はじめての
S-PLUS for Windows

S-PLUS 8.0 入門

株式会社数理システム

http://www.msi.co.jp/

目次

1 はじめに....4

1.1 S-PLUS の歴史....4
1.2 この本について....5

2 データを準備する....6

2.1 データを入力する....6
2.2 データのインポート....9
2.3 データの名前....13
2.4 データを保存する....14
2.5 データを管理する....14
2.6 S-PLUS のデータベース....17
2.7 データの操作....18
2.8 exenvirn データウィンドウから新しいデータを作る操作....25

3 グラフ....36

3.1 S-PLUS グラフ....36
3.2 グラフの簡単な描画....36
3.3 グラフの保存....39
3.4 プロットパレット....40
3.5 グラフの編集....42
3.6 プロットの追加....46
3.7 グラフの追加....49
3.8 さらにプロットの編集....49
3.9 グラフ編集のためのツールバーボタン....54
3.10 グラフツールパレット....54
3.11 いろいろな 2 次元グラフ....57
3.12 マルチパネルグラフ....62
3.13 マルチパネルの編集....64
3.14 軸タイプリストボックス....66
3.15 3 次元グラフ....68
3.16 いろいろな 3 次元グラフ....70
3.17 3 次元グラフと 2 次元グラフを重ね合わせる....73
3.18 3 次元グラフの編集....73

3.19 オブジェクト・エクスプローラとグラフ 75
3.20 グラフへデータを埋め込む 77
3.21 メニューからグラフを選択する 77
3.22 ブラッシングとスピン 78

4 統計 79

4.1 統計メニューでの操作 80
4.2 基本統計量の計算 80
4.3 検定 81
4.4 実験計画 86
4.5 回帰 87
4.6 その他の回帰分析 93
4.7 その他の統計モデル 95
4.8 オブジェクトエクスプローラと統計解析 98

5 スクリプトの利用 101

5.1 スクリプトウィンドウ 101

6 S-PLUS のカスタマイズ 104

6.1 オブジェクトエクスプローラによるデータ管理 104
6.2 異なるデータベースを利用する 105
6.3 オプションの指定 108

1 はじめに

S-PLUS は統合的統計解析ソフトウェアです。その機能を大きく 3 つに分けると、

- データ操作
- グラフ作成
- 統計解析

となります。この 3 つの機能はお互いに密接に係わり合い、実際の解析シーンでは、「データ操作をして、グラフを作成したら、データにおかしなところがあったので、もう一度データ操作に戻り、再びグラフ化し、最後に統計解析機能を用いた」というように、無意識のうちに行ったり来たりしているはずです。この本では、この 3 つの機能を独立した章としていますので、必要に応じて前に戻ったり、読み飛ばして先に進んでください。

1.1 S-PLUS の歴史

S-PLUS の前身は、米国ベル研究所で開発された、データ解析言語 S です。S は 1970 年代に開発された、当時としてはまったく革新的な言語です。その特長としては、

- オブジェクト指向言語である
- 対話型言語である

ことがあげられます。多くのデータ解析用の関数(プログラム）が既に用意されていることも、コンピュータではなく、データ解析を専門とした研究者には大変便利なものでした。そうは言っても、インストール、プログラミングなど、その敷居の高さから、当時は一部研究者だけが使うことのできる言語であることも事実でした。

その後、WindowsOS が搭載された PC の普及により、簡単なプログラムにより、手元のデータをグラフ化して眺めたり、計算によりその性質を確かめるツールへの要求が高まり、1980 年代には、米国の StatSci という会社が S 言語を商用化したソフトウェアを作り、販売する権利を獲得し、S-PLUS の誕生となります。その後、開発会社は変わるものの、S-PLUS の開発は続けられ、現在では Insightful 社が、S-PLUS バージョン 8 をリリースするまでに至っています。幾たびかのバージョンアップの間に、多くの関数(プログラム)、機能が追加され、Windows 版ではよりユーザーにとって使いやすいものになるよう、グラフィカル・ユーザー・インターフェース（以降 GUI とします）が追加されました。

日本では、1980 年代の販売開始当時から、株式会社数理システムが日本語化を行い、日本国内での販売を行っています。

1.2 この本について

この本は、S-PLUS for Windows バージョン 8 に準拠していますが、それ以前のバージョンであっても、ほぼ同じ操作を行うことができます。また、特に必要がない限り、GUI のみを用いて操作を行っています。S 言語を使っていろいろなプログラムを組んでみたい、グラフを描きたい、解析を行ってみたいという方は、S 言語に関する書籍をご覧頂くことをお勧めします。書籍は

http://www.msi.co.jp/splus/tips/books2002.html

に一覧があります。フリーウェア R の本を参考にすることもできます（ただし、プログラムは完全には一致しません)。

2　データを準備する

S-PLUS 上でデータを作成します。作成には、数字などを入力する方法と、既存のファイルからデータをインポートする方法とがあります。

2.1 データを入力する

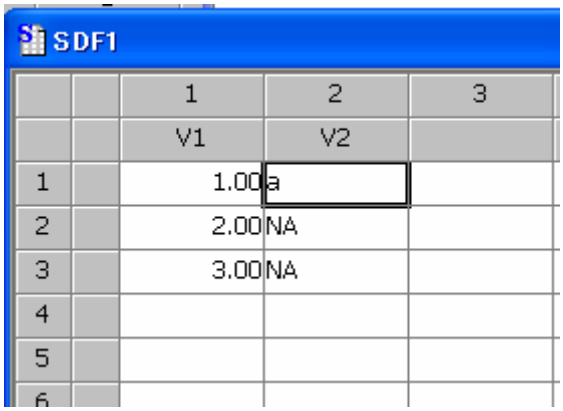
Excel など多くのスプレッドシート型ソフトウェア同様、新しいウィンドウを開いて、データを入力することができます。

新しいデータを作り、データを入力する

操作

- ツールバー **ボタン** をクリック
- 1 列目の 1 つ目のセルから順に、下に向かって、1,2,3,…と入力
- 2 列目の 1 つ目のセルに、a と入力

図 2-1 データ入力画面



分かったこと

- データのウィンドウに勝手に名前がついた（SDF1, SDF2,…）
- 列の最上部に V1,V2 と名前がついた
- 1 列目の数字が、小数点第 2 位で表示された

- 2 列目の 1 つ目に a と入れると、その下のセルが NA,NA となった

解説

新しいデータを作ると、そのデータには自動的に連続な名前（SDF1, SDF2,…)が付けられ、保存されます。

V1, V2,… を列名、あるいは変数名と呼びます。列名は必須なので、新しい列のセルに入力を行うと、自動的に連続な名前が付けられます。

数値を入力すると、セル上で表示される桁数は特に設定を変更していなければ、小数点第 2 位です。ただ、内部的には、倍精度の数値として保存されていますので、有効な桁数は 15 桁程度あります。

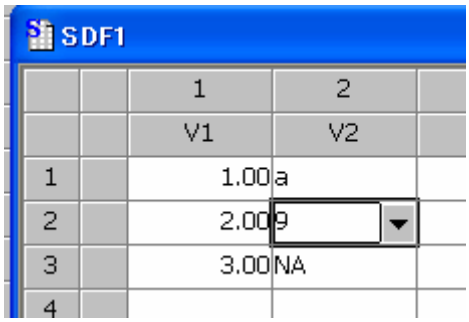
新しい列の 1 つ目のセルに入力を行うと、その下のセルは、データ全体の行数まで NA が埋められます。これは、データの各列は常に同じ長さであるためです。NA は Not Available、欠損を表し、文字ではなくシンボルで、NA と表示された場所にはデータがないことを意味します。

ここで新しく作られたデータは、データフレームという種類のデータです。

S-PLUS には、他にもいくつかの種類のデータがありますが、GUI から操作をする場合、基本となるのはこのデータフレームです。この本では、特に断りがない限り、データといった場合、データフレームを指します。

