

## College Analysis を使い易くする追加機能 2

福井正康<sup>\*1</sup>, 細川光浩<sup>\*2</sup>, 奥田由紀恵<sup>\*2</sup>

<sup>\*1</sup> 福山平成大学経営学部経営学科

<sup>\*2</sup> 福山平成大学大学教育センター

**要旨**：著者ら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、新しく追加した分析の報告を行ってきたが、最近ではプログラムの使い易さを改良することにも力を入れるようになった。ここでは、学生の卒業論文作成に焦点を合わせ、いかに少ない作業で分析結果を論文に載せられるかについて検討した結果を報告する。また、相関分析や回帰分析の改良、データエラーを見つけるための追加検索機能など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

著者ら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 を作成してきたが<sup>1)</sup>、開発を進めるに従って、過去にあまり手を加えていない機能的に弱い部分も見えてくる。そのため最近ではプログラムの使い易さの向上にも力を注いできた<sup>2)</sup>。特に卒業論文での利用を考えて、データ処理に不慣れな学生が論文を作成する際、簡単に使えるように改訂を進めている。プログラムの使い易さは、学術的な価値にはつながらないが、利用者にとっては重要な問題である。この報告では、以下の改良点について説明する。

- ・ 量的データの集計の「先頭列で群分け」において基本統計量の表示順を変更できるようにした。
- ・ 量的データの集計の基本統計量の表示をそのまま卒論に貼り付けることができるように、簡易表示が選べるようにした。
- ・ 複数の 2 次元分割表を同時に表示する出力で、これまで行名は 1 度だけ使っていたが、そのまま切り取って論文に貼り付けられるように、分割表ごとに行名を付ける表示に変更した。
- ・ 複数の変数を選んだ場合の相関係数の表示で、当該の相関係数を与える 2 つの変数間での欠損値除去も選べるようにした。それに伴い、相関係数ごとに異なるデータ数を表示するようになった。

- ・ 相関と回帰分析のメニューで、相関係数や回帰係数の比較検定の部分を切り離し、相関と回帰分析のメニューを初心者に分かり易くした。
- ・ グリッドエディタの検索で、数値以外のデータの検索ができるようにした。
- ・ グリッドエディタでワイルドカードによる検索と置換ができるようにした。
- ・ グリッド出力での行名や列名の編集がそのまま可能なように、表示形式の切り替えができるようにした。

## 2. 基本統計量の表示の変更

メニュー [分析-基本統計-量的データの集計] を選択すると図 1 のような実行画面が表示される。

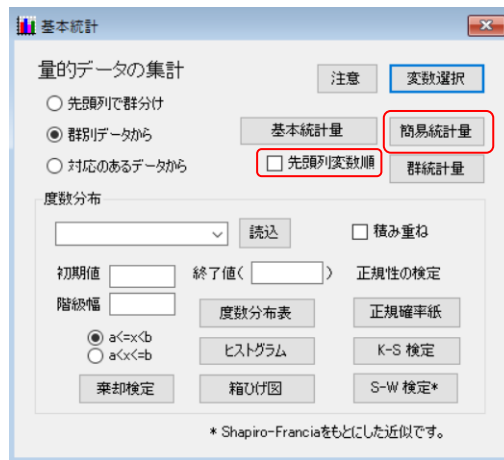


図 1 量的データの集計実行画面

この中の「先頭列で群分け」では、「基本統計量」の表示で、デフォルトとして図 2 のように群分け変数、変数の順に表示が並ぶようになっている。

	地域1>年取	地域1>支出	地域2>年取	地域2>支出
データ数	94	94	106	106
最小値	356.000	11.000	426.000	13.000
最大値	986.000	75.000	1432.000	85.000
平均値	608.149	45.106	650.821	48.283
中央値	603.000	46.000	618.000	48.500
レンジ	630.000	64.000	1006.000	72.000
分散	9254.999	185.201	27054.260	217.524
標準偏差	96.203	13.609	164.482	14.749

図 2 先頭列で群分け (デフォルト)

