

## 第10章 データの統計解析

数ナビには, "Stats/List Editor" というアプリケーションが標準装備されています。このツールを利用すると確率・統計の主要な計算を簡単に行うことができます。自分で統計上のいろいろなシミュレーションを行うことも容易です。以下では, このアプリケーションの使い方について説明します。

なお, ここでは, すでに分析すべきデータは収集されていること, そして, 前章で説明した表計算を行う "Data/Matrix Editor" の使い方についてある程度知っていること, そして5年の確率統計や品質管理を学習中か学習済みであることを前提とします。

### 10.1 "Stats/List Editor" の概要

確率・統計の学習では, 与えられたデータから平均や分散を求めることが必要です。単純な四則計算が連続するため, まともに計算しようとするとうるさな計算が必要になりますが, その種々の計算部分を数ナビに任せれば, 確率・統計の本質部分に意識を集中させることができるのではないかと思います。

"Stats/List Editor" は, list 形式のデータに対して "Data/Matrix Editor" の操作性を加味し, さらに確率・統計での利用に焦点をおいて機能強化を図ったツールといえるでしょう。以下に, "Stats/List Editor" で可能になることを箇条書きします。次のような機能があります。

#### 統計機能の概要

- (1) 種々の統計計算をさせたとき, その値を保持する変数があり, それらの変数は「statvars」というフォルダーに保存されます。したがって, これらの値を以後の計算でも使用することができます。たとえば1変数の場合は, 次の値が保存されます。2変数の場合は, 変数  $y$  に関する値の他に積和  $\sum xy$  も保存されます。

平均 $\bar{x}$	合計 $\sum x$	2乗和 $\sum x^2$	標本標準偏差 $s_x$	母集団標準偏差 $\sigma_x$
データ数 $n$	最小値 MinX	第1四分位 $q_1$	メディアン Me	第3四分位 $q_3$
最大値 MaxX	偏差 $\sum(x - \bar{x})^2$			

- (2) 豊富な統計回帰機能があります。たとえば, 次のような統計回帰を行うことができ, その変数や残差がフォルダー「starvars」に保存されます。特に, この回帰による予測値と実測値との誤差が, 自動的に list データとして保存されます。

$$y = a + bx$$

$$y = ax^2 + bx + c$$

$$y = a + b \log x$$

$$y = \frac{c}{1 + a \cdot e^{-bx}}$$

$$y = ax + b$$

$$y = ax^3 + bx^2 + cx + d$$

$$y = a \cdot b^x$$

$$y = \frac{a}{1 + b \cdot e^{cx}} + d$$

メディアンメディアン法

$$y = ax^4 + bx^3 + cx^2 + dx + e$$

$$y = a \cdot x^b$$

$$y = a \sin(bx + c) + d$$

重回帰分析 (10 変数まで)

- (3) 種々の乱数を発生させることができます。たとえば、次のような乱数を発生させることができます。左側は、その乱数を発生させるための関数です。

rand83( $n$ ) : 区間 (0, 1) 内の値を持つ  $n$  個の乱数を列データで発生させる。

randInt( $k, m, n$ ) : 整数  $k$  から整数  $m$  ( $k < m$ ) までの間の  $n$  個の整数乱数を列データとして発生させる。

.randNorm( $\mu, \sigma, n$ ) :  $N(\mu, \sigma)$  にしたがう  $n$  個の乱数を列データとして発生させる。

randBin( $n, p, k$ ) : 成功確率が  $p$  ( $0 < p < 1$ ) である試行を  $n$  回行ったときの成功回数を  $k$  個の列データとして発生させる。

- (4) 複数のリストデータ (20 個まで指定可能) を指定して、2 つのリストデータの間の相関係数を計算して行列 (相関行列) を作成することができます。その行列により、複数のデータ間の相互関係を知ることができます。