では、操作を続けます。a と入力したあと、次のセルに、9 と入力します。

図 2-2 文字の入力



分かったこと

9 の横には、下向きの三角が現れ、数字の小数点以下が表示されない

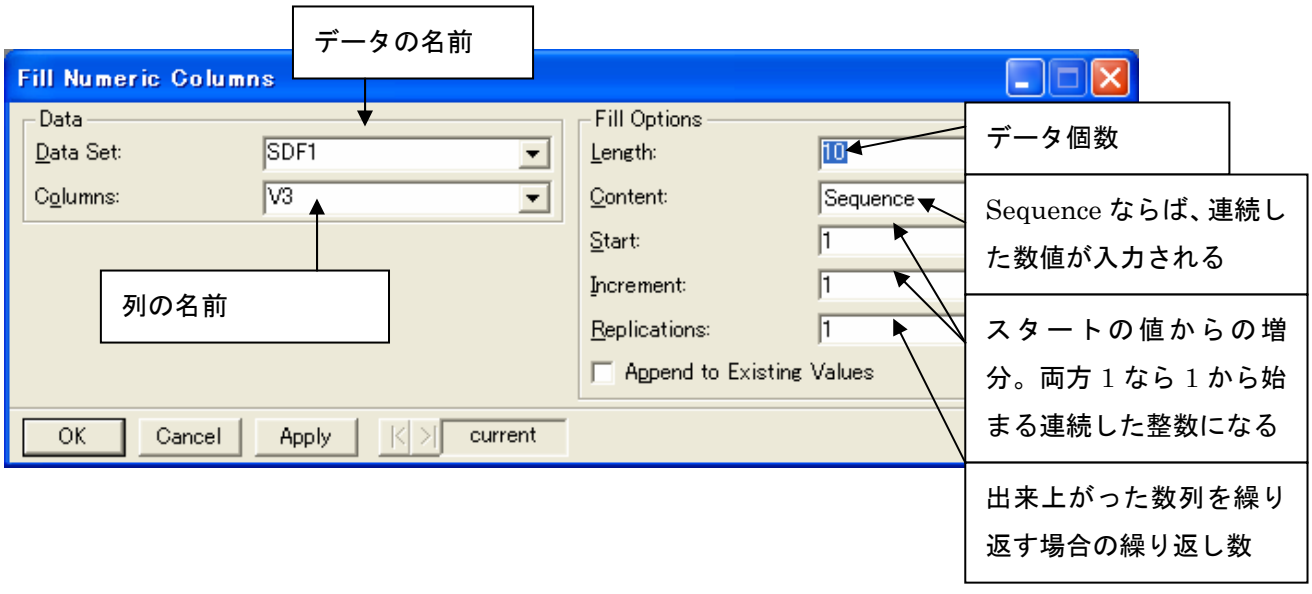
解説

S-PLUS のデータは、列ごとにひとつの種類のデータしか入力することができません。したがって、この 9 は数字の 9 ではなく、”a”と同じように、文字の 9 として扱われます。そのため、小数点以下は表示されません。もちろん、9.99 とすれば、小数点 2 位まで表示されます。下向き矢印をクリックすると、a と 9 からなるリストが現れます。このリストには、既に入力を行った値が表示されます。実際には、文字を入力した場合のデータの種類は、「文字」（character）ではありませんが、後で詳しく説明します。

連続値の入力

Excel ならば 1,2,3 と入力し、セルを選択、その先のセルにドラッグすると、連続した数値が入力される便利な機能がありますが、S-PLUS では、これに代わる機能として、メニュー **データ / 埋める** があります。メニューを選択すると、次のようなダイアログウィンドウが現れます。

図 2-3 メニュー 埋める のダイアログ



上記のように指定すると、1 から始まり、1 ずつ増える、全部で 10 個の数値が SDF1 の V3 列に入力されます。他の列にも、1 列に 10 個のデータが存在するように、NA が入力されます。

ここで、ダイアログのボタンについての説明です。

ダイアログ上の共通ボタン

ボタン	説明
OK	実行をしてダイアログを閉じる
Cancel	実行をせずにダイアログを閉じる
Apply	実行をしてダイアログを閉じない
|<>|	ダイアログのロールバック（ダイアログを元の状態に戻す）

2.2 データのインポート

Excel や他のソフトウェアで入力済みのデータがあれば、そのデータのインポートも可能です。特に、Excel の場合、インポート対象となるデータにはいくつかの注意が必要です。

注意点

- データは、m 行 n 列の形になっている必要があります（S-PLUS のデータが、m 行 n 列になっているためです）
- いくつかのセルが空白の場合は欠損（前述：NA）としてインポートされます
- 数式などのままインポートされるわけではありません（計算結果のみがインポートされます）
- S-PLUS V8 は Excel2007 形式ファイル(*.xlsx)には未対応です。Excel でセーブする際に、Excel97-2003 ファイル形式で保存してください。

次のようなデータは、インポートのダイアログでオプションを指定することにより、インポート可能です。

インポート可能なデータ例

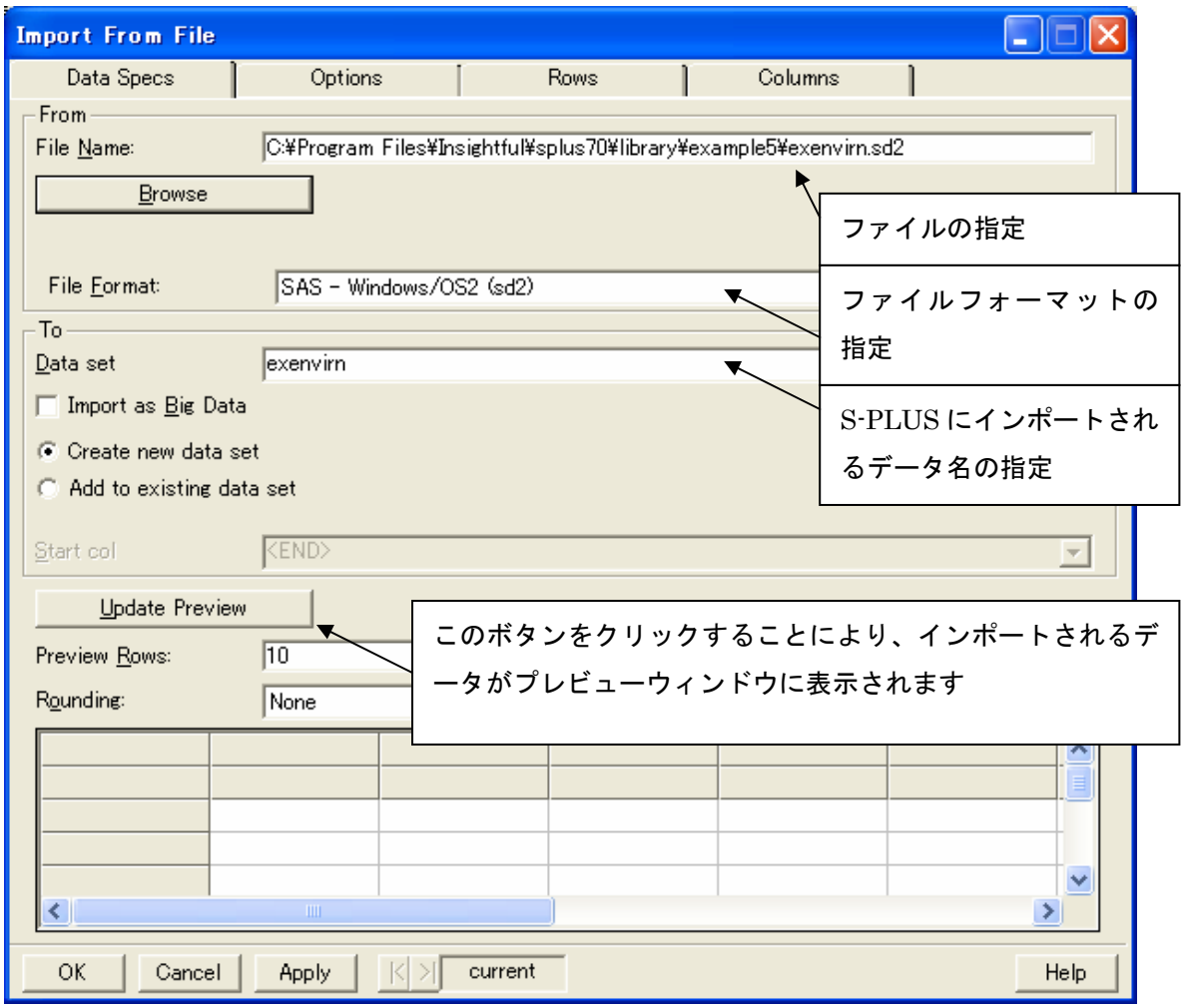
- Excel ブック形式で、複数シートにデータが保存されている
- データが 1 行目、あるいは 1 列目のセルから始まっていない
- データの途中からインポートしたい、あるいは途中までインポートしたい