しかしこの形式では、変数が多い場合、変数を群別に比較したいとき、あまり見易いとはいえない。比較は隣り合った方が分かり易いので、メニューに「先頭列変数順」というチェックボックスを設け、チェックを入れると図 3 のように、変数、群分け変数の順に表示が並ぶようにした。比較したいものが隣に並ぶので見易くなる。

## College Analysis を使い易くする追加機能 2

	地域1>年収	地域2>年収	地域1>支出	地域2>支出
データ数	94	106	94	106
最小値	856.000	426.000	11.000	13.000
最大値	986.000	1432.000	75.000	85.000
平均値	608.149	650.821	45.106	48.283
中央値	603.000	618.000	46.000	48.500
レンジ	630.000	1006.000	64.000	72.000
分散	9254.999	27054.260	185.201	217.524
標準偏差	96.203	164.482	13.609	14.749

図 3 先頭列で群分け (先頭列変数順)

基本統計量の表示は様々な統計量が出力されるが、基礎的な報告書などでは、データ数、平均値、標準偏差が重要である。そのため、図 1 の量的データの集計メニューで、新たに「簡易統計量」ボタンを設けた。「群別データから」と「先頭列で群分け」の出力結果を図 4a と図 4b に示す。

	N	平均値	標準偏差
▶ 年収	200	630.765	138.702
支出	200	46.790	14.348

図 4a 群別の簡易統計量

群分け 地域	1>N	平均値	標準偏差	2>N	平均値	標準偏差
▶ 年収	94	608.149	96.719	106	650.821	165.263
支出	94	45.106	13.682	106	48.283	14.819

図 4b 先頭列で群分けの簡易統計量

これは結果を示す最低限の表示である。卒論などでも使う機会が多いと考える。

### 3. 複数分割表の表示法の変更

これまでメニュー「分析-基本統計-質的データの集計」の中の「分割表の作成」ボタンでは、3つ以上の変数を選択した場合、先頭の変数によって群分けされ、図 5a のように後ろの変数の分割表が示されていた。

	意見1-1	意見1-2	合計	意見2-1	意見2-2	意見2-3	合計
▶ 地域-1	32	62	94	33	26	35	94
地域-2	53	53	106	38	30	38	106
合計	85	115	200	71	56	73	200

図 5a 3つ以上の変数選択の分割表 (旧)

先頭の変数は共通であるから、1回表示されればよいと考えてきたが、この中の1つ (例えば後ろの意見2) を取り出して報告書に貼り付けることは意外に面倒である。そのため、今回の改訂では、図 5b のように、間に先頭の項目をはさんで表示することにした。

	意見1:1	意見1:2	合計	地域1	意見2:1	意見2:2	意見2:3	合計
▶ 地域1	32	62	94	地域1	33	26	35	94
地域2	53	53	106	地域2	38	30	38	106
合計	85	115	200	合計	71	56	73	200

図 5b 3つ以上の変数選択の分割表 (新)

これによって、表示は重複するが、Excel に貼り付け、個々の分割表を取り出そうとした場合、その部分だけを抜き出すことができ処理が非常に簡単になる。何のために分割表を並べて表示するかを考えた場合、当然図 5b のような表示にするべきであった。

