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煩雑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わ れます。

"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 $ar{x}$	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 σ_x
データ数 n	最小值 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 + b
y = ax² + bx + c
y = a + b log x
y =
$$\frac{c}{1 + a \cdot e^{-bx}}$$

y = a + b log x
y = a + b log x

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(μ, σ, n):N(μ, σ) 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) である試行を <math>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 α 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 (Probavility density function) は,正規分布 N(μ , σ^2),二項分布 B(n,p), t分布, F分布, π 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 TIstat.invNorm(α , μ , σ) という関数を利用することになり ます。

正規分布 TIStat.normPdf (x, μ, σ)	t分布 TIStat.tPdf (x, df)
χ^2 分布 TIStat.chi2Pdf (x, df)	F分布 TIStat.FPdf (x, n, m)
二項分布 TIStat.binomPdf (n, p, x)	ポアソン分布 TIStat.poissPdf (λ, x)
幾何分布 TIStat.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 χ^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 β に関する検定 (t 検定)

(7) 有意水準を自分で指定して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 β 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

[<u>APPS</u>] 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 *l*1, *l*2, *l*3 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)。



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いかもしれません。

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」フォルダーにそ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 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いと思いま す。確率・統計の問題を考えるとき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。





図 6 ファイル選択

10.2.2 列データに関する操作と統計量 F3 List

指定した列に関しては, F3「3: Math」を押すことで最大値,最小値,平均値などの様々な統計 量を求めることができます(図7)。ただし,計算結果を置くセルに事前にカーソルを移動させておく 必要があります。これらの値をまとめて求めるには, F3 ではなく F4 を利用するのがよいでしょ う。求められた値は変数としてフォルダー statvars に保存されます(図8)。

\min	max	mean	median	sum
product	stdDev	variance	stDevPop	varPop

この Editor は,個々の列に何らかの計算式を定義しても,計算式は保存されずに計算結果の数値 列だけが残されます。計算式を残すには,F3「4: Attach List Formula」を利用して式を定義する 必要があります。この計算式を定義しておくと,使用した列の値を変えると,計算式を定義した列の 内容も自動的に修正されます。

10.2.3 いろいろな統計計算 [F4] Calc

基本統計量の計算やいろいろな回帰を行わせる機能は F_4 に登録されています (図 7)。種々の乱数 の発生は , 4: Probavility 」の箇所に登録されていす (図 8)。

また,これらの機能を利用した結果は,statvarsというフォルダーに保存されます(図 6)。



10.2.4 いろいろな確率分布 [F5]

範囲を指定して確率密度曲線とx軸との間の面積 (確率)を求めたり,逆に $(-\infty, p)$ における面積 からpの値を求めるには, F5 を利用してください (図 9)。

たとえば,正規分布 N(μ , σ) で範囲を指定して,その範囲を塗りつぶし曲線下の面積を求めるには, [F5]1の「1: Shade Normal」を利用します (図 10)。 μ , σ の他に範囲を求める画面が現れるので (図 11),必要な値を入力して最後にある「Auto-Scale」で「Yes」を選択すると,グラフが表示されて指 定した範囲が塗りつぶされ,下段にその面積が表示されます (図 12)。 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 α を与えて $\int_{-\infty}^p f(x) dx = \alpha$ となる α の値を求めることができます。



10.2.5 いろいろな検定 [F6] Tests

帰無仮説を指定して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ます (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。



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。



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, χ^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。



このマニュアルは,独立行政法人日本科学技術振興会の平成22年度~24年度科学研究費補助金 基盤研究 (C) 課題番号22500830 (研究代表者:梅野善雄)の支援を受けて作成されたものです。