ここでは、サンプルデータとして用意されている、SAS データ（exenvirn.sd2）のインポートを行います。

- メニュー **ファイル / インポート / ファイルから読み込み** を選択

- **Browse ボタン** をクリック
- **Select file to import** ダイアログでファイルの指定、ファイルの場所をクリックし、S-PLUS インストールフォルダの下、**library¥example5** フォルダにある **exenvirn.sd2** ファイルを選択

図 2-4 インポートのダイアログ



ファイルフォーマットは、自動的に、指定されたファイルの拡張子（ここでは*.sd2）から決定されます。インポートされるデータ名はファイル名の拡張子を取ったものになります（もちろん、変更も可能です）。

さらに細かい指定

SAS などのデータに関しては、この指定でデータをインポートすることができます。ただし、前述のように、Excel ブック形式のデータで、複数シートを指定する場合、あるいはデータの途中からインポートする場合、途中までをインポートする場合などは、オプションの指定が必要となります。オプション指定画面は複数のページに分かれていて、

- シート番号などを指定する場合　　**Options**
- 行に関連した指定　　**Rows**
- 列に関連した指定　　**Columns**

のページから指定します。

例えば、1 つのブック形式ファイルから、複数のシートをインポートするには

- **Optinos タブ** をクリック
- **Worksheet Number** に、インポートするシートが左から何番目に位置するか、半角数字で入力（シート名がリストされるわけではありません)、指定がないと、一番左にあるシートがインポートされます

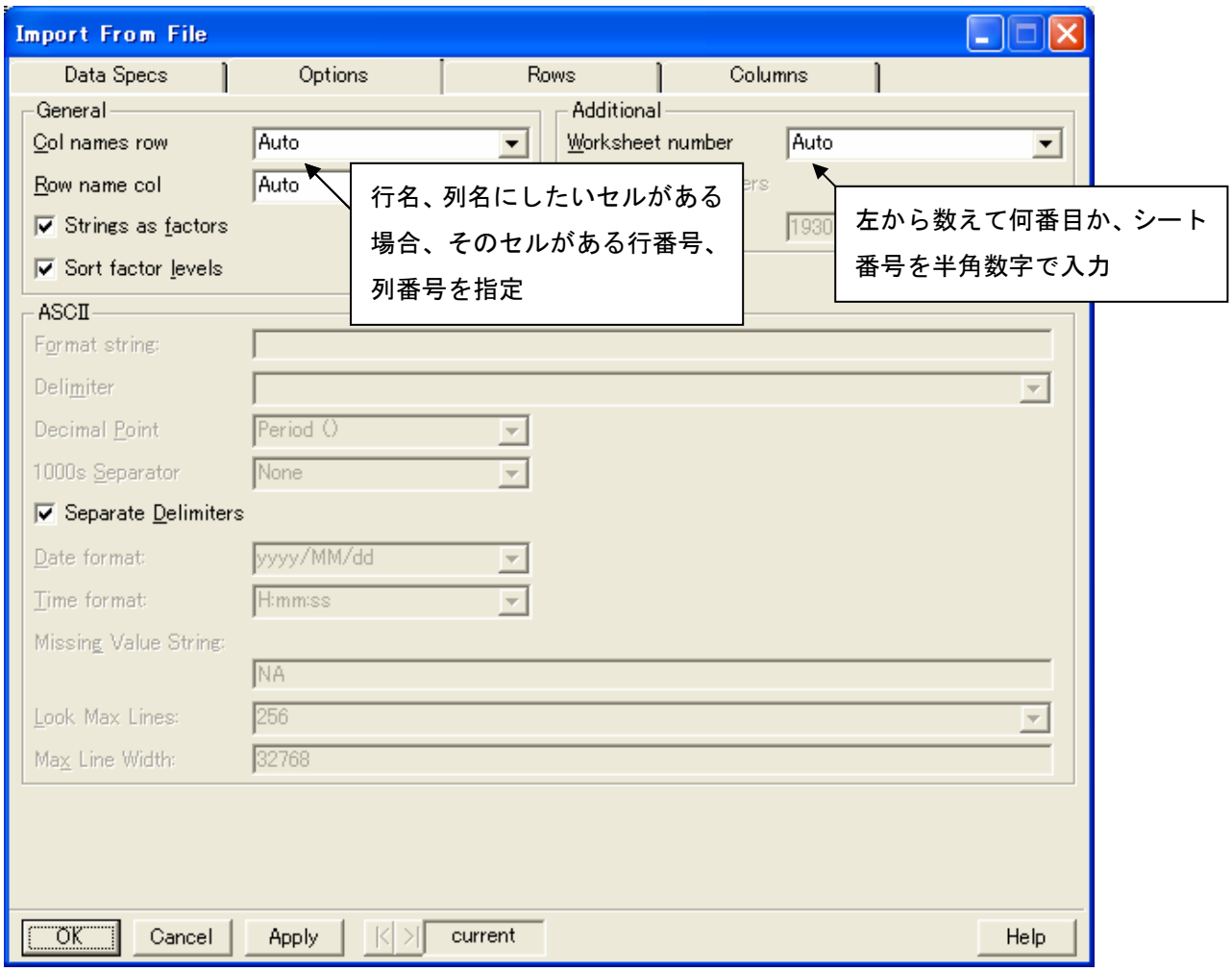
同じファイルからインポートをする場合は、データが上書きされないよう、あらかじめ、**Data Specs ページ** の **Data set フィールド**に新しいデータ名を入力してください。

インポート開始行や終了行を指定するには

- **Rows タブ** をクリック
- **Start row** にインポートを開始したい行の番号を指定
- **End row** にインポートを終了したい行の番号を指定

列も同様です。指定がないと、1 行目からデータがある最後の行までをすべてインポートします。

図 2-5 Import ダイアログの Options タブページ



OK あるいは、**Apply** ボタンをクリックすると、次のように `exenvirn` という名前でデータがインポートされます。

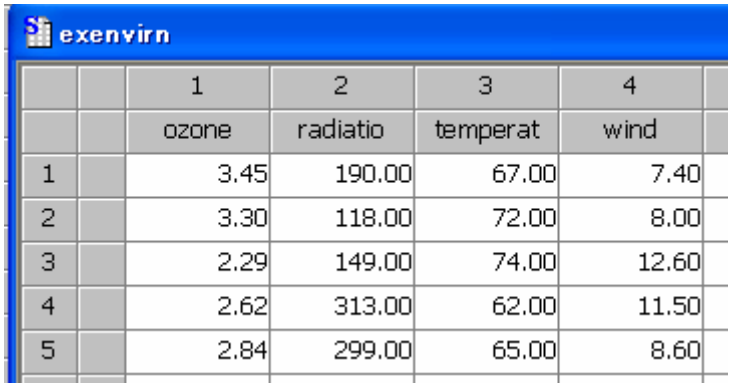
分かったこと

`exenvirn` という名前の付いたデータ・ウィンドウが開き、各列には名前がついている

解説

SAS データの場合は、SAS の項目名(変数名）が S-PLUS の各列名となります。特に設定をしないと、インポートされたデータはインポート時にデータ・ウィンドウ上に表示されます。これもまた、データフレームという種類のデータです。