## 4. 相関と回帰分析の変更

### 4.1 欠損値の除去

メニュー [分析 - 基本統計 - 相関と回帰分析] を選択すると、図 6 のような新しくなった実行画面が表示される。

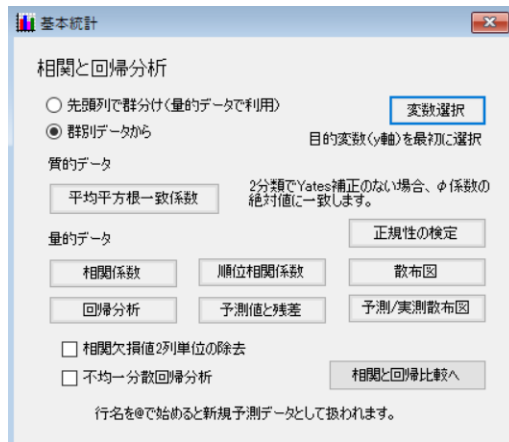


図 6 相関と回帰分析実行画面

元々の相関と回帰分析のプログラムでは、3 つ以上の変数を選び、「相関係数」ボタンをクリックすると、図 7a のような形式で結果が出力されていた。

相関係数	卒業試験	入試点数	内申点数	勉強時間	出席率
▶ 卒業試験	1.000	0.404	0.074	0.895	0.926
入試点数	0.404	1.000	0.265	0.378	0.318
内申点数	0.074	0.265	1.000	-0.018	0.006
勉強時間	0.895	0.378	-0.018	1.000	0.898
出席率	0.926	0.318	0.006	0.898	1.000
検定確率					
卒業試験		0.0072	0.6375	0.0000	0.0000
入試点数	0.0072		0.0862	0.0125	0.0376
内申点数	0.6375	0.0862		0.9072	0.9711
勉強時間	0.0000	0.0125	0.9072		0.0000
出席率	0.0000	0.0376	0.9711	0.0000	

図 7a 3 変数以上の相関係数表示 (旧)

この形式の問題点はいくつかある。1 つは欠損値の除去がレコード単位に固定されていたことである。しかし、データを有効に活用しようと考えれば、相関を取る 2 つの変数ごとに欠損値を抜き出し、除去することが望ましい。これは利用者の方からの指摘であった。また、データ数が表示されていないため、検定確率の根拠が明確に示せない。

以上のことから、新しいメニューでは、「相関欠損値 2 列単位の除去」チェックボックスを付けた。これにチェックが入っている (デフォルト) と図 7b のような結果が表示される。

## College Analysis を使い易くする追加機能 2

	卒業試験	入試点数	内申点数	勉強時間	出席率
相関係数					
卒業試験	1.000	0.412	0.131	0.888	0.917
入試点数	0.412	1.000	0.267	0.365	0.328
内申点数	0.131	0.267	1.000	-0.007	0.027
勉強時間	0.888	0.365	-0.007	1.000	0.880
出席率	0.917	0.328	0.027	0.880	1.000
両側確率					
卒業試験		0.0040	0.3803	0.0000	0.0000
入試点数	0.0040		0.0693	0.0137	0.0246
内申点数	0.3803	0.0693		0.9661	0.8553
勉強時間	0.0000	0.0137	0.9661		0.0000
出席率	0.0000	0.0246	0.8553	0.0000	
データ数					
卒業試験	48	47	47	44	47
入試点数	47	48	47	45	47
内申点数	47	47	49	45	48
勉強時間	44	45	45	46	46
出席率	47	47	48	46	49

図 7b 3 変数以上の相関係数表示 (新)

これは相関係数を求める 2 変数間だけで欠損値を取り除いた結果である。この方法では、変数の組ごとに、利用するデータ数が異なるため、データ数の項目を新たに設けている。また、レコード単位の除去の場合も、全体を 1 行下にずらして、空いた先頭の行にデータ数を表示するように訂正した。

## 4.2 相関係数と回帰式の比較

これまで College Analysis の相関と回帰分析では、相関係数と回帰係数は 0 との比較の場合だけを考えてきた。しかし相関係数や回帰式を他の値と比べることも多くなると考え、指定値との比較や 2 群間の比較の検定を相関と回帰分析のメニューに加えた。ただ相関と回帰分析は初心者も多用する手法であるので、メニューが複雑になり過ぎることは控えたい。そこで、著者らは相関と回帰分析のメニューの中に入れていた相関係数と回帰分析の比較のメニューを別途作り、図 6 の「相関と回帰比較へ」ボタンで図 8 に示す実行画面を表示するようにした。

相関と回帰分析

相関と回帰比較

先頭列で群分け
 変数選択

群別データから(2列ずつ)
 目的変数(y軸)を最初に選択

指定値との比較

比較値 r

比較値 a  b

簡易計算 n  r  r0  → p

2群間の比較

簡易計算 n1  r1  n2  r2  → p

図 8 相関と回帰分析実行画面

データの種類として群別データを使う場合は、指定値との比較の検定が利用でき、先頭列で群分けを使う場合は、指定値との比較の検定と2群間の比較の検定が両方利用できる。

図9の「専門試験」と「SPI」のデータを用いて、「群別データから」とし、「指定値との比較」のグループボックス内で、比較値を $r=0.5$ として「相関係数比較」ボタンをクリックすると図10のような分析結果が表示される。