- (5) いろいろな確率密度関数  $f(x)$  に対して、次の値を求めることができます。これらの値は、フォルダー「Statvars」にも保存されます。

(a)  $x$  の値を指定したとき  $f(x)$  の値

(b)  $x$  の範囲  $[a, b]$  を指定したときの定積分の値  $\int_a^b f(x) dx$

定積分の値だけを表示するか、グラフの該当部分を塗りつぶして表示するかを選択できる。

(c) 定積分の値  $\alpha$  を指定したとき  $\int_{-\infty}^p f(x) dx = \alpha$  となる  $p$  の値

取り扱える確率密度関数 (Probability density function) は、正規分布  $N(\mu, \sigma^2)$ 、二項分布  $B(n, p)$ 、 $t$  分布、 $F$  分布、ポアソン分布、そして幾何分布です。そのコマンドは `CATALOG` にも追加されており、`CATALOG` `F3` で参照することができます。たとえば、(a) は次のような関数により実現できます。(b) は下記の関数の「Pdf」の箇所を「Cdf」に変更することで実現できます。(c) は、正規分布の場合は `Tlstat.invnorm( $\alpha, \mu, \sigma$ )` という関数を利用することになります。

正規分布 `Tlstat.normPdf( $x, \mu, \sigma$ )`

$t$  分布 `Tlstat.tPdf( $x, df$ )`

$\chi^2$  分布 `Tlstat.chi2Pdf( $x, df$ )`

$F$  分布 `Tlstat.FPdf( $x, n, m$ )`

二項分布 `Tlstat.binomPdf( $n, p, x$ )`

ポアソン分布 `Tlstat.poissonPdf( $\lambda, x$ )`

幾何分布 `Tlstat.geomPdf( $p, x$ )`

- (6) 与えられた list データに対して、いろいろな検定を行うことができます。具体的には次のような検定を行うことができ、それぞれについて両側検定、右側検定、そして左側検定のいずれの方法で検定するかを指定できます。 $p$  の値が直接返されるので、有意水準を指定する必要はありません。

平均の検定：母分散既知の場合，母平均未知の場合  
 等平均の検定：母分散既知の場合，母平均未知の場合  
 比率の検定：母比率の検定，母比率の差の検定  
 $\chi^2$  検定：適合度の検定，分割表の検定  
 F 検定：等分散の検定  
 線形回帰  $y = \alpha + \beta x$  の  $\beta$  に関する検定 ( $t$  検定)

(7) 有意水準を自分で指定しているいろいろな区間推定を行うことができます。

母平均の推定：母分散既知の場合，母平均未知の場合  
 平均の差の推定：母分散既知の場合，母平均未知の場合  
 比率の推定：母比率の推定，母比率の差の推定  
 線形回帰  $y = \alpha + \beta x$  の  $\beta$  の推定

(8) 1 要因分散分析や 2 要因分散分析の機能もあります。

## データ形式

[APPS] に登録されている”Data/Matrix Editor”では、「data」「matrix」「list」の3種類のデータを扱うことができます。

data 形式のデータは EXCEL と同様の操作が可能です。99 列 999 行のデータまで扱うことができ、列に計算式を定義して自動計算させることができます。その計算に数式処理を取り混ぜることもできるという意味では、EXCEL より強力な機能といえるでしょう。しかし、列データに対して合計  $\sum x_i$  や平方和  $\sum x_i^2$  などの値を計算させると、単に計算結果を画面に表示するだけであり、その結果は後で利用できるような形では保存されません。

matrix 形式のデータは、計算式を定義することはできません。

list 形式のデータは、1 行だけのデータです。CBL を利用して収集されたデータは、単なる数値列として  $l1$ ,  $l2$ ,  $l3$  などのファイル名で list 形式のデータとして数ナビに転送されます。そのデータを”Data/Matrix Editor”で読み込むと列データとして読み込まれ、列に対して種々の計算を施して列の内容を修正することができます。しかし、その修正結果を再度 list 形式のデータとして保存するには、多少の煩雑な手続きが必要になります。