図 2-6 インポートされたデータ（データフレーム）



exenvirn

		1	2	3	4
		ozone	radiatio	temperat	wind
1		3.45	190.00	67.00	7.40
2		3.30	118.00	72.00	8.00
3		2.29	149.00	74.00	12.60
4		2.62	313.00	62.00	11.50
5		2.84	299.00	65.00	8.60

Excel やテキストファイルなどで、1 行目から数値データが入力されており、列名が定義されていない場合は、S-PLUS が連続的な名前をつけます（Col1, Col2,…)。

2.3 データの名前

S-PLUS では、データの名前についていくつかのルールがあります。

- データの名前に使うことができるのは、英数文字、全角かな（日本語版のみ)、数字、ピリオド（アンダーバー _ は使うことができません）
- データの名前は、英数文字、全角かな（日本語版のみ）で始める（数字から始めることはできません。また、ピリオドも特殊なケースのために用いられるので、通常のデータ名の始まりには用いないようにしてください）
- 大文字と小文字は区別される

データ名の長さには特に規定がありません。特に、Excel やテキストファイルの場合、アンダーバーはファイル名に比較的よく用いられますが、S-PLUS では、ピリオドに変換されたデータ名となります。これを、改めてアンダーバーに直すことのないよう、注意してください。`exenvirn` の列名は、SAS データの変数名ルール（8 文字まで）に従ったものです。

また、これらデータの名前に関するルールは、同様にデータの各列の名前にも適用されます。

2.4 データを保存する

Excel などのアプリケーションでは、データはファイルに名前をつけて保存しないと失われますが、S-PLUS では、キーボード入力して作成したデータでも、他のアプリケーションのファイルをインポートして作ったデータでも、すべてデータベースに自動的に保存されます。ここで S-PLUS をシャットダウンしても、再起動すれば、自動的にそのデータベースが利用可能となり、以前作ったデータを確認したり、そのデータを用いたグラフ作成、計算ができるようになります。したがって、データの保存という操作は不要です。

ただ、次のようなケースでは、ファイルに名前をつけて保存するのが便利です。

- あるデータのバックアップを取りたい
- 他の S-PLUS ユーザーにデータをメールで送りたい

操作は、他のアプリケーション同様、保存したいデータのウィンドウをアクティブなウィンドウにして、メニュー **ファイル / 名前をつけて保存** を選択、保存したいファイル名やフォルダを指定します。

バックアップ、あるいはメールで受けとったファイルを開くには、そのファイルをダブルクリック、あるいは、メニュー **ファイル / 開く** で開きたいファイルを指定します。

次章では、閉じてしまったデータウィンドウを再度開く方法を説明しますので、一旦、アプリケーションの閉じるボタン、あるいは、メニュー **ファイル / 終了** によって、S-PLUSをシャットダウンします。

2.5 データを管理する

ここでは、既存のデータを開く、削除するなどのデータの運用について紹介します。
データ管理については、最後に、さらに進んだ方法を説明します。

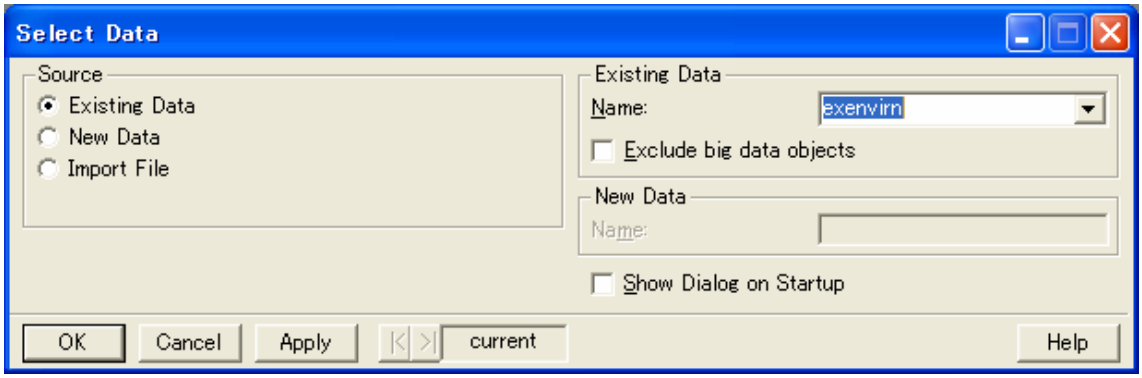
データウィンドウにデータを開く

再度、S-PLUSを起動します。前回インポートした `exenvirn` データはS-PLUSを起動した直後はウィンドウに開かれていませんが、保存はされています。
一旦、閉じたデータをウィンドウ上に開くには

- メニュー **データ / データの選択**

- **Existing Data** リストボックスから exenvirn を選択し、**OK** をクリック

図 2-7 データの選択をするダイアログ

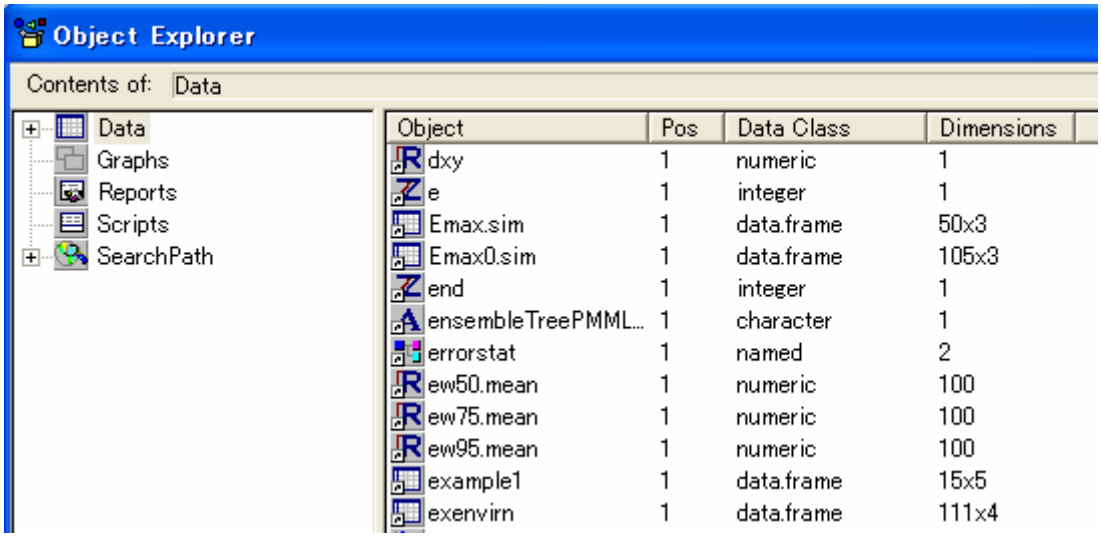


続けて他のデータをウィンドウに開くには、**OK** の代わりに、**Apply** をクリックします。連続して他のデータの選択が可能となります。

データベースにあるデータを検索する

前の章で、S-PLUS のデータはすべてデータベースに自動保存されると説明しました。では、そのデータベースには現在、どんなデータがあるのかを確認するには、ツールバーの**オブジェクト・エクスプローラボタン** をクリックします

図 2-8 オブジェクト・エクスプローラを開いたところ



Windows のエクスプローラと似ていますが、左側から、**Data** を選択すると、右側には現在、ユーザーのデータベースにあるデータが一覧表示されます。ここから、exenvirn を選び、ダブルクリックで開きます。

データが多くて、見つけるのが大変な場合は、ツールバーの **検索ボタン** をクリック、**Pattern** に検索したいデータ名を入力します。ワイルドカード*も用いることができます。exenv* とすれば、exenv で始まるデータを検索することが可能です。

