

基礎から学ぶシリーズ 3

College Analysis で学ぶ 集計と検定

福井正康

福山平成大学経営学部経営学科

はじめに

このシリーズ、基礎からの統計学では、データの集計方法と検定・推定について少し理論に踏み込んで勉強しました。その際処理はすべて Excel を使い、何を計算しているのか分かるようにしました。ただこの本は経済・経営系の大学院に進もうとする人に基礎を学んでもらう目的で作ったもので、実用を目的とした人向きではありません。

そこで我々は徹底した実用を目的に本を作ることにしました。計算はすべて統計処理のソフトを使い、解説はできるだけやさしく、初心者に応用力をつけることが目的です。数式も定義をはっきりさせるために出てきますが、もちろん飛ばしてもらっても結構です。とにかく「習うより慣れろ」でやってみて下さい。最後には統計処理を見渡せる力がつくことと思います。

さて、統計ソフトにはいろいろなものがあります。SPSS, SAS, S-PLUS, R のように世界的に評価されているものや比較的使い易い STATISTICA, R-Commander 等、数多くのものが開発されています。これらの単独ソフトの他にも Excel の機能を利用するために VBA で記述されたマクロ的なソフトもあります。どれを利用するかは個人の好みでしょうが、一般に上中級者用のものは非常に高価で、初心者用のものでもある程度費用がかかります。またフリーのものでも、R は文系の学生にはちょっと難しいし、他のソフトはインターフェースが今ひとつという感じがします。R-Commander も良くできてはいるんですがインターフェースが練り上げられていないという感じがします。

そこで我々は、学生に自由に使ってもらうために、分かりやすい初心者向けの統計ソフトを開発することにしました。せっかくですからその当時開発中だった OR 関係の分析ソフトに統合させ、できたものが「College Analysis」です。「分析」という大げさな名前ですので、今後より多くの分析手法を加えて充実させていかなければなりません。これはインターネット上で公開していますので、いつでも最新のものを自由に利用することができます。

この教科書では、我々の開発した分析ソフト College Analysis を使って統計処理をする方法を学びます。特別なソフトではありますが、ここで学んだことを基礎にして、他のソフトを利用すると理解が速いと確信しています。

福山平成大学 福井正康

1 章 集計

1.1 質的データの集計

データには性別や賛成・反対・どちらともいえないなどの分類データと身長、体重、テストの点などの数字が大切なデータとがあります。一般に前者は質的データ、後者は量的データと呼ばれ、統計では処理の方法を決める重要な性質です。但し、成績の5段階評価などのように、分類だけど点数として平均を取ったり、点数を優・良・可・不可に分けたりと同じデータを異なった見方で見る場合もありますので2つのデータが完全に分離されるわけではありません。ここではまず質的データについて、その代表的な集計方法を学びます。

最初に以下のような例について考えてみます。これはあるアンケート調査で性別（1：男性，2：女性）と、ある意見に対して賛成かどうか（1：賛成，2：反対，3：どちらともいえない）を聞いたものと考えていただければよいでしょう。これらはどちらも分類データですので質的データです。

例

20人にアンケートを取ったところ、以下のような結果が得られた。以下の問いに答え、結果は文書にまとめよ。

性別	回答	性別	回答	性別	回答	性別	回答
1	1	2	3	1	2	2	1
2	1	1	1	1	2	1	2
1	2	2	2	2	1	1	1
1	2	2	3	2	1	2	3
2	1	1	1	1	3	1	1

注) 性別： 1：男性 2：女性

回答： 1：賛成 2：反対 3：どちらともいえない

- 1) 回答に関する1次元分割表を描け。
- 2) 性別と回答に関する2次元分割表を描け。
- 3) 回答に関する分割表を用いて棒グラフと円グラフを描け。
- 4) 性別と回答に関する分割表を用いて積み重ね棒グラフを描け。

解答

College Analysis を起動させると、以下のような画面になります。

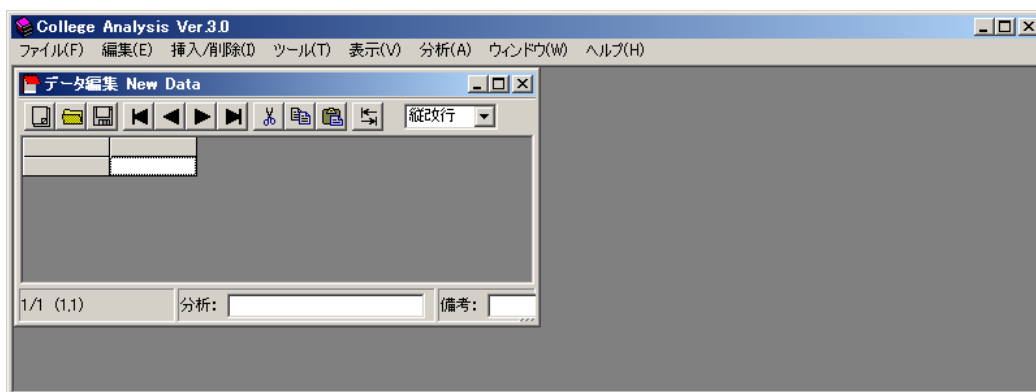


図 1.1.1 College Analysis 起動画面

起動画面上に現れているのはデータエディターです。ここではデータ入力から始めます。通常は Excel などからコピーしてきますので、実際大きなデータを打ち込むことはあまりありませんが、一応やっておきましょう。