## 10.2 ”Stats/List Editor”の利用法

### 10.2.1 Editor の起動とデータ入力

#### Editor の起動

数ナビの電源を入れると、デスクトップに”Stats/List Editor”のアイコン(図1)が表示されるので、それを選択して [ENTER] を押します。すでに電源を入れて他の操作中の場合は、[APPS] を押すことでデスクトップ画面になります。[MODE] [F3] で「Apps desktop」の箇所を「Off」にしている場合は(図2)、[APPS] を押してもデスクトップは表示されません。そのときは、[APPS] を押した後に「1: FlashApps」を選択すれば”Stats/List Editor”が選択できます(図3)。



図 1: アイコン

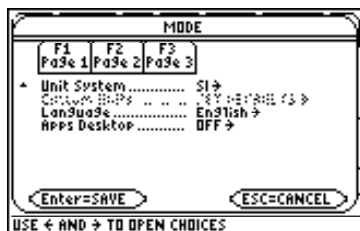


図 2: desktop OFF



図 3: FlashApps

起動すると、最初にフォルダーの選択画面が現れます。確率統計では多数のデータを取り扱うので、ファイルの数が増えていきます。できれば専用のフォルダーを作成して、確率統計に関することはそのフォルダーに保存されるようにした方が良くもありません。

以下では、そのような方針で、仮に「stats」というフォルダーを新規に作成することとして説明します。フォルダーを新たに作成するには、「variable」の箇所には適当なフォルダー名(ここでは、「stats」とした)を入れます。フォルダーを指定すると、List1, List2, ..., List6 からなる空のファイルが表示され(図 4)、指定したフォルダーにこれらのファイルが生成されます。新たにフォルダーを指定しない場合は、通常使用している「main」フォルダーにこれらのファイルが生成されます。



図 4 空のリスト

List1, List2, ... は list 形式のファイル名です。すでにその名前のファイルが存在していれば、その内容が表示されます。この Editor は個々のリスト形式のデータを1つの画面にまとめて表示しているだけであり、「Data/Matrix Editor」のように表全体が1つのファイルになるわけではありません。新たな列を作れば、その列につけた名前でも list 形式のファイルが生成されます。

CBL2 を利用して収集した実験データは、1つのファイルとして保存しておいた方が良くもありません。確率・統計の問題を考えると利用するのであれば、問題ごとにファイルを作らなくても、必要に応じて List1 ~ List6 のファイルに上書きしていけばよいでしょう。

#### データ入力

データを入力するには、個々のセルに1つずつ値を入れていきます。あるいは、上段のファイル名の箇所にカーソルをおいて **ENTER** を押すと、下段に中括弧で囲まれた空のファイルが表示されるので、その括弧内にデータをカンマ区切りで入力していくこともできます(図 5)。

事後の参照例として、次のデータを List1, List2 のデータとして入力してください。

No.	1	2	3	4	5	6	7	8	9	10
List1	20	70	30	45	30	55	85	50	70	60
List2	15	50	25	50	20	50	80	40	85	75

list 形式で保存されている既存のファイルを読み込むには、読み込もうとしている列の最上段にカーソルをおいて **2nd** **[ ]** (var-link) を押します。保存されているファイルの一覧が表示されるので(図 6)、読み込もうとしているファイルにカーソルを当てて **ENTER** を押せばその内容が読み込まれます。



[F5] の箇所では、 $x$  の値を与えて確率密度関数  $f(x)$  の値、範囲  $[a, b]$  を与えたときの確率密度関数の積分値  $\int_a^b f(x) dx$ 、そして、面積  $\alpha$  を与えて  $\int_{-\infty}^p f(x) dx = \alpha$  となる  $\alpha$  の値を求めることができます。



図 9: [F5]