データを削除する

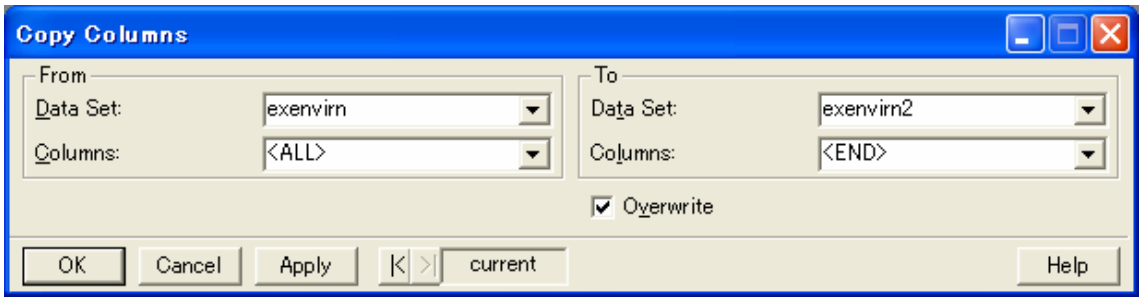
オブジェクト・エクスプローラから、不要なデータを選択し、**Delete** ボタンをクリック、あるいはメニュー **編集 / 切り取り** とします。

データをコピーする

- メニュー **データ / コピー / 列** を選択
- **From** の **Data Set** にデータ名を指定、**Columns** にコピーしたい列を CTRL キーを押しながら選択、あるいは、**<ALL>** の場合は、すべての列がコピーされます

- **To** の **Data Set** にコピーしたい新しいデータ名を指定

図 2-9 列コピーのダイアログ



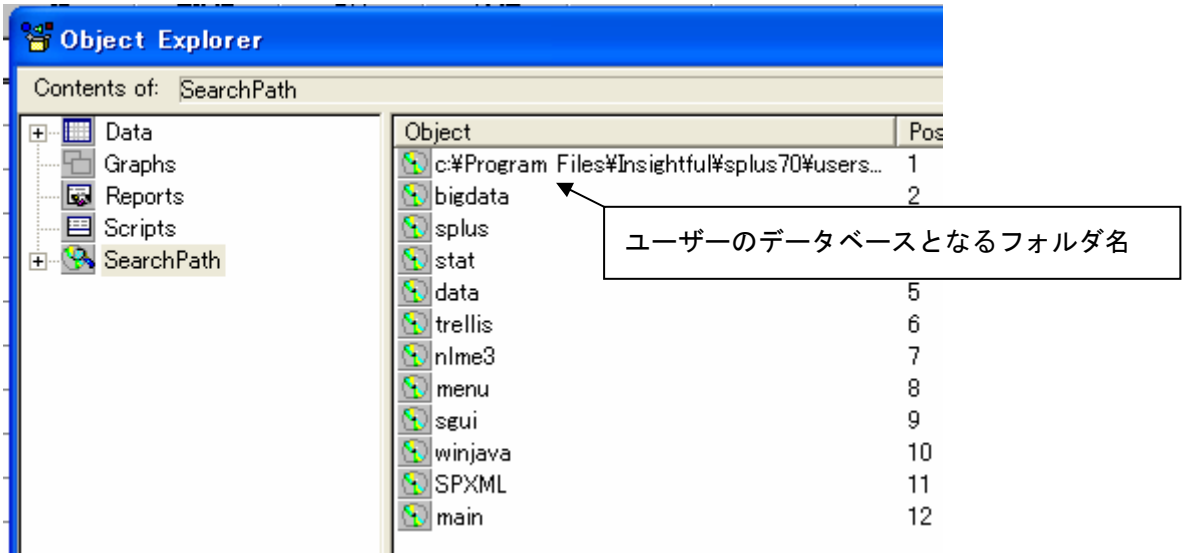
データの名前を変える

- オブジェクト・エクスプローラから、名前を変えたいデータを選択、右ボタン
- メニューから、**Properties** を選択
- **Name** フィールドに、新しく用いたい名前を入力

2.6 S-PLUS のデータベース

先ほどから、S-PLUS のデータはデータベースに保存されると説明しました。そのデータベースはどこに作られるのか調べるには、オブジェクト・エクスプローラの左側にある **SearchPath** をクリックします。右側の 1 番目には、フォルダの名前が表示されます。このフォルダがユーザーのデータベースが作られる場所です。

図 2-10 オブジェクト・エクスプローラからデータベースの確認



このデータベースを**作業データベース**と呼びます。

2.7 データの操作

exenvirn データウィンドウの操作

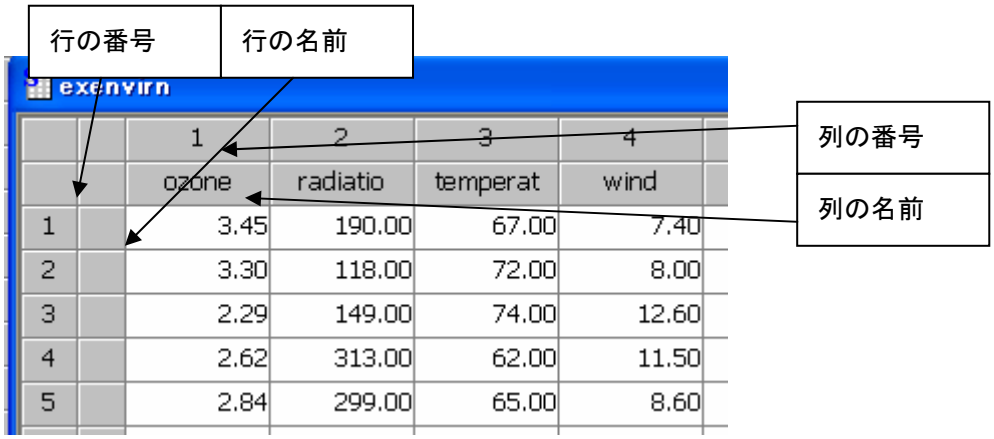
データウィンドウを使って、データ操作をします。

操作に失敗した場合、ツールバーの **元に戻す** **ボタン** で、1 回の操作を元に戻すことが可能です。

データウィンドウについて

入力、あるいはインポートされたデータ、またはオブジェクト・エクスプローラなどから開いたデータはウィンドウに表示されます。このウィンドウをデータウィンドウと呼びます。まず、データウィンドウについて、簡単に説明します。

図 2-11 データウィンドウ



データウィンドウは灰色のセルと白のセルから構成されています。灰色のセルは、列や行の名前と番号を示しています。特に、列の名前(変数名)は必須です。

ある列のデータ種類を確認する

データ入力の際に、データは列ごとに 1 つの種類（数値なら数値、文字なら文字）にしか入力できないというルールについて説明しました。ウィンドウ上から、データの種類を簡単に確認するには、列の番号を示す灰色のセルにマウスを移動させます。すると、その列がなんという種類のデータ表示させる **DataTips** が表示されます。この **DataTips** に表示される、データの種類は次の表のいずれかになります。

データの種類と、対応するインポートファイル

データの種類	説明	対応するインポートファイルの種類
`character`	文字列(ダブルクォーテーションでくくられます)	
`complex`	複素数	
`timeDate`	日付	Excel の日時データをインポートした場合
`double`	倍精度数値	すべてのファイルにおける数値データ
`factor`	因子	すべてのファイルにおける文字データ