	判定	専門試験	SPI
▶ 1	2	55	40
2	2	57	45
3	2	63	47
4	1	71	59
5	1	72	61
6	1	64	53
7	2	59	42
8	2	42	45
9	1	76	66

図9 相関係数と回帰係数の比較データ

指定相関係数との比較結果

変数1: 専門試験  
 変数2: SPI  
 データ数: 40  
 実測相関係数: 0.711  
 指定相関係数: 0.500  
 z統計値: 2.0624  
 両側確率P: 0.0392  
 有意水準 $\alpha$ : 0.05  
 P< $\alpha$ より、指定相関係数と差があるといえる。

図10 相関係数の指定値との比較結果

すべての変数を選び、「先頭列で群分け」として、「指定値との比較」グループボックス内で、「相関係数比較」ボタンをクリックすると図11のような分析結果が表示される。

指定相関係数との比較検定

変数1: 専門試験  
 変数2: SPI

群#1  
 データ数: 19  
 実測相関係数: 0.854  
 指定相関係数: 0.500  
 z統計値: 2.8896  
 両側確率P: 0.0039  
 有意水準 $\alpha$ : 0.05  
 P< $\alpha$ より、指定相関係数と差があるといえる。

群#2  
 データ数: 21  
 実測相関係数: 0.454  
 指定相関係数: 0.500  
 z統計値: -0.2538  
 両側確率P: 0.7936  
 有意水準 $\alpha$ : 0.05  
 P $\geq$  $\alpha$ より、指定相関係数と差があるといえない。

図11 相関係数の比較結果

次に図9の「専門試験」と「SPI」のデータを選び、「群別データから」とし、「指定値との比較」グループボックス内で、比較値 $a$ を0.5、 $b$ を20として「回帰式比較」のボタンをクリックすると図12のような分析結果が表示される。

## College Analysis を使い易くする追加機能 2



図 12 回帰式の指定回帰式との比較

すべての変数を選び、「先頭列で群分け」として、「2 群間の比較」グループボックス内で、「相関係数比較」のボタンをクリックすると、図 13 のような結果が表示される。

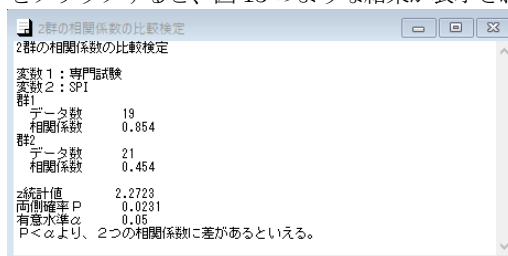


図 13 2 群の相関係数の比較結果

同じく、「回帰式比較」のボタンをクリックすると、図 14 のような結果が表示される。

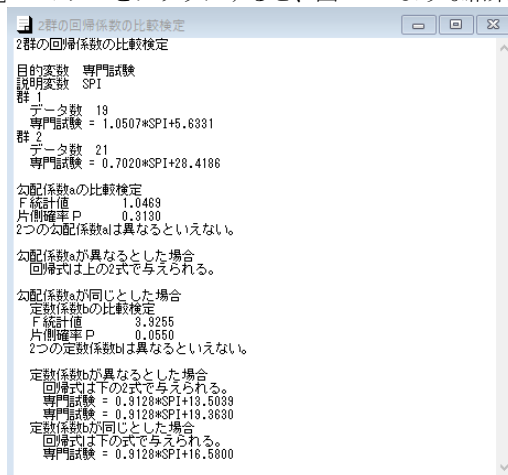


図 14 2 群の回帰式の比較結果

以下に上のプログラムで用いた、相関係数と回帰分析の比較の理論を簡単に述べておく<sup>[3]</sup>。相関係数と母相関係数の比較では、データ数を  $n$ 、標本相関係数を  $r$ 、母相関係数を  $\rho$  として、以下の関係を利用する。

$$T = \frac{\frac{1}{2} \log \frac{1+r}{1-r} - \frac{1}{2} \log \frac{1+\rho}{1-\rho}}{1/\sqrt{n-3}} \sim N(0,1)$$