まずエディターメニュー [ファイルー新規作成] で新規設定用の Window を開きます。



図 1.1.2 新規設定メニュー

この例題の場合、データは4列に表されていますが、入力は縦に長く入力します。これは Excel などと同じで、横方向のデータの並びを行またはレコードと呼び、縦方向のデータの並びを列または変数と呼びます。一人分のデータは横1行に入力して行きます。アンケート調査の場合、行数は調査人数、列数は質問項目数になると考えて下さい。

新規設定 Window のシート数はそのままにして行数を20、列数を2にして「OK」ボタンをクリックします。これでエディターに20行2列のデータ領域ができ上がりました。

た。

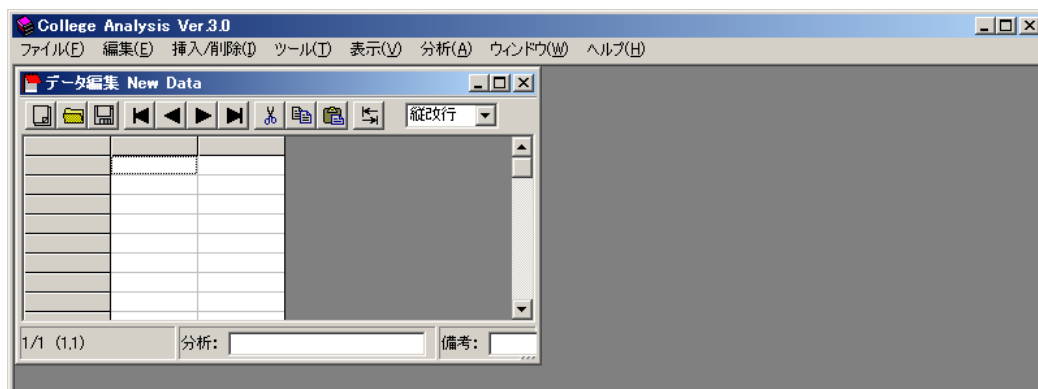


図 1.1.3 データ入力画面

シート数を増やすと行数と列数をシートごとに決めることができます。しかし、実際の処理ではページの追加機能を用いることが多いので、初期設定でページを指定することはまずありません。全体のシート数と現在のシート番号は画面下のステータスバーに“現在のシート/全シート数”として表示されています。その隣の“(1,1)”はカーソルのある行と列の番号です。

この中に問題で与えられたデータを入力します。入力が終わったら、続いてエディターメニュー [編集→列名入力] で現れた列名入力の **Window** に「性別」「回答」と入力します。このとき日本語入力が画面の上方に出ますので注意して下さい。



図 1.1.4 列名入力メニュー

この列名入力は必ずしも必要でなく、省略した場合、変数名は **var1** と **var2** となります。また同様にして行名（レコード名）も入力することができます。1 から連続した数値

にする場合は「連番」コマンドボタンをクリックし、ある文字列を先頭文字列として連番を付けることもできます。「OK」ボタンをクリックすると図のように列名（変数名）が確定されます。

