

College Analysis を使い易くする追加機能

福井正康^{*1}, 細川光浩^{*2}

^{*1} 福山平成大学経営学部経営学科

^{*2} 福山平成大学大学教育センター

要旨 : 我々は社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成し、新しい分析を追加することに報告してきたが、ここでは入力や出力用のツールに対して、使い勝手を向上させるために行った機能の改訂について報告する。また、これまでの分析プログラムに追加したいいくつかの処理についても報告する。

キーワード : College Analysis、ユーティリティ、追加機能

URL : <http://www.heisei-u.ac.jp/ba/fukui/>

1. はじめに

我々は教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた^[1]。しかし、プログラム作成を続けていると、分析で共通に利用するユーティリティの機能不足や分析間の不整合が目につくようになる。今回これらの問題に対処するために、これまで作成してきたプログラムに利用者の使い勝手を向上させる機能の追加を行った。ここではそれについて報告する。

我々が行った主な機能追加は以下の通りである。

- ・ グリッドエディタとテキストエディタの表示桁数を変更できるようにした。
- ・ グリッドエディタの下部にある分析と備考のコメント欄を各ページ独立に変更できるようにした。
- ・ 2次元グラフ表示で、表示後に文字を追加できる機能を加えた。また、グラフの色、文字の大きさ、追加した文字の大きさ、色を変更できるようにした。
- ・ グラフィックエディタの、要素を表す図形と関係を表す線について、簡易的に色や太さを変更できる機能を加えた。また、要素の強調機能とコメントの追加機能を加えた。
- ・ 株式会社朝日ネットの manaba のアンケート集計データを簡単に処理するための2つの機能を加えた。1つは複数選択のカンマ区切りデータを 0/1 型のデータに変換する機能、もう一つは列名 (行名) の先頭に付く "# "を取り除く機能である。
- ・ 数量化 I 類、判別分析、数量化 II 類の結果の表示を、他の分析に合わせ、変数が縦になるように変更した。
- ・ 数学の 3 次元パラメータ関数の中の陰関数表示を新しいアルゴリズムに変更した。

College Analysis を使い易くする追加機能

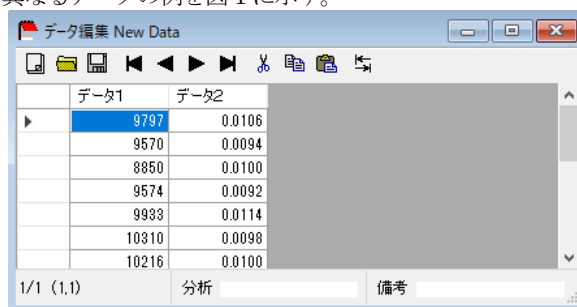
- ・ 量的データの検定のメニューや実験計画法に、平均値を比較するエラーバー付き棒グラフを描くボタンを追加した。
- ・ 複数のヒストグラムを一度に描く場合に積み重ねる表示を追加した。
- ・ 生存時間分析でパラメータ推定後の生存関数、密度関数、ハザード関数を表示する機能を追加した。混合分布の場合は混合された関数を分解して表示することができる。

2. 機能追加の詳細

ここでは節に分けて、新しく追加された機能について説明する。図はその節の中で参照するだけなので、図の番号は節の中で付けることにする。

2.1 グリッド出力とテキスト出力の表示桁数の設定

グリッド出力とテキスト出力の表示方法、特に小数点以下の桁数については、これまで必要な桁数を想定して多めに出力してきた。しかし、初心者に分かり易くということ考えた場合、あまり大きな桁数の表示では問題がある。また、データの大きさによってはさらに精度のよい結果を求められることもある。そこで我々は、グリッド出力については出力後に、テキスト出力については、出力後がテキストの性質上難しいので、設定を変えて再出力する形で対応することにした。デフォルトの出力桁数は、平均値や標準偏差の場合はグリッド表示では小数点以下3桁、テキスト表示では4桁、確率値の場合は、四捨五入して $p < 0.001$ の結果を見極めるために、小数点以下4桁にしている。出力数値によっては、必要性から変更できないものもある。極端に大きさの異なるデータの例を図1に示す。