2群の相関係数の比較では、データ数を $n_1, n_2$ 、標本相関係数を $r_1, r_2$ として、以下の関係を利用する。

$$T = \frac{\frac{1}{2} \log \frac{1+r_1}{1-r_1} - \frac{1}{2} \log \frac{1+r_2}{1-r_2}}{\sqrt{1/(n_1-3)+1/(n_2-3)}} \sim N(0,1)$$

回帰係数と母回帰係数の比較では、データ数を $n$ 、標本回帰式を $y = ax + b$ 、母回帰式を $y = \alpha x + \beta$ として、以下の関係を利用する。

勾配係数の比較  $T_a = (a - \alpha) \sqrt{SS_x / V_E} \sim t_{n-2}$

定数係数の比較  $T_b = \frac{b - \beta}{\sqrt{V_E (1/n + \bar{x}^2 / SS_x)}} \sim t_{n-2}$

ここに

$$\bar{x} = \frac{1}{n} \sum_{\lambda=1}^n x_{\lambda}, \quad \bar{y} = \frac{1}{n} \sum_{\lambda=1}^n y_{\lambda}$$

$$SS_x = \sum_{\lambda=1}^n x_{\lambda}^2 - n\bar{x}^2, \quad SS_y = \sum_{\lambda=1}^n y_{\lambda}^2 - n\bar{y}^2, \quad SS_{xy} = \sum_{\lambda=1}^n x_{\lambda} y_{\lambda} - n\bar{x}\bar{y}$$

$$V_E = \frac{1}{n-2} \left[ SS_y - (SS_{xy})^2 / SS_x \right]$$

2群の回帰係数の比較では、データ数を $n_1, n_2$ 、標本回帰式を $y = a_1 x + b_1$ ,  $y = a_2 x + b_2$ として、まず、以下の関係を利用して勾配係数の比較を行う。

$$F_a = \left[ (\Delta_2 / \Delta_1) - 1 \right] (n_1 + n_2 - 4) \sim F_{1, n_1 + n_2 - 4}$$

勾配係数が異なるとすると、回帰式はそのまま使われ、勾配係数が等しいとすると、以下の関係を利用して定数係数の比較を行う。

$$F_b = \left[ (\Delta_3 / \Delta_2) - 1 \right] (n_1 + n_2 - 3) \sim F_{1, n_1 + n_2 - 3}$$

ここで、定数係数が異なるとすると  $a = (SS_{xy1} + SS_{xy2}) / (SS_{x1} + SS_{x2})$ ,  $b_i = \bar{y}_i - a\bar{x}_i$ として、回帰式は以下を与える。

$$y = ax + b_1, \quad y = ax + b_2$$



定数係数が同じとすると  $a = SS_{xy}/SS_x$ ,  $b = \bar{y} - a\bar{x}$  として、回帰式は同じとして以下で与える。

$$y = ax + b$$

ここに、 $i = 1, 2$  として以下の関係を用いた。

$$\bar{x}_i = \frac{1}{n_i} \sum_{\lambda=1}^n x_{i\lambda}, \quad \bar{y}_i = \frac{1}{n_i} \sum_{\lambda=1}^n y_{i\lambda}$$

$$SS_{x_i} = \sum_{\lambda=1}^n x_{i\lambda}^2 - n_i \bar{x}_i^2, \quad SS_{y_i} = \sum_{\lambda=1}^n y_{i\lambda}^2 - n_i \bar{y}_i^2, \quad SS_{xy_i} = \sum_{\lambda=1}^n x_{i\lambda} y_{i\lambda} - n_i \bar{x}_i \bar{y}_i$$

$$\Delta_1 = \left[ SS_{y_1} - (SS_{xy_1})^2 / SS_{x_1} \right] + \left[ SS_{y_2} - (SS_{xy_2})^2 / SS_{x_2} \right]$$