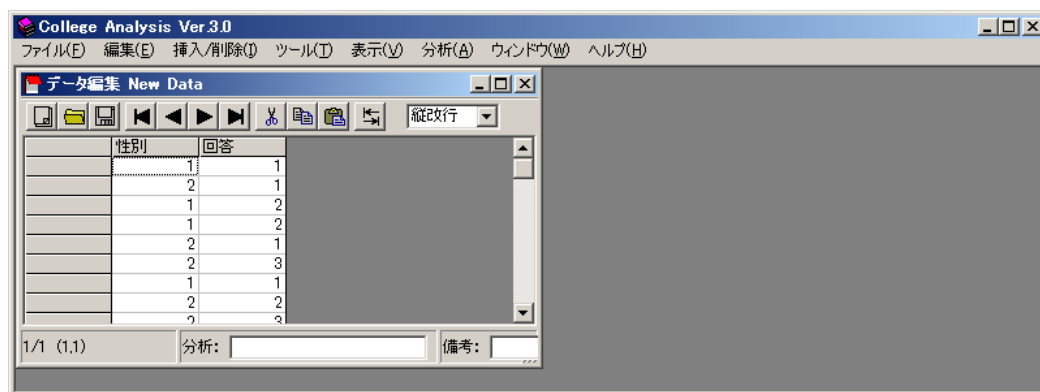


図 1.1.5 データ入力完了画面

それではいよいよ例題の解説をします。

1) 回答に関する 1 次元分割表を描け。

これは回答の 3 つの選択肢をそれぞれ何人が選んだかを数える問題です。質的なデータにおいては最も基本的な集計です。

まずエディターメニュー「分析－基本統計－質的データの集計」を選択します。そうすると質的データ集計 Window が開きます。

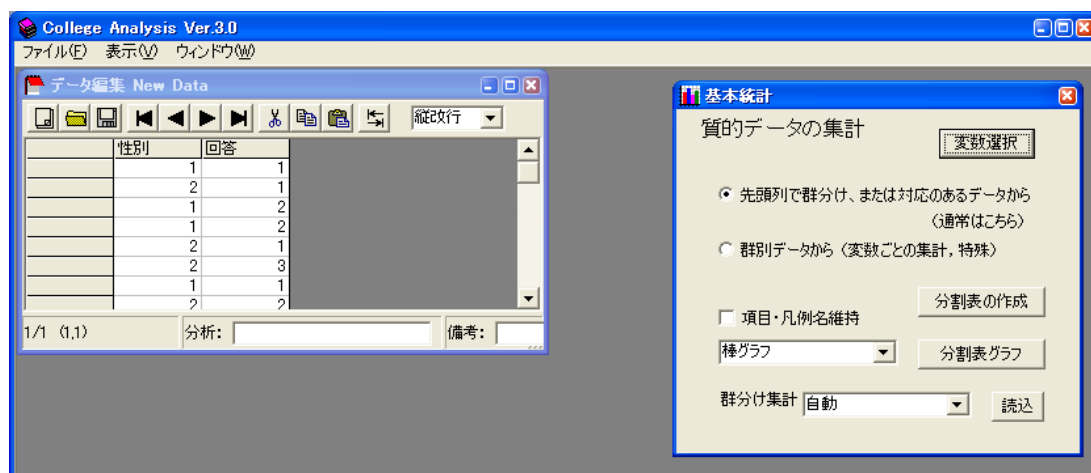


図 1.1.6 質的データの集計メニュー画面

このときメニューバーを見て下さい。たくさん項目のあった前のときと違っていません。メニューは各 Window について決まっていますので、どの Window が選ばれているかで表示されるメニューが違います。例えば最初に現れる「分析」を含むメニューはデータ編集 Window に付属しますので、これが表示されていない場合、これを表示するにはデータ編集 Window をクリックします。

ここではまず「変数選択」ボタンによりどの変数について集計を取るか決めます。「変数選択」ボタンをクリックすると以下の変数選択画面が現れます。



図 1.1.7a 変数選択画面

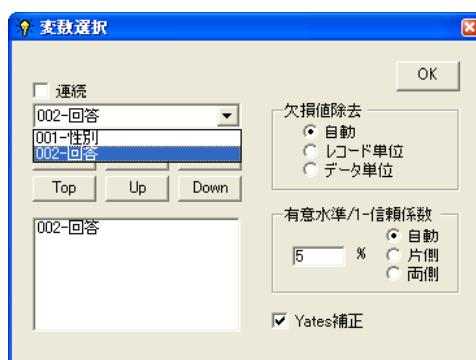


図 1.1.7b 変数選択

左図の左上のコンボボックスをクリックすると変数名が現れますので、その中から必要なものを選びます。選んだ変数名は下のリストボックスに表示されます。リストボックスの整理は「All」「Delete」「Reset」「Top」「Up」「Down」などのコマンドボタンで行いますが、今回は回答だけなので回答を選択して、「OK」ボタンをクリックします。

これで変数選択画面が消え準備が整いましたので、質的データの集計メニューの中の「分割表の作成」ボタンをクリックします。すると以下のような 1 次元分割表の画面が表示されます。

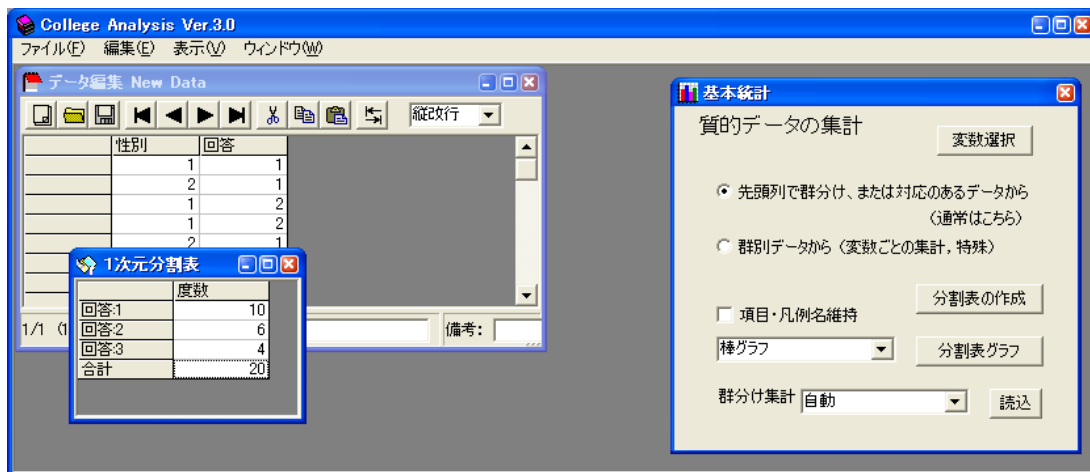


図 1.1.8 1次元分割表表示画面

通常 1 次元分割表は横並びで出力されますが、ここでは分割数が多い場合を考えて縦に表示しています。横に表示する場合は、[編集→行列交換] メニューで以下になります。