データ1	データ2
9797	0.0106
9570	0.0094
8850	0.0100
9574	0.0092
9933	0.0114
10310	0.0098
10216	0.0100

図1 データの大きさが異なる例

このデータの基本統計量を求めるとグリッド出力では図2のように表示される。変更を必要とする部分を選択し、グリッド出力のメニュー [表示-書式] を選択すると図3のような設定画面が表示されるので、適当な設定をして「再描画」ボタンをクリックする。

College Analysis を使い易くする追加機能

	データ1	データ2
データ数	17	17
最小値	8229.000	0.008
最大値	11932.000	0.013
平均値	10097.706	0.010
中央値	9933.000	0.010
レンジ	3703.000	0.005
分散	1005600.000	0.000
標準偏差	1002.799	0.001
不偏分散	1068500.000	0.000
標準偏差	1033.662	0.001
標準誤差	243.215	0.000
歪度	-0.146	0.366
尖度	-0.607	0.424

図2 基本統計の出力結果

書式設定

標準(自動) 再描画

小数指定 %表示
 小数点以下 桁

浮動小数点指定(3.15E+4)
 小数点以下 桁

分数 帯分数
 分母(<=10000)

範囲をあらかじめ指定して下さい。

図3 書式設定の画面

この設定によって図4のような結果に変更することができる。

	データ1	データ2
データ数	17	17
最小値	8229	0.00770
最大値	11932	0.01260
平均値	10098	0.00989
中央値	9933	0.01000
レンジ	3703	0.00490
分散	1.006E+06	1.246E-06
標準偏差	1003	0.00112
不偏分散	1.069E+06	1.324E-06
標準偏差	1034	0.00115
標準誤差	243	0.00027
歪度	-0.146	0.36602
尖度	-0.607	0.42444

図4 書式変更結果

テキスト表示の場合は、表示画面の中にデータの詳細な情報を組み込むことができないので、書式を設定する場合は、予め指定しておいて、再出力する。メニュー [分析-テキスト小数点桁数] の中で、小数点以下の桁数か、浮動小数点を選択する。図5に標準の結果、図6に浮動小数点表示の結果を示す。

回帰分析結果

目的変数 データ2

説明変数 データ1

データ数 17

データ2 = 0.0000*データ1 + 0.0146

寄与率 0.175

(重)相関係数 -0.418

図5 標準のテキスト出力

回帰分析結果

目的変数 データ2

説明変数 データ1

データ数 17

データ2 = -4.65118E-7*データ1 + 1.45307E-2

寄与率 0.175

(重)相関係数 -0.418

図6 浮動小数点表示のテキスト出力

変更が適用される数値とそのままの数値があるが、そのままの数値は例えば寄与率や検定確率など、数値の範囲が予め明らかで、変更があまり必要でないと思われる数値である。変更は適用される数値について、すべて同じ書式で出力される。

2.2 グリッドエディタのコメント欄の拡張

グリッドエディタには画面の下に図 1 のようなコメントの入力欄がある。



図 1 グリッドエディタのコメント欄

例えば AHP のプログラムなどは複数のページを使うことから、我々には元々 1 つの分析データに 1 つのファイルという考え方があった。そのためコメント欄は 1 つのファイルに 1 つと考えてきた。しかし後に、1 つのファイルに複数のデータを保存しておく方が扱いやすいということが分かり、ファイルを複数ページの形式で使い始めた。そのためコメントを各ページ独立に付けられるようにしたかったが、エディタのデータはデータ構造の基本であり、修正をするとこれまでのデータ形式に変更を加えることになり、新しいデータ形式は古いバージョンのプログラムで読むことは通常不可能になる。そのため我々は長くデータの形式を変えずにきた。しかし最近、新しい形式のデータでも古いバージョンのプログラムで読むことができるような簡単な方法が分かり、コメント欄をページごとに独立に記述することができるようになった。これは見た目にはあまり変わらないが、プログラムとしては大きな変更になった。