$$\Delta_2 = SS_{y_1} + SS_{y_2} - \frac{(SS_{xy_1} + SS_{xy_2})^2}{SS_{x_1} + SS_{x_2}}$$

$$\Delta_3 = SS_y - (SS_{xy})^2 / SS_x$$

## 5. 検索機能の強化

グリッドエディタのメニュー [ツールー検索] と [ツールー置換] で表示される検索と置換の実行画面を図 15 と図 16 に示す。

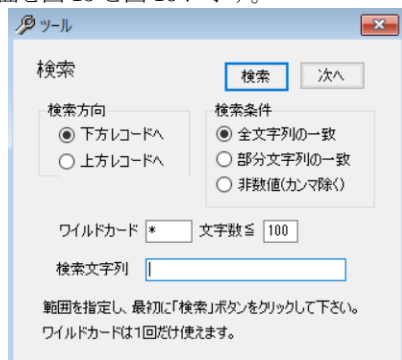


図 15 検索実行画面

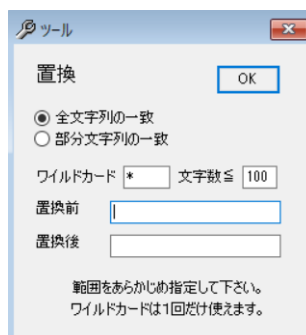


図 16 置換実行画面

検索の実行画面に新たに追加された機能が非数値の検索である。これまでの経験から、卒業論文などのデータ入力には必ずと言っていいほど間違いがあり、その発見が意外と面倒であった。そのため、検索条件のグループボックスに「非数値(カンマ除く)」ラジオボタンを加えた。これを選択すると、検索文字列などとは無関係に数値以外のデータを検索する。但し、数値の中にカンマが含まれている場合は、その位置に関わらず、Visual Basic の関数が数値と判断するため、数値と判断をする。検索は、検索範囲を選択したのち、最初の 1 回は「検索」ボタンをクリックし、その後から「次へ」ボタンをクリックして求めて行く。

次に、これまではなかったワイルドカードも、検索と置換の文字列中で使えるようにした。ワイルドカードはデフォルトでは「\*」であるが、変更も可能である。また、対応する文字数も設定できるようになっている。例えば、実際にあった入力ミスであるが、「123/34」などのデータから「/」より前の部分を取り出す必要が生じた場合、置換前を「\*/」、置換後を「\*」とすると、ワイルドカード「\*」に「123」が対応し、置換後に「123」が取り出されることになる。本来のワイルドカードの使用はもっと綿密な処理が必要であるが、ここでは簡易的な使い方として取り入れている。

## 6. グリッド出力での表示枠の切り替え

グリッド出力では、行名や列名の部分をコピーしようと思う場合、そのままではコピーをすることができない。そのため、コピーのためだけに一度グリッドエディタに貼り付けるか、Excelに貼り付けることになるが、少し手間を感じる。また行名や列名を直接編集したい場合も同様である。そこで、グリッド出力での表示枠を切り替える機能をグリッド出力に追加した。グリッド出力は通常、図 17.1 のように表示されるが、グリッド出力には行名と列名の編集機能がないために、グリッド出力のメニュー [編集-表示枠変更] を選択することによって、図 17.2 のように変更される機能を追加した。



	地域1の年取	地域1の支出	地域2の年取	地域2の支出
データ数	94	94	106	106
最小値	356.000	11.000	426.000	13.000
最大値	986.000	75.000	1432.000	85.000
平均値	608.149	45.106	650.821	48.283
中央値	603.000	46.000	618.000	48.500
レンジ	630.000	64.000	1006.000	72.000
分散	9254.999	195.201	27054.260	217.524
標準偏差	96.203	13.609	164.482	14.749
不偏分散	9354.515	187.193	27311.920	219.595
標準偏差	96.719	13.682	165.263	14.819
標準誤差	9.923	1.404	15.976	1.433
四分位数1	543.500	36.250	565.000	40.000
四分位数3	650.500	55.750	684.750	57.000
歪度	0.746	-0.219	2.612	0.068
尖度	1.860	-0.571	8.473	0.115