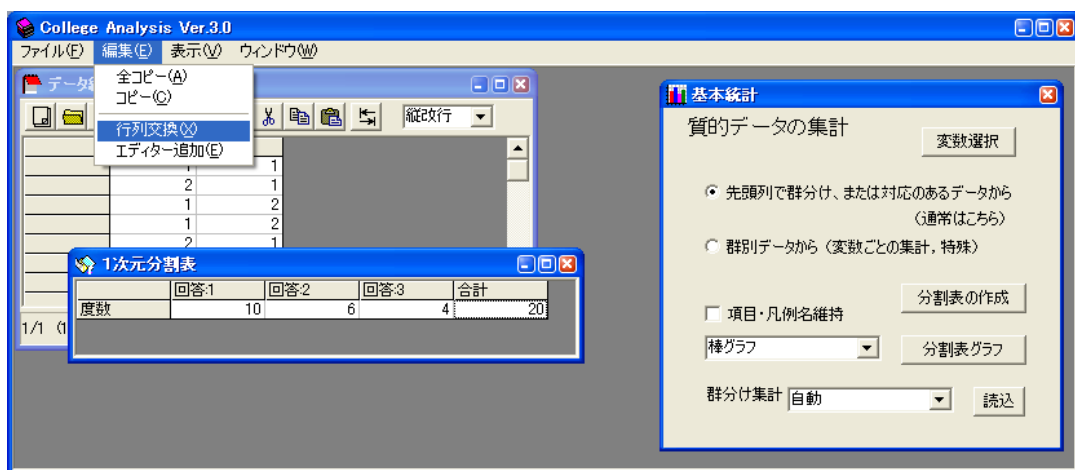


図 1.1.9 1次元分割表横表示画面

これで回答のところで 1, 2, 3 と答えた人数が分かりました。

2) 性別と回答に関する 2 次元分割表を描け。

次は男女別に回答状況を見てみたい場合です。このような集計には 2 次元分割表が利用されます。

先ほどの変数選択で性別と回答をこの順番で選んでみます。

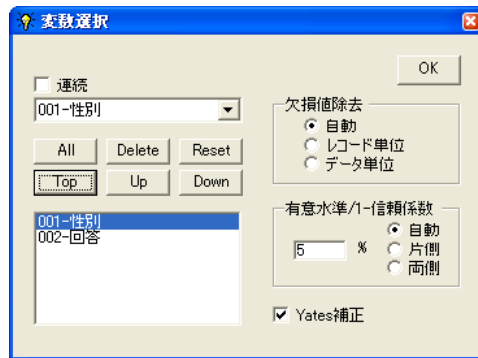


図 1.1.10 2次元分割表のための変数選択

その後、「分割表の作成」ボタンをクリックすると、以下のような結果が表示されます。

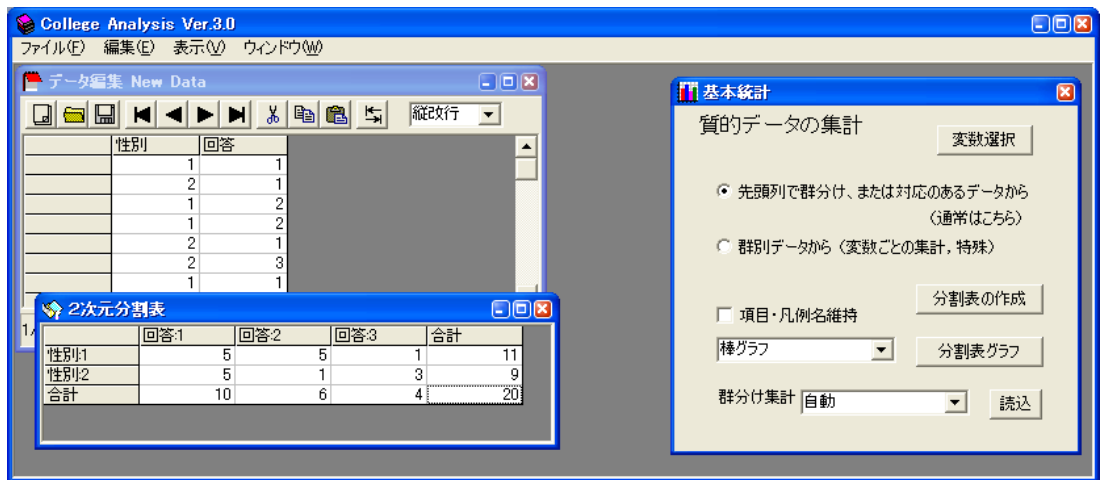


図 1.1.11 2次元分割表表示画面

ここで行が性別になっていますが、変数選択の順番を逆にすると行が回答になります。これを見ると男性の方が賛成が多いようです。

我々はこの結果を Word や Excel に貼付けて利用することが多いので、その方法をここで見ておきましょう。

この出力結果が選ばれた状態で [編集-全コピー] メニューを選択すると結果がコピーされます。Excel を起動して [編集-貼り付け] メニューで以下のようにデータが貼り付きます。