2.3 グラフへの文字の追加等の変更

Excel のグラフの表現力は非常に優れており、我々は見栄えのするグラフについては、データを Excel に与えられるように、グラフ表示のメニューに [編集—データ表示] の機能を付け、Excel 側にデータをコピーして処理することができるようにしていた。しかし、多くの分析を組み込むと、Excel では作れない独自のグラフも作るようになり、College Analysis のグラフにもグラフのタイトルや軸の名前などを付加する機能を追加する必要が生じてきた。そこで我々はグラフィックエディタの手法を用い、グラフに文字列を追加して表示する機能を加えた。また、タイトルを大きくする必要から、追加した文字列の大きさと色を設定できるようにした。まだ機能的には弱いですが、これによって独自のグラフが論文などで使えるようになったと考える。

独自のグラフに文字列を追加した例として図 1 に生存時間分析の Kaplan-Meier グラフとメタ分析のフォレストプロットを示す。

College Analysis を使い易くする追加機能

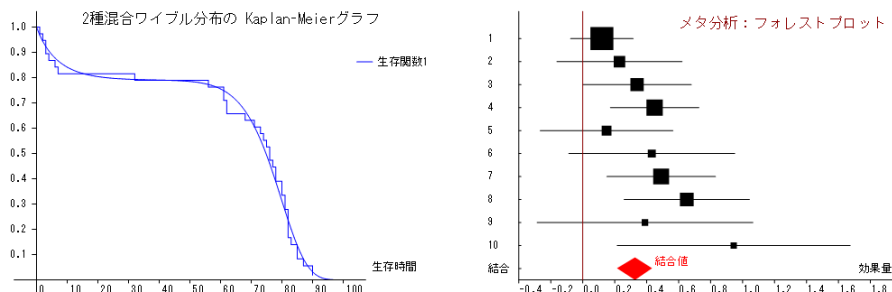


図1 文字列追加とサイズ変更の例

ここで基本のグラフに文字を書き込むには、グラフのメニュー [グラフ表示-文字追加] を選択し、挿入したい場所をクリックする。文字位置はマウスで自由に動かせ、文字のサイズや色は、文字の上で右クリックすることにより、図2のようなメニューが表示され、変更することが可能になる。

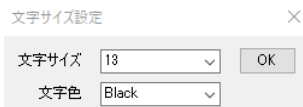


図2 挿入文字列のサイズと色の設定画面

グラフ自身の文字サイズや、グラフの配色はグラフメニューの [設定-文字サイズ・色設定] を選択することによって、変更可能である。その実行画面を図3に示す。



図3 グラフの文字サイズ・色設定

グラフに塗られる色は、予め指定された 19 種類の色パターンの中から選ばれる。但し、各色パターンの中には 16 種類の色が配置されており、その中から「色入替」で前半と後半の 8 種類を交換するので、合計 32 種類（実際は前半と後半同じ色のものがあるので 30 種類）の色パターンから選択できる。