図 10: [F5] 1, 1

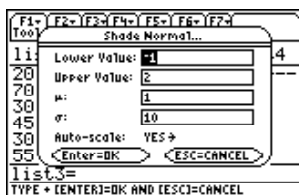


図 11: 範囲入力

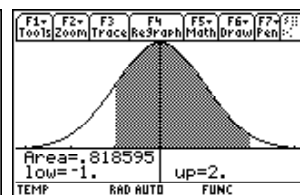


図 12: グラフ描画

### 10.2.5 いろいろな検定 [F6] Tests

帰無仮説を指定しているいろいろな検定を行うには [F6] を利用する (図 13)。検定の仕方は細かく指定することができます。列データからの計算ばかりではなく、でに求められている値を利用した検定を行うこともできます。また、単に結果を表示するばかりではなく、指定されて検定の仕方により棄却域が確率密度曲線のどの部分になるかをグラフ表示することもできます。

たとえば、 $t$  検定を行う場合は [F6]2 を押すと、データのタイプが問われる (図 14)。list 形式のデータを利用するか、それともすでに計算済みの平均値などを利用するかを指定します。「data」利用を指定すると、検定すべき母平均の値と list 形式のデータの入力が求められる (図 15)。両側検定か片側検定であるかを指定する。最後に「Results」で「Draw」を指定すると結果がグラフ表示される (図 16)。この場合は、 $p = 0.226176$  であるので仮説は採択されることが分かります。



図 13: [F6]

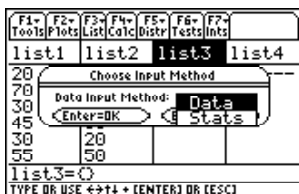


図 14: データタイプ



図 15: 検定条件入力

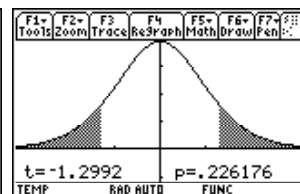


図 16: 結果表示

### 10.2.6 いろいろな推定 [F7] Ints

区間推定は [F7] を利用する (図 17)。指定の仕方は検定の場合と同様である。列データからの計算ばかりではなく、すでに求められている値を利用した区間推定を行うこともできます。信頼度のレベルは、自分で指定します (図 18)。画面に表示される結果 (図 19) はフォルダー statvars に出力されるので、事後に参照することができます。



図 17: [F7]



図 18: 推定条件入力

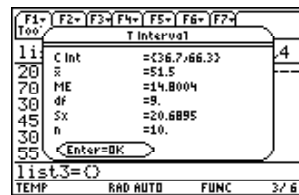


図 19: 結果表示

### 10.2.7 いろいろな確率密度関数

Stats/List Editor は標準添付されている APPS である。この Editor が利用する種々の確率・統計に関するコマンドは、**CATALOG** を押して現れる画面の **F3** (Flash Apps) に登録されています (図 20)。**F3** を押すと、この Editor で利用している種々のコマンドや関数の一覧が表示される (図 21)。カーソルを移動させると、そのコマンドや関数の引数が下段に表示されます。

これらの関数は、基本画面でも普通に利用することができます。たとえば、 $\chi^2$  分布で自由度が増えた場合のグラフを簡単に確かめることができます。図は、 $df = 1, 3, 5$  の場合です。

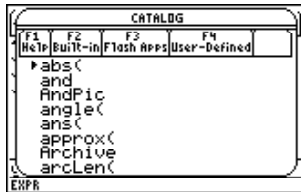


図 20: **CATALOG**

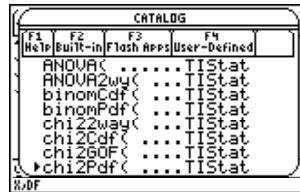


図 21: Flash Apps



図 22: 関数定義

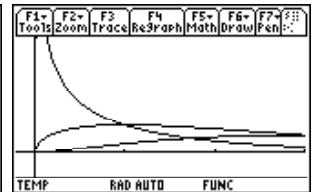


図 23:  $\chi^2$  分布