図 17.1 グリッド出力



	地域1の年取	地域1の支出	地域2の年取	地域2の支出
データ数	94	94	106	106
最小値	356.000	11.000	426.000	13.000
最大値	986.000	75.000	1432.000	85.000
平均値	608.149	45.106	650.821	48.283
中央値	603.000	46.000	618.000	48.500
レンジ	630.000	64.000	1006.000	72.000
分散	9254.999	195.201	27054.260	217.524
標準偏差	96.203	13.609	164.482	14.749
不偏分散	9354.515	187.193	27311.920	219.595
標準偏差	96.719	13.682	165.263	14.819
標準誤差	9.923	1.404	15.976	1.433
四分位数1	543.500	36.250	565.000	40.000
四分位数3	650.500	55.750	684.750	57.000
歪度	0.746	-0.219	2.612	0.068
尖度	1.860	-0.571	8.473	0.115

図 17.2 グリッド出力の枠変更

通常の数値と同じ枠に表示されると、コピーも編集も自由である。また元の枠に戻したい場合は、同じメニュー [編集-表示枠変更] をもう一度クリックする。ただこの変更では表示された文字列はそのまま移るが、小数点以下の桁の表示されていないデータは使えなくなるので、注意を要する。

## 7. おわりに

College Analysis を使い易くする機能として、この報告では、基本統計量の表示法の変更、複数項目を選択した場合の分割表の表示法の変更、相関と回帰分析について、欠損値除去法の変更と相関係数と回帰式の比較メニューの独立化、検索機能の強化、グリッド出力での表示枠

## College Analysis を使い易くする追加機能 2

の変更機能について説明した。これらの変更の背景には、卒業論文などの指導を行い易くし、報告書を書き易くするという目的がある。漫然と機能強化を考えても思い浮かばない改良点が、このような方向性を考えることで見えてくる。

基本統計量の群分けの出力は何のために行うかを考えると、平均は異なる群で並べて表示した方が違いは明白になる。そのため表示順を変えることも考えなければならない。多くの分割表を並べて出力する場合も、結果が分かればよいのではなく、結果が使えることが重要である。そのため今回の変更は重要である。

相関係数の出力については、利用者からの提案で改訂を行ったが、作成した当初は、レコード単位の除去を行って、必要があれば2変数だけで見ればよいと考えていた。これも出力結果をそのまま活用するという観点からすると大きな思い違いである。相関と回帰直線の比較のメニューの独立化については、利用者が初心者の場合を考えての措置であるが、メニューの見た目がきれいに収まったので成功ではないかと考えている。

検索機能の強化については、学生指導の実践から必要性を感じた。元々質的データの入力ミスについて分割表を用いて調べ、その位置を特定するために検索を用いていたが、これでは予測できない値が入力された場合の量的データに対して通用しない。より早い対応を求められた結果、この方法を考えた。データを処理する段階で、数値以外を選び出してもよいが、エラーを連続して見つけるには検索の中に含める方が便利である。

最後にグリッド出力の枠の変更も、学生の指導中に考えた。グリッド出力の変数名を利用しなければならぬことが起こり、不便を感じた結果の改良である。元々、グリッドエディタには、行名と列名を編集する機能もあったし、全コピーと全貼り付け、コピーと貼り付けという2つの編集方法があり、これによって表示枠の変更も可能であった。しかし、グリッド出力についてはこのような機能がなかったため、最も手軽に編集ができる方法を導入した。

今後様々な機能拡張や機能変更がなされると思うが、これからは卒業論文や報告書を作るときにできるだけ手間をかけずにできること、学生がどこで困るかを第一に考えると、道筋が見えてくるものと思う。

## 参考文献

- [1] 福井正康, College Analysis 総合マニュアル,  
<http://www.heisei-u.ac.jp/ba/fukui/analysis.html>
- [2] 福井正康, 細川光浩, College Analysis を使い易くする追加機能, 福山平成大学経営研究, 第14号, (2018) 107-118
- [3] 丹後俊郎, 古川俊之監修, 新版 医学への統計学, 朝倉書店, 1993.