2.4 グラフィックエディタの描画変更機能

グラフィックエディタは、共分散構造分析、デシジョンツリー、PERT など、多くの分析に

College Analysis を使い易くする追加機能

使われているグラフィック入力のツールであるが、黒色で単純にグラフ構造を表すツールであった。その共分散構造分析についての例を図 1 に示す。

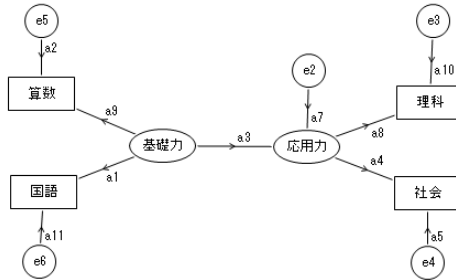


図 1 グラフィックエディタの描画面

この図にもう少し変化をつけるために、図形の線の太さと色を変更する機能を加えた。グラフィックエディタのメニュー [表示-線幅・色設定] を選択すると、図 2 に示す線幅・色設定画面が表示される。グラフィックエディタは、基本的に四角の図形、円形の図形、それを繋ぐ線、で描画される、これらそれぞれの線の幅と色の変更を可能にするのがこの画面である。これを利用すると図 1 は図 3 のように変更される。

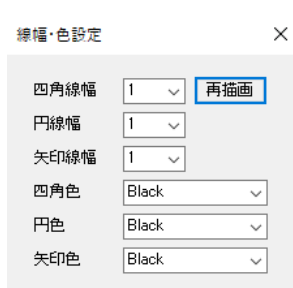


図 2 線幅・色設定画面

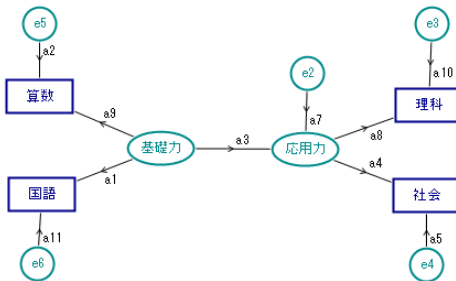


図 3 グラフィックエディタの変更例

これに伴い、重要な部分を黄色で強調する機能とタイトルなどのコメントを挿入する機能をグラフィックエディタのメニュー [表示] の中に加えた。その適用例を図 4 に示す。

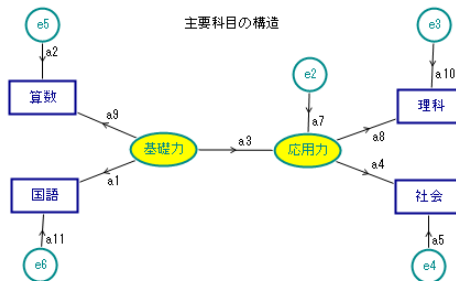


図 4 強調とコメント挿入の例

コメントは画面上では枠が付いているが、コピーするときにはそれがなくなる。ただ、現在は

コメントもエディタの構成要素であるので、分析用には消す必要がある。

2.5 アンケート集計の補助機能

株式会社朝日ネットの提供している manaba は多くの大学で採用されている教育支援システムである²⁾。その機能の一部にアンケート収集・集計機能があるが、そのデータは Excel 形式で保存される。その際、複数回答の問題では、図 1 の枠線のように、選択した回答の番号がカンマ区切りで入力されている。

	# 提出日時	# 回答1.1	# 回答1.2	# 回答1.3	# 回答1.4	# 回答1.5	# 回答1.6	# 回答1.7
5								
6	20xx-05-09 11:47:12	6	1,9	6	3	1,4,5	2	1
7	20xx-05-09 11:46:11	1	1,4	3	1	2,4	3	1
8	20xx-05-09 11:46:11	1	1,4,8	1,2	1	5	1	1
9	20xx-05-09 11:45:11	1	9	6	3	2,5	1	1
10	20xx-05-09 11:46:11	2	1	4	3	3,4,5	3	1
11	20xx-05-09 11:45:11	1	1,4,8	2,3	1	2,4,5	1	3
12	20xx-05-09 11:45:11	3	1,2	5	2	6	1	3
13	20xx-05-09 11:48:12	2	1,4	2	3	4,5	2	2
14	20xx-05-09 11:46:11	3	1,2,3	2,3,4,5	3	2,4,5	1	2
15	20xx-05-09 11:44:11	3	1,2,4	1,2,5	2	1,2,4	3	1