integer	整数	
logical	論理値（T,F）	
single	単精度数値、およそ 7 桁の数値表現	

既存の列の対数を新しい列とする

既存の列から計算した結果を新しい列として挿入するには、

- メニュー **挿入 / 列**

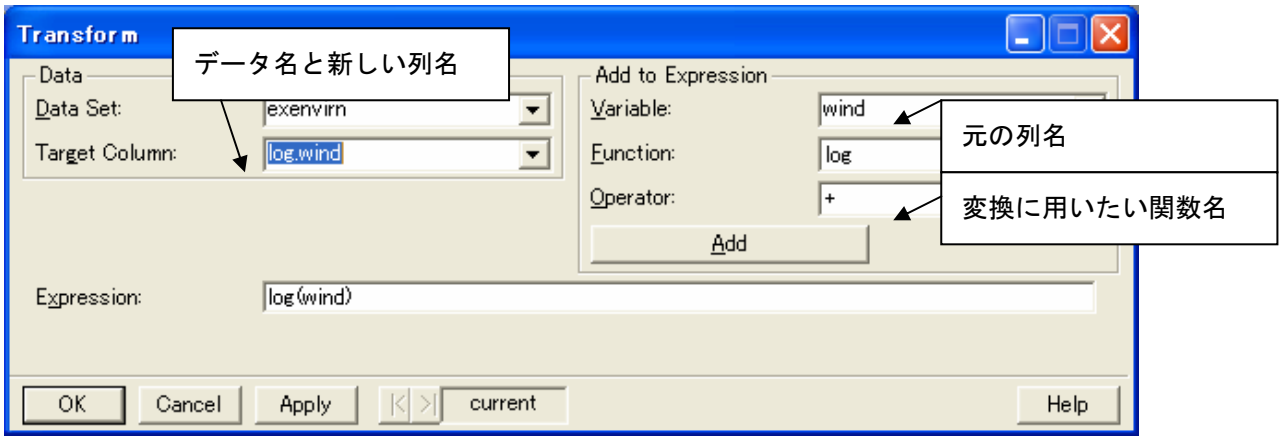
あるいは

- **データ / 新しい列を変換式で作成**

を用います。特に、S 言語に詳しくなければ、簡易的な式ビルダーのついている **新しい列を変換式で作成** が便利でしょう。

exenvirn の 列 wind の自然対数を計算して、log.wind とします。

図 2-12 メニュー データ /新しい列を変換式で作成 を選択したダイアログ



- **Data Set** のリストボックスから、exenvirn を選択
- **Target Column** に新しい列の名前を入力（既存の列を選択すると、上書きされてしまうので注意）
- **Variable** から、wind を選択
- **Function** から、log を選択
- **Add** ボタンをクリック

Expression フィールドに、log(wind) が表示されることを確認します（これが計算に用

いられる S 言語の式になります)。さらにこの値をある値で割るなど、四則演算を行う場合は、**Operator** フィールドから選択します。

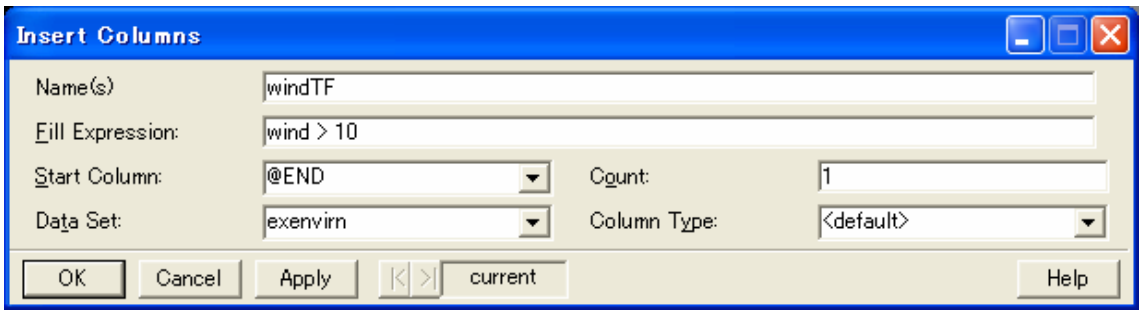
ここで注意が 1 つあります。

Excel であれば、元のデータを変えると、このデータを参照している他のセルの値も変わりますが、S-PLUS では、例えば上の `log.wind` 列はもう `wind` 列を参照していませんので、`wind` 列の値を変えても、`log.wind` の値は変わりません。列削除をして、再度同じ操作をする必要があります。

論理値列の挿入

では、さらに、挿入メニューを用いて、`wind` の値が 10 より大きいかどうかを示す列を挿入します。

図 2-13 メニュー 挿入 / 列 を選択したダイアログ



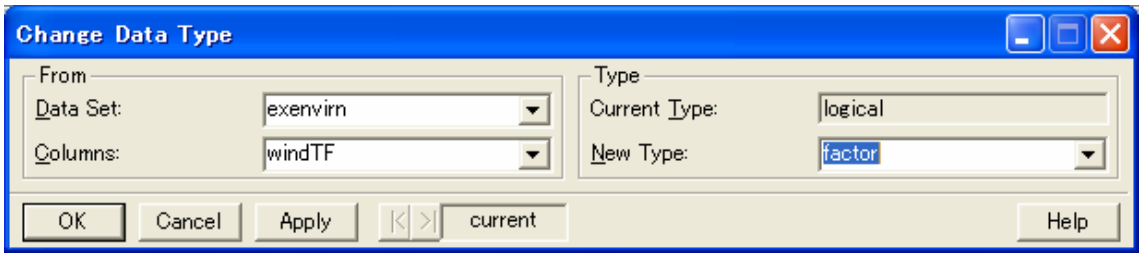
- **Name(s)** フィールドに、新しい列名 `windTF` と入力
- **Fill Expression** フィールドに、`wind>10` と入力
- **OK** をクリック

新しい列 `windTF` が `exenvirn` 内に作られます。比較演算子のみを指定すると、正しい場合に対応するセルに `T` が、間違っている場合には `F` が挿入されます。この列のデータの種類を、列番号のセルをマウスでポイントして見ると **logical（論理値）** となります。

データの列形式を変える

上記のようにして生成した列 windTF は論理値ですが、これは、T,F だけからなる最も原始的なデータです。これを、例えば「はい、いいえ」のように表示するには、論理値のままではできませんから、データの列の形式を変換します。

図 2-14 メニューバー データ / 列形式の変換　を選択



- **Data Set** に、exenvirn を選択
- **Columns** に、windTF を選択
- **Current Type** は **logical** となっているので、**New Type** で、**factor（因子）** を選択
- **OK** をクリック

多くの場合、文字データを表すのに、**factor（因子）**を用いることが望ましいです。**character（文字列）**は、人名など、ユニークになるときに用います。

気がついたこと

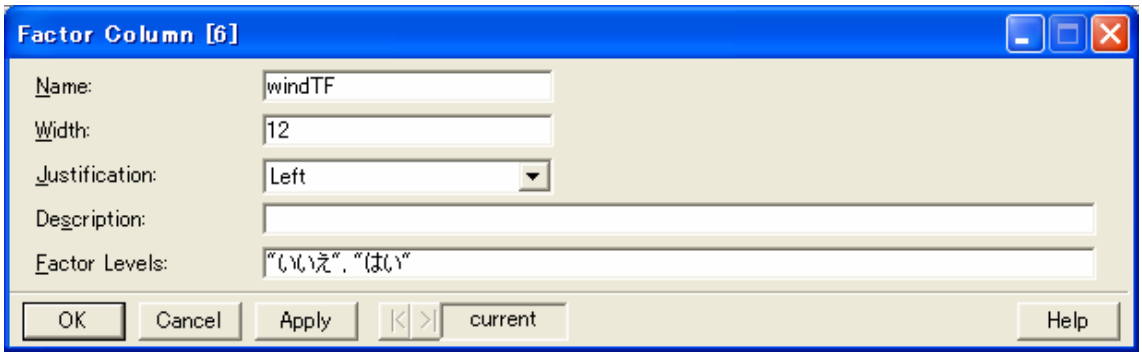
windTF の T が TRUE に、F が FALSE に変換されている。

文字の置換

windTF の TRUE を「はい」に、FALSE を「いいえ」に置き換えます。
TRUE, FALSE それぞれを、メニュー **データ / 置換** を用いて置き換えることも可能ですが、windTF の場合、TRUE, FALSE のデータを一度に置き換える、次の方法を用いると便利です。

- 列 windTF を選択（列全体が反転表示されている状態）し、マウスの右ボタンクリック、現れたメニューから **Properties** を選択
- "FALSE", "TRUE"を、"いいえ","はい" に変える
- OK をクリック