	A	B	C	D	E
1		回答:1	回答:2	回答:3	合計
2	性別:1	5	5	1	11
3	性別:2	5	1	3	9
4	合計	10	6	4	20

図 1.1.12 Excel へ表の貼り付け

同様に Word への貼り付けも行えます。但し、Word の場合はデータがタブ区切りで貼り付けられますので、以下のように選択して、

	A	B	C	D	E
1		回答:1	回答:2	回答:3	合計
2	性別:1	5	5	1	11
3	性別:2	5	1	3	9
4	合計	10	6	4	20

図 1.1.13 Word への表の貼り付け

「罫線ー挿入ー表」メニューで以下のように変換します。

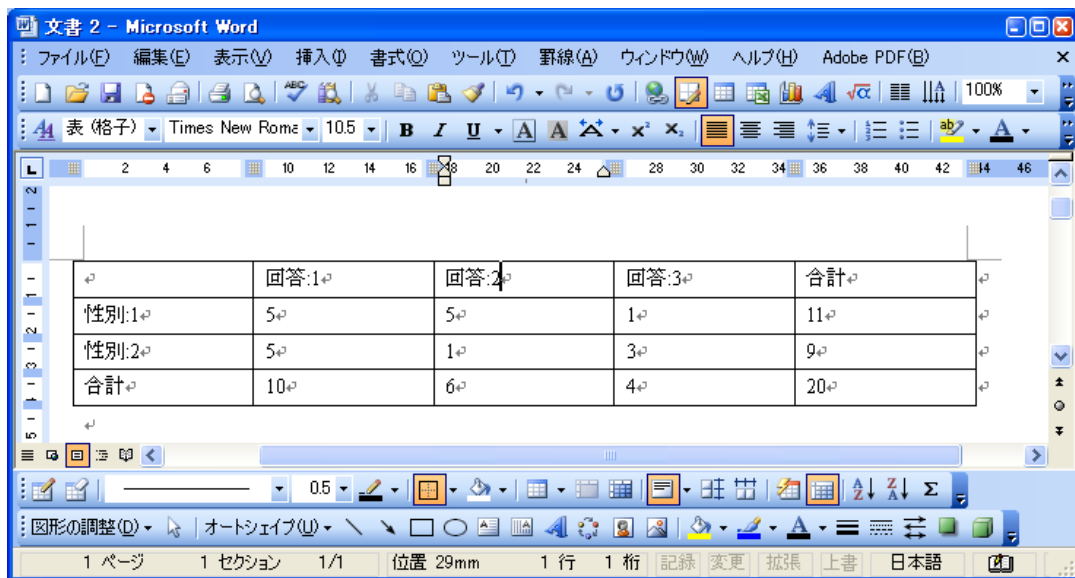


図 1.1.14 Word の表への変換画面

後の罫線の飾りは Word の機能を活用して下さい。

表の中の一部をコピーする場合は、全コピーではなく [編集→コピー] メニューを利用して下さい。

3) 回答に関する分割表を用いて棒グラフと円グラフを描け。

1 次元分割表を視覚的に表示するために棒グラフか円グラフがよく使われます。データの度数を重視する場合は棒グラフ、割合を重視する場合は円グラフで表示するのが標準的です。

まず質的データの集計メニューの左下のところにあるグラフ名のコンボボックスから棒グラフを選択します。これはデフォルトで設定されています。次にその右の分割表グラフをクリックすると以下のように棒グラフが表示されます。

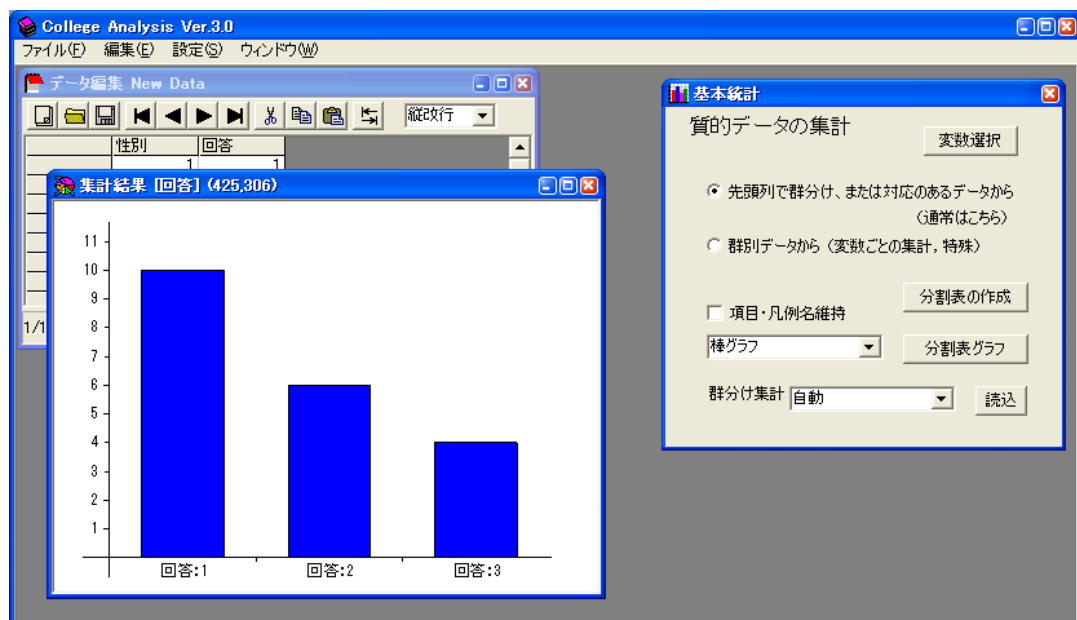


図 1.1.15 棒グラフ表示画面

ここで横軸の項目名を変更したい場合は、集計結果を選んだ状態でメニューから「編集－項目名変更」機能を選択します。その際に以下のような変更画面が現れます。

The '項目名入力' dialog box has a table with the following content:

	項目名
1	回答:1
2	回答:2
3	回答:3

Buttons on the right include 'OK', '全消去', '先頭文字列' (with an input field), '連番', 'コピー', and '貼り付け'.