図 1 manaba のアンケート回答形式

統計ソフトでは複数回答の場合、提示された質問項目ごとにチェックしたかどうかの 0/1 回答で入力し、そのデータを元に処理を実行するものが多い。College Analysis でも同様にして処理を行うので、グリッドエディタのメニュー [ツール-0/1 変換[含カンマ]] で、選択して指定された列の前に、0/1 データに変換された複数列のデータを挿入する機能を追加した。

また、図 1 を見ると変数名の最初に「#」記号が付いている。College Analysis では「#」をコメント行やグラフで項目名を非表示にする記号として使うので、これを取り除く機能をエディタに [ツール-列名#除去] または [ツール-行名#除去] として加えている。

2.6 分析による変数列並びの整理

これまでのグリッド出力では分析により、変数の並びが縦になったり、横になったりしていた。元々グリッド出力には縦と横を交換する機能があり、あまり気にすることもなかったが、最初に揃って出力された方が初心者には見易いと考え、行と列の表示の方針を揃えることにした。特に、判別分析、数量化 I 類、数量化 II 類について、表示の並びの縦と横を他の指標の配置も考えながら変更した。図 1 に数量化 I 類の改訂前の表示法、図 2 に改訂後の表示法を示す。

	地域1	地域2	気候1	気候2	気候3	定数項
▶ カテゴリウエイ	3.5167	1.8917	0.0000	-0.3750	-1.4667	0.0000
重回帰 ウェイト	0.0000	-1.6250	0.0000	-0.3750	-1.4667	3.5167
基準化 ウェイト	0.4875	-1.1375	0.6992	0.3242	-0.7675	2.3300
重相関係数	0.968	調整済	0.951			
寄与率	0.937	調整済	0.905			
有効性F値	29.621	自由度	3.6			
参考p値	0.0005					

図1 数量化I類の改訂前の実行結果

	カテゴリウエイ	重回帰ウエイ	基準化ウエイ
▶ 地域1	3.5167	0.0000	0.4875
地域2	1.8917	-1.6250	-1.1375
気候1	0.0000	0.0000	0.6992
気候2	-0.3750	-0.3750	0.3242
気候3	-1.4667	-1.4667	-0.7675
定数項	0.0000	3.5167	2.3300
重相関R	0.968	調整済R	0.951
寄与率R ²	0.937	調整済R ²	0.905
有効性F値	29.621	自由度	3.6
参考p値	0.0005		

図2 数量化I類の改訂前の実行結果

変数が少ない場合は、図1の方が見易いと思うが、変数が多くなると横の表示にも限界がある。

2.7 3次元パラメータ表示関数の陰関数表示

メニュー [分析-数学-グラフ-3次元パラメータ表示関数] で表示される、実行画面の中にある3次元陰関数 $f(x, y, z) = 0$ の表示のアルゴリズムは良い方法が見つからないまま、3方向の2変数関数グラフを重ね合わせるという方法でごまかしてきたが、この度、2次元の等高線を描く方法を3次元に拡張した方法で描くことが可能になった。即ち、3次元空間を直方体のブロックに細かく区切り、 $f(x, y, z)$ の値がその辺の2点で符号を変えるか否かでその辺を横切るかどうかを判断して切断面を描く方法である。図1にその例を示す。

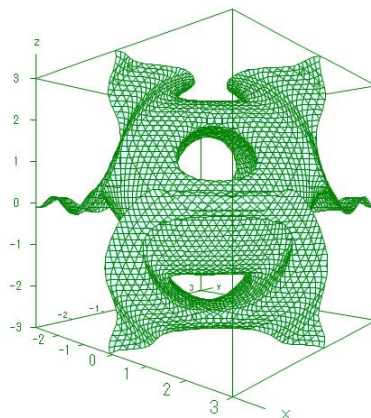


図1 3次元陰関数の例

ここでは切断面のすべての辺を表示しているが、例えば、xy 平面に平行な平面上にある辺だけを表示することもでき、等高線のイメージを出すことも可能である（現在その機能はない）。図の形状から、以前の方法とあまり差がないように思われがちだが、アルゴリズムとしては格段の進歩であり、拡張性も高い。