図 2-15 Properties を選択したダイアログ



ダブルクォーテーションを削除しないよう、注意してください。これで、FALSE が"いいえ"、TRUE が"はい"に置き換えられます。

データのソート

列 ozone の値の小さい順にデータを並べ替えます。

- 列 ozone を選択（列全体が反転表示されている状態）し、**ツールバー ボタン**をクリック

気がついたこと

データが並べ替えられますが、行の名前を表す灰色のセルに数値が入力されている。また、横並びになっているデータは一緒にソートされている。

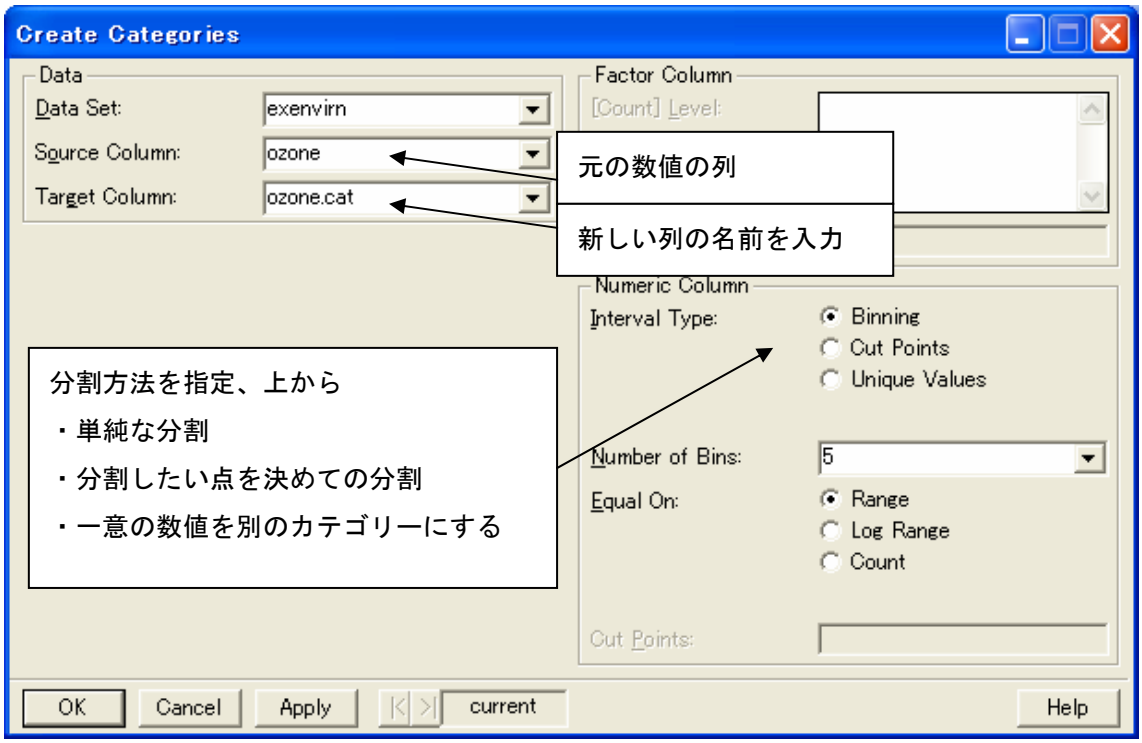
行の名前にある数値は、元のデータの何行目にあったかを示す数値です。また、ソートは指定された列をキーとして、データ全体がこのキー列の順に並べ替えられます。つまり、指定された 1 列だけが並べ替えられるわけではありません。並べ替えのキーにする列は複数選択することができます。

数値を因子に変換

年齢などの数値を「10 代」、「20 代」のようなカテゴリーにしたいことがあります。例えば、列 ozone を何段階かに分けて濃度が濃い場合、薄い場合でどのような違いがあるか

見たいとします。

図 2-16 メニュー データ / 数値を因子に変換 を選択



分割方法

Binning:

データを **Number of Bins** で指定した数のグループに分けます。デフォルトは5グループへ、**Equal On** で指定されている**Range**(等間隔）に分けます。**Log Range** なら、対数を取った等間隔に、**Count** なら、件数が大体等しくなるように分割する点を決めます。

Cut Points:

Binning と似ていますが、分割したい両端点をユーザーが決めることが可能です。**Cut Points** にチェックをすると、下の **Cut Points** フィールドに 5 つの数値が現れますが、これは、データの四分位範囲（最小値、25%点、中央値、75%点、最大値）を表します。つまり、デフォルト時には、データは等間隔に 4 分割されますが、**Cut Points** フィールドの数値を変更することにより、自分で分割したい両端点を決めることも可能です。例えば、このフィールドに、0,30,60,100 とあれば、「0 以上 30 以下」「30 より大きく 60 以下」「60 より大きく 100 以下」の 3 分割になります。「以上、未満」ではなく、一番小さい場所を除いて「より大きく、以下」なので注意してください。この値範囲に含まれない数値は **NA** とな

りますので、注意してください。

Unique Values:
1 つの数値を 1 つのカテゴリーにします。アンケート調査の回答番号、医薬品の臨床データにおける subject 番号などは **Unique Values** を選び、因子にしておくと便利です。反対に、年齢や測定値のような連続した数値には向きません。

このダイアログを使って、`ozone` 列をいろいろなカテゴリーに分けます。

分かったこと

新しい列は **因子（factor）**になり、1 つ 1 つの水準は 0.9548215+ thru 1.876463 のような名前になる。

解説

水準名には、区分点の値が用いられます。もちろん、列を選択、右ボタンクリックで **Properties** を選び、変更することも可能です。

この方法を用いて、`ozone.cat` の因子水準を順に、"低","やや低","中","やや高","高"とします。ダブルクォーテーションやカンマの数に注意してください。

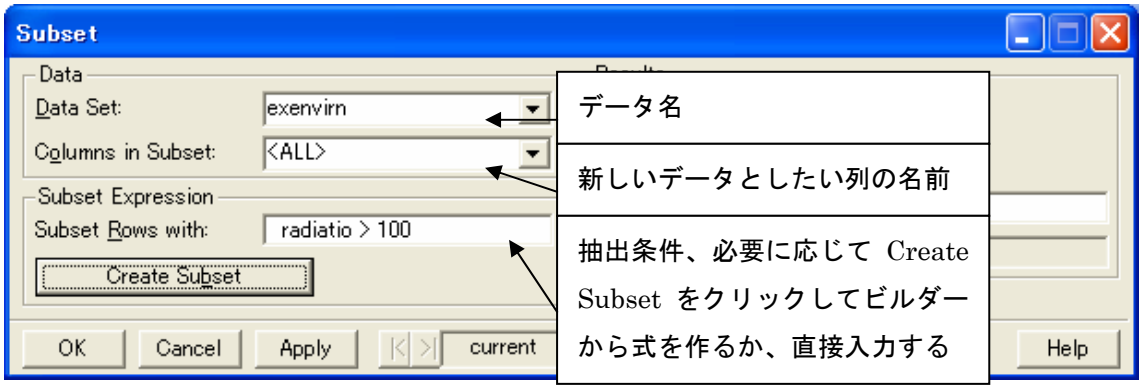
2.8 exenvirn データウィンドウから新しいデータを作る操作

今までは、Excel で行う操作のように、1 つのデータウィンドウ内で操作が完結していました。新しいデータが作られ、ウィンドウが開かれるようなことはありませんでした。
S-PLUS では、新しいウィンドウにデータをコピーペーストする操作をしなくても、メニューから条件に沿った新しいデータを作ることができます。