図 1.1.16a 項目名入力画面

The '項目名入力' dialog box after editing, with the following table content:

	項目名
1	賛成
2	反対
3	どちらとも

The buttons and layout are identical to the previous screenshot.

図 1.1.16b 項目名入力後画面

項目名を右図のように入力した後「OK」ボタンをクリックすると、入力画面は消え、以下のように変更されたグラフが表示されます。

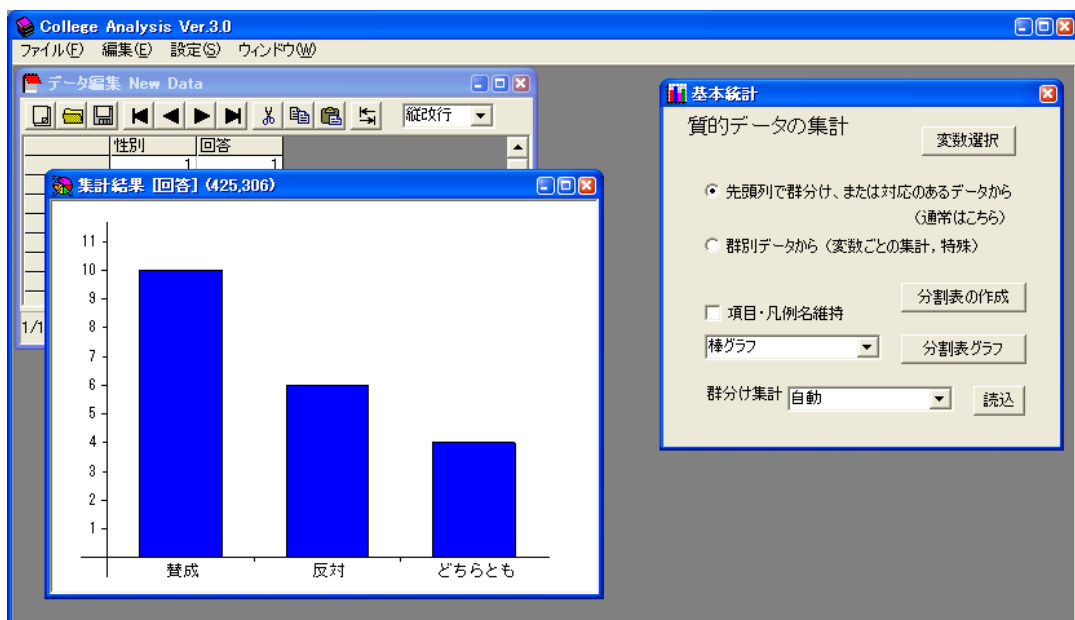


図 1.1.17 項目名変更後の棒グラフ

このソフトはグラフ機能が弱いので、これ以上きれいなグラフは望めません。そこできれいに出力させるためにグラフを描いたデータを Excel に渡せるように考えています。グラフを選択した状態で、メニュー [編集-データ表示] 機能を選択すると以下のようにグラフを描いた際のデータが表示されます。これをコピーして Excel に貼り付けてグラフを描くこともできます。

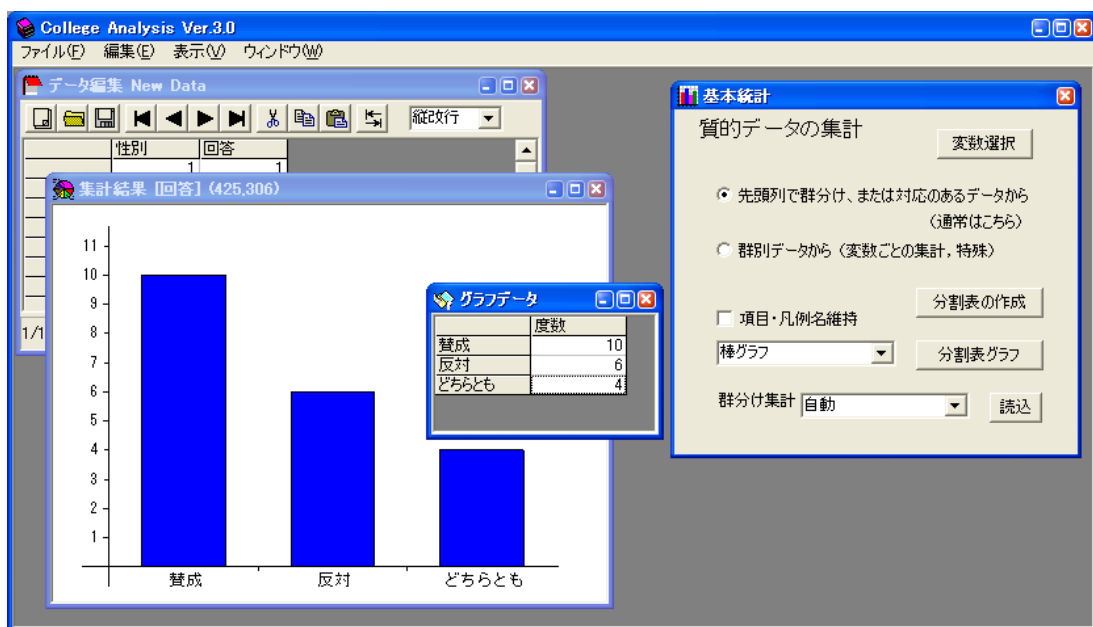


図 1.1.18 グラフデータ表示画面

つぎに円グラフについてですが、せっかく項目名を書き換えたのでそのまま残しておくことにします。質的データの集計メニューの「項目・凡例名維持」チェックボックスにチェックをいれて、グラフ選択のコンボボックスから円グラフを選び、「分割表グラフ」ボタンをクリックすると以下の円グラフが表示されます。