2.8 各種分析での細かな修正

2.8.1 比較のためのエラーバー付き棒グラフ

2 群間の量的データの比較や 1 元配置実験計画法で、データのグラフによる比較機能が望まれることがある。その場合、エラーバー付きの棒グラフが選ばれることが多いが、このグラフはこれまでのプログラムでも描画可能であった。しかし、すぐに結果を表示できなかったため、量的データの検定実行画面と実験計画法の実行画面の中にボタンを設けることにした。

図 1 に実験計画法の実行画面を示す。赤い線で囲まれたところが平均値の比較グラフのボタンである。エラーバーは標準誤差が用いられる場合と標準偏差が用いられる場合があるので、「偏差」チェックボックスを加えて選択できるようにした。

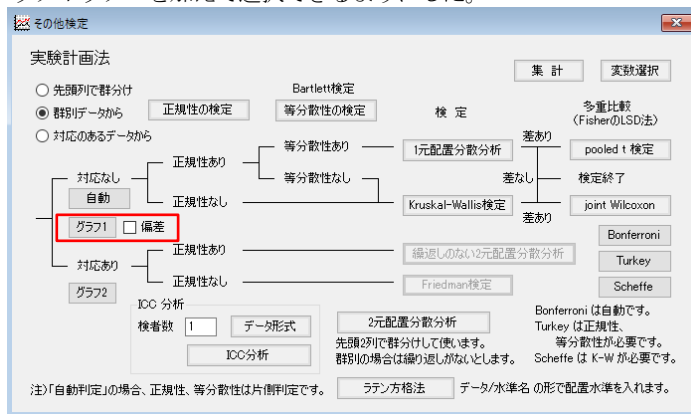


図 1 実験計画法での比較グラフ

対応のない場合とある場合の標準誤差付きの比較グラフを描いた結果が図 2 と図 3 である。

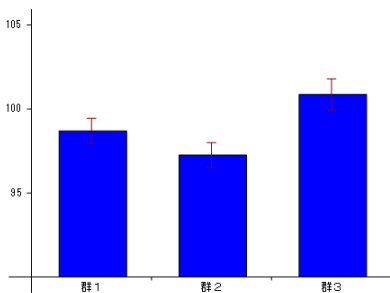


図 2 対応のない場合の比較グラフ

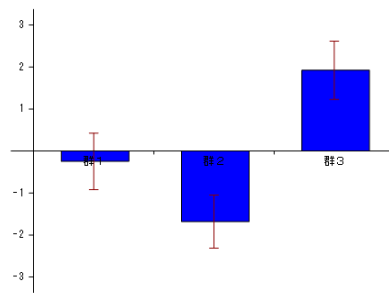


図 3 対応のある場合の比較グラフ

対応のある場合は、レコードごとに、「群 i の値ーすべての群の平均」を計算して平均と標準誤差を取ったものである。

2.8.2 複数変数のヒストグラム

2つの条件でヒストグラムを比較する場合、これまでは横に並べた表示をしていた。これは単に棒グラフを並べたようで、面積グラフであるはずのヒストグラムの意味が曖昧であった。そのため、図4のように量的データの集計実行画面に「積み重ね」チェックボックスを設け、積み重ねる形式でヒストグラムを表示することを試みた。

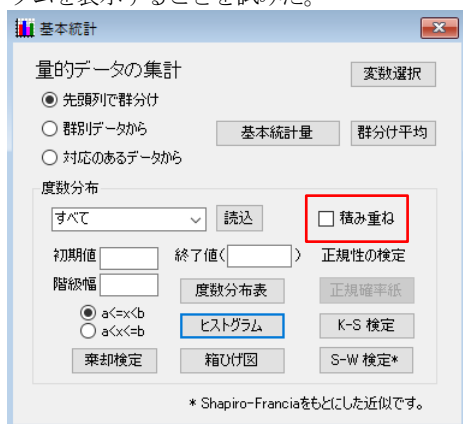


図4 質的データの集計メニュー

図5に横に並べた形式、図6に積み重ねた形式のヒストグラムを描く。横に並べると比較がしやすいが、面積グラフの意味が曖昧になる。また、積み重ねると面積グラフの意味は十分に分かるが、高さの比較が少し見づらい。

