AIC = 100.515865673785

	条件 1	条件 2
該当度数	23	7
総度数	50	22

図6 $p_1 = p_2$ の制約条件での AIC の値

制約条件「 $p_2 = p_3$ 」の下での AIC を求めるときは図7のようにデータを設定する。

	条件 1	条件 2
該当度数	15	15
総度数	27	45

図7 $p_2 = p_3$ の制約条件での AIC を求めるためのデータの設定

図 7 のデータの場合は、表 2 における条件 2 と条件 3 のデータが併合されて条件 2 のデータとして設定されている。AIC の値は図 8 のようになる。

	条件 1	条件 2
該当度数	15	15
総度数	27	45

図 8 $p_2 = p_3$ の制約条件での AIC の値

制約条件「 $p_1 = p_3$ 」の下での AIC を求めるときは図 9 のようにデータを設定する。

	条件 1	条件 2
該当度数	22	8
総度数	49	23

図 9 $p_1 = p_3$ の制約条件での AIC を求めるためのデータの設定

図 9 では、表 2 における条件 1 と条件 3 のデータが併合されて条件 1 のデータとして設定されている。AIC の値は図 10 のようになる。

	条件 1	条件 2
該当度数	22	8
総度数	49	23

図 10 $p_1 = p_3$ の制約条件での AIC の値

以上の計算結果を表 3 にまとめる。

表 3 AIC の値

制約条件	AIC の値
制約なし	1 0 0 . 3 4
$p_1 = p_2$	1 0 0 . 5 2
$p_2 = p_3$	9 8 . 3 8
$p_1 = p_3$	1 0 1 . 1 4

表 3 より、制約条件「 $p_2 = p_3$ 」における AIC の値が最小である。したがって、表 2 のデータにおいては、条件 2 における該当カテゴリの観測確率 p_2 と条件 3 における観測確率 p_3 は区別できない(等しい)が、条件 1 における観測確率 p_1 はそれらより小さいといえる。