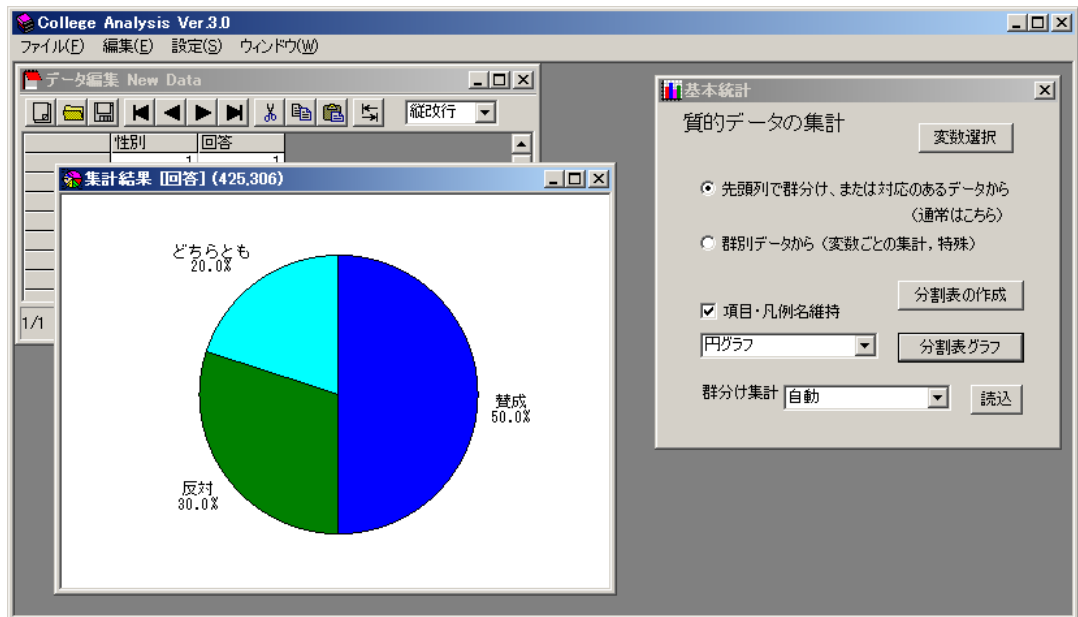


図 1.1.19 円グラフ表示画面

その他にもグラフのメニューはいくつかありますが必要になったときに説明します。

4) 性別と回答に関する分割表を用いて積み重ね棒グラフを描け。

この問題ではまず「変数選択」ボタンで、性別と回答をこの順番に選択します。メニュー右下のグラフ名のコンボボックスから積み重ね棒グラフを選択し、「分割表グラフ」ボタンをクリックすると以下のようなグラフが表示されます。