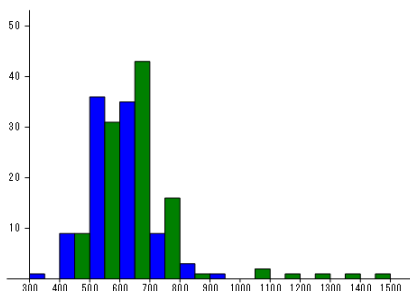


図5 横に並べたヒストグラム

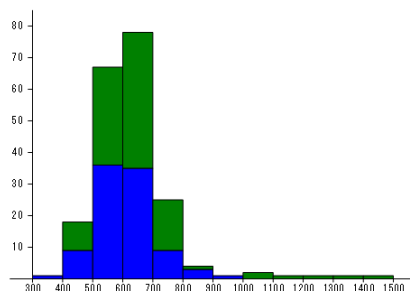


図6 積み重ねたヒストグラム

2.8.3 生存時間分析での予測関数表示

生存時間分析では、データに適合するワイブル分布や混合ワイブル分布を推定する。その際、それらの生存関数や密度関数などを表示することはデータがどのような分布に従うかを知る上で重要である。特に混合分布の場合、どのような生存関数や密度関数の合成になっているのか

ということは重要である。図 7 に今回新しくした生存時間分析の実行画面を示す。選択したデータ形式のラジオボタンに応じて表示が変わっているので 2 種類の画面を提示してある。赤い枠で囲んでいる部分が、関数描画の部分である。