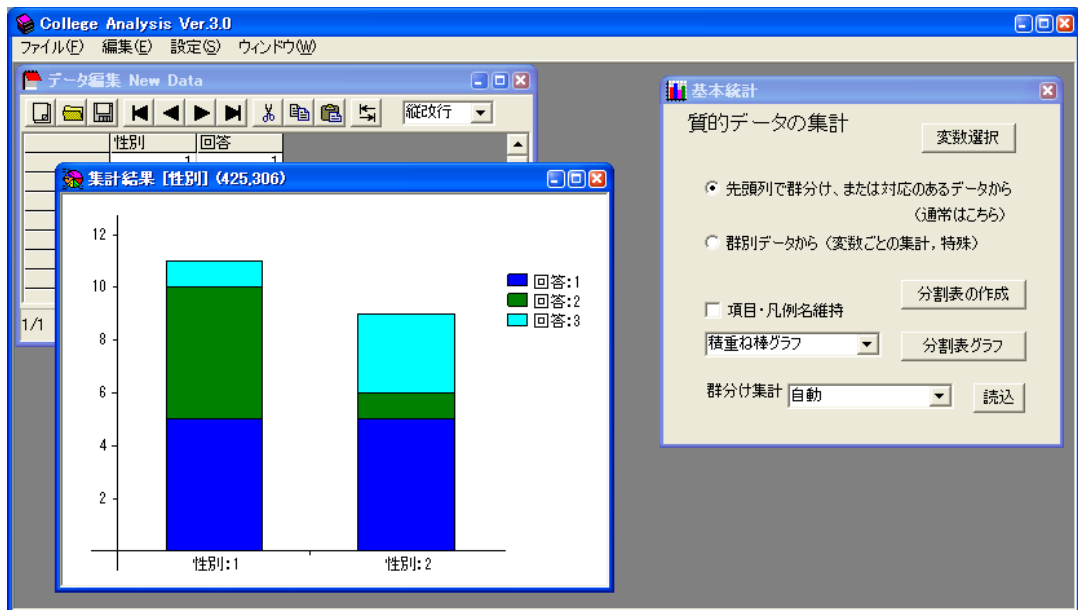


図 1.1.20 積み重ね棒グラフ表示画面

横軸の下項目名は、グラフのメニュー〔編集―項目名変更〕機能を使って訂正できますが、右上の凡例はグラフメニューの〔編集―データ・凡例名変更〕機能で訂正できます。

上の例は男女の数を基準にしていますが、変数選択の順番を変えると、回答を基準にグラフを描くこともできます

。

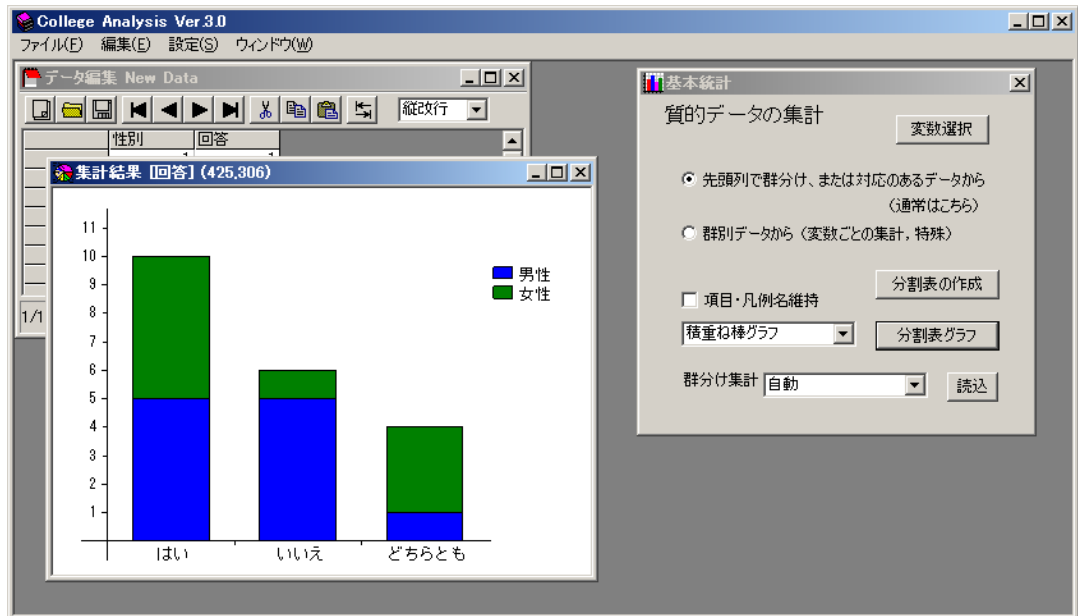


図 1.1.21 積み重ね棒グラフ表示画面 2

ファイルについて

次の問題ではファイルを読み込んで回答しますのでファイルの読み書きについて簡単に説明します。ファイルを読み込む場合はエディターメニュー「ファイルを開く」で標準的なファイル選択ウィンドウが開きます。ファイルの読み込み形式は、このプログラム独自の形式「名前.txt」と csv ファイル「名前.csv」です。どちらもカンマ区切りを用いていますが、独自の形式はシートの複数ページに対応しています。csv ファイルは 1 ページのみで、先頭列が行名の場合とデータから始まる場合の 2 種類から選択します。ファイルへの保存は独自の形式で行う方が良いでしょう。csv 形式にはメモ帳などで簡単に移せます。

問題 1

Samples¥テキスト 9.txt を用いて以下の問いに答え、結果は文書にまとめよ。但し、地域について 1：市街、2：郊外、意見 1 について 1：賛成、2：反対、意見 2 について 1：はい、2：いいえ、3：どちらとも（いえない）とする。

1) 地域に関する 1 次元分割表を描け。

市街	郊外	合計

2) 意見 1 に関する 1 次元分割表を描け。

賛成	反対	合計

3) 意見 2 に関する 1 次元分割表を描け。

はい	いいえ	どちらとも	合計

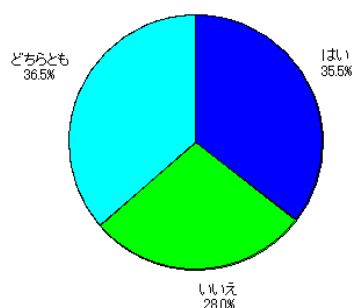
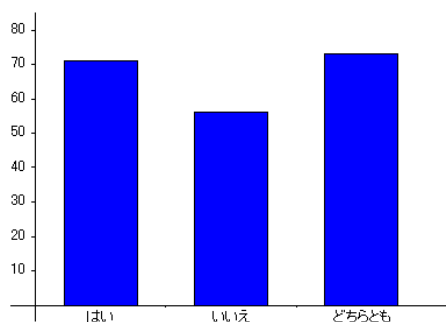
4) 地域と意見 1 に関する 2 次元分割表を描け。

	賛成	反対	合計
市街			
郊外			
合計			

5) 地域と意見 2 に関する 2 次元分割表を描け。

	はい	いいえ	どちらとも	合計
市街				
郊外				
合計				

6) 意見 2 に関する棒グラフと円グラフを描け。



7) 地域と意見 2 に関する積み重ね棒グラフを描け。