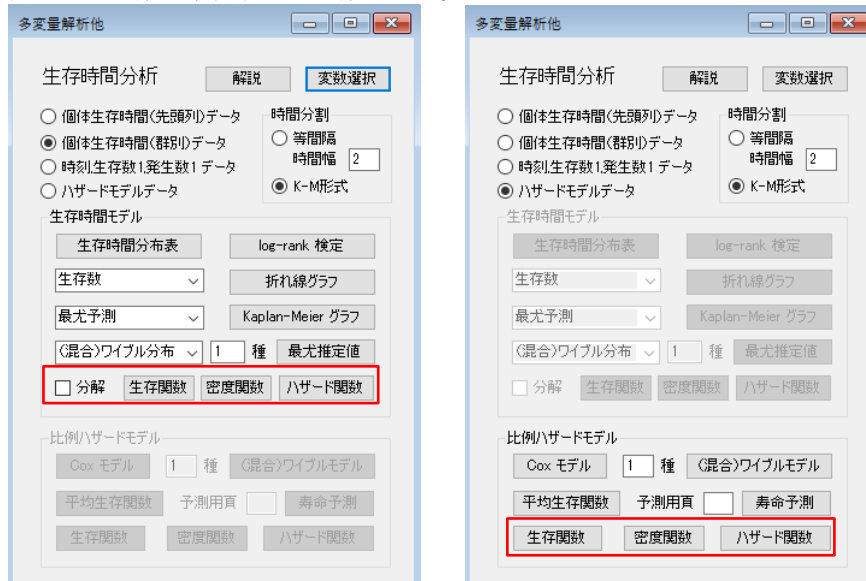


図 7 生存時間分析の実行画面

左側の図で、これらのボタンをクリックした実行例として、図 8 に 2 種混合ワイブル分布の生存関数、密度関数、ハザード関数の例を与える。

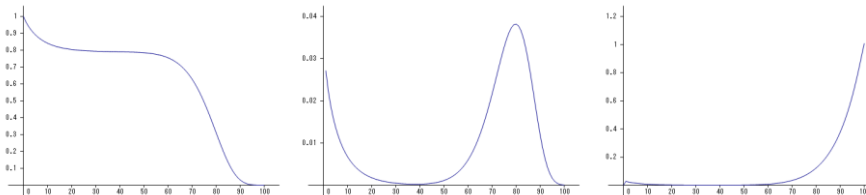


図 8 2 種混合ワイブル分布の生存関数、密度関数、ハザード関数

これを分解したグラフを図 9 に示す。

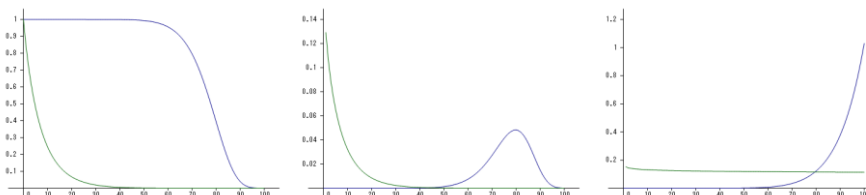


図 9 2 種混合ワイブル分布の分解された生存関数、密度関数、ハザード関数

図 9 の分解に際しては、2 種類の混合の程度を表す確率は掛けず、分解された 2 種類の分布として表示している。

3. おわりに

我々は College Analysis を少しでも使い易くするために、いくつかの細かな機能の改訂や追加を行った。改訂は 2.1 節から 2.6 節のような全体的な機能向上に関するものと、2.7 節と 2.8 節のような具体的な分析での機能追加があるが、後者についてはここで述べたこと以外にも常に改訂を行っている。

グリッド出力とテキスト出力について、今回は 2 度目の改訂である。1 度目の改訂を行った際には気付かなかったが、今回テキスト表示のパラメータの桁数の変更の仕様に問題を感じた。そのため、もう 1 度プログラムを書き直し、ここで述べたように訂正した。

グリッドエディタのコメント欄については、以前から変更の方法を考えていたが、良い方法を思いつかなかった。この度、保存データに全く影響を及ぼさない拡張法を思いついたので変更を加えた。ホームページ上で公開しているために、このようなデータの形式に関する部分は、一度作ると大幅な修正が難しい。

グラフの表現力の問題は、College Analysis の弱点であった。今もそれは変わっていないが、自由に文字が追加できることになって多少は改善された。しかし、例えばタイトルを書こうとすればグラフ全体の位置を下げるなどの配慮が必要であろうが、そのようなことはまだ不可能である。今後も少しずつ変更を加えて行くことになるだろう。

グラフと同様に、College Analysis に独自に組み込まれているグラフィックエディタについても、作成当初、どのぐらいの速さで処理できるのか分からなかった。今ではかなり複雑な構造のものでも十分扱えることが分かっているため、表現力を上げることは可能である。しかし、これにもデザイン性と開発時間という問題がある。この程度の修正しかできないことが残念である。

3 次元パラメータ表示関数についての今回の新しいアルゴリズムは、見た目はあまり変わらないが、大きな進歩であった。今後は、他の 3 次元関数についても、等高線表示など、異なった表示法を導入して行きたいと考える。数学分野での発展はこれからの課題である。

College Analysis ではここで述べられていないような細かい改良や修正も行っている。それによりこれまで眠っていたエラーが発見されることもあり、これは怠ることができない作業である。しかし、逆に修正によりエラーが発生することも考えられる。十分注意をするつもりであるが、利用者の方で、もしそのような異常な動きを見つけられたら、作者宛に知らせていただければ幸せである。

参考文献

- [1] 福井正康, College Analysis リファレンスマニュアル, <http://www.heisei-u.ac.jp/mi/fukui/>
- [2] manaba, 株式会社朝日ネット, <http://manaba.jp/>