条件にあったデータを抽出して新しいデータとする

`radiatio` 列が 100 を超えたデータだけを抽出します。

図 2-17 メニュー データ / 行の部分集合 を選択

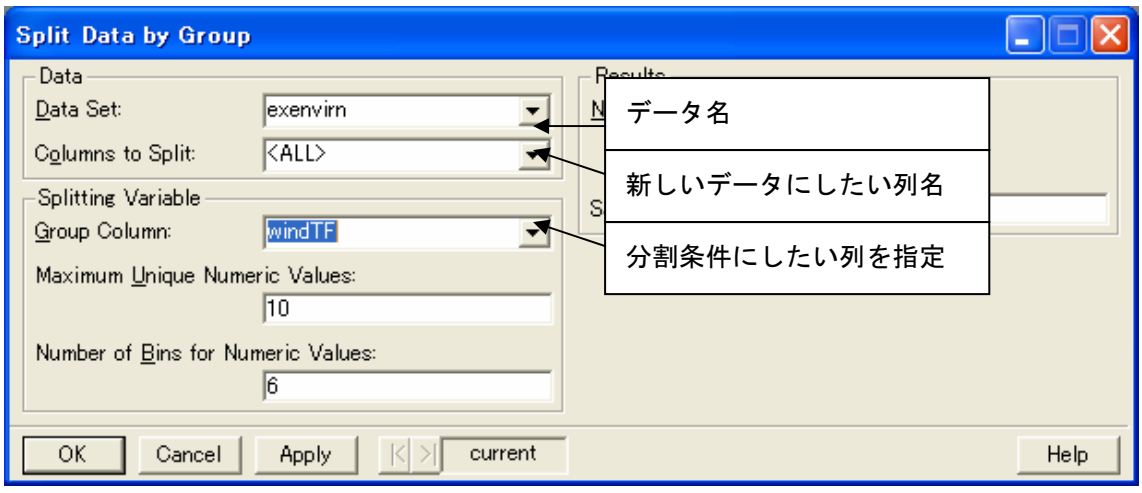


Subset Rows with に **radiatio > 100** と指定することにより、データが抽出され、新しいデータウィンドウに表示されます。このようにして作られたデータも、もちろんデータベースに保存されます。**Save In** フィールドにデータ名を指定すれば、抽出されたデータにはその名前がつきます。

データの分割

先ほど挿入した列、`windTF` が"はい"の場合と、"いいえ"の場合でデータを 2 分割します。

図 2-18 メニュー データ / 行の分割 を選択



Group Column に因子（factor）列を指定した場合はその値ごとにデータが作られますが、数値の場合は：

Maximum Unique Numeric Values:

アンケートの回答「1,2,3」のように、**Group Column** で指定した数値が、このフィールドで指定された数（デフォルトは 10）よりも種類が少なかった場合、数値ごとに分割します。つまりここで大きな数値を指定すれば、異なる数値ごとに分割される可能性が高くなります。

Number of Bins for Numeric Values:

測定値のように、**Maximum Unique Numeric Values** の数値を超えた種類の数値（連続量）だった場合、いくつに分割するかを指定します。「ある値からある値を 1 つのグループとして、1 つのデータにしたい」という決まりがある場合は、あらかじめ 23 ページの **数値を因子に変換** メニューで因子列を作っておいた方が便利です。

分かったこと

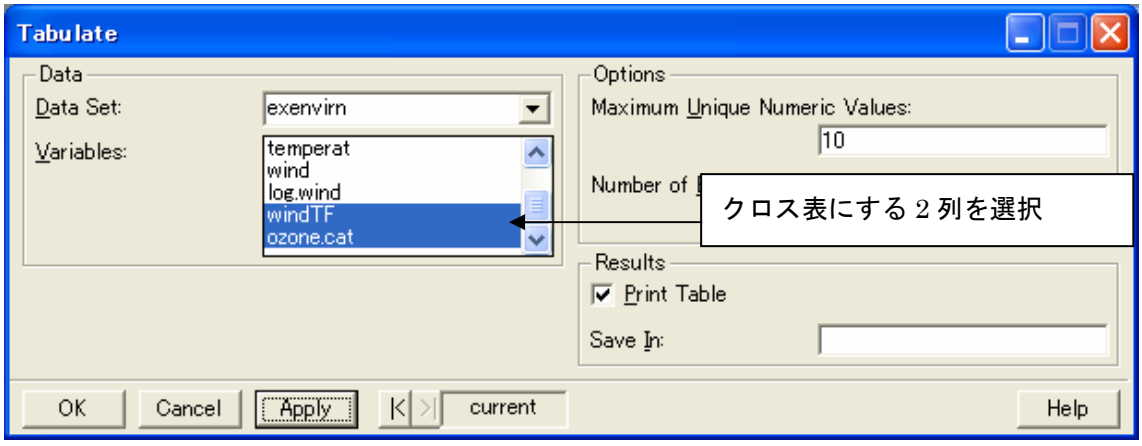
`exenvirn.はい` と `exenvirn.いいえ` の 2 つのデータができた

テーブル

`windTF` と、`ozone.cat` の間でクロス表集計を行います。

- メニュー **データ / テーブル** を選択
- **Variables** で `windTF` と `ozone.cat` をこの順番で CTRL キーを押しながら選択
- **Apply** をクリック

図 2-19 クロス表作成のダイアログ



分かったこと

Report1 というウィンドウに集計結果が出力された

さらに、**Save In** フィールドに適当な名前、例えば、`mytab` と入力し、それ以外は同じ内容で **OK** をクリックします。

分かったこと

`mytab` というデータが作られた。ただ、データは縦 x 横のクロス表形式ではなく、1 列目、2 列目にそれぞれ `windTF,ozone.cat` があり、その 2 列のすべての水準の組み合わせが作られ、`Count` 列に発生した件数が表記されている。

解説

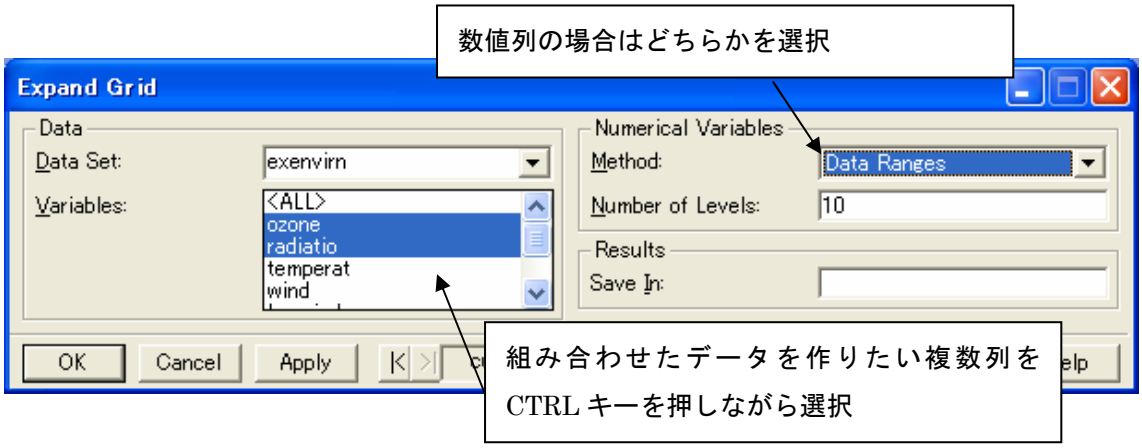
元の m 行 n 列のクロス表ではなく、このデータ形式（原因となる列のすべての水準の組み合わせに対し、発生件数が他の列として作られる）が S-PLUS の統計解析の基本になります。

組み合わせデータフレームを作る

選んだ列のすべての組み合わせからなるデータフレームを作ります。これは例えば、回帰分析の結果から予測をしたいときなどに、分析の結果と組み合わせて用いると、便利です。

- メニュー **データ / 組み合わせデータフレームを作成** を選択

図 2-20 組み合わせデータフレーム選択画面



Variables から 2 列以上を選ぶと、そのすべての組み合わせからなるデータを作ります。例えば、「はい」「いいえ」と、「A」「B」「C」なら、2x3=6 行のデータが作られます。数値列を選んだ場合は注意が必要です。右上の **Method** で指定された方法で組み合わせのデータが作られますが、デフォルトはすべての組み合わせを作る **Data Values** なので、数値データですべての値が異なれば、それだけの組み合わせができます。100 個の乱数からなる 2 列を選択すれば、100x100=1 万行のデータになります。データ範囲 **Data Ranges** を選べば、数値データを **Number of Levels** で指定した数で分割して代表値を作り、その組み合わせからなるデータフレームとします(デフォルトは 2 列を選んだ場合は 10x10 で 100 行、3 列を選べば、10x10x10 で 1000 行のデータになります)。

- **Data Set** で `exenvirn` を、
- **Variables** で `ozone` と `radiatio` を、
- **Method** で **Data Ranges** を選択

Apply をクリック、どのようなデータができるか、確認します。

分かったこと

`ozone, radiatio` ともに 10 回、同じ値が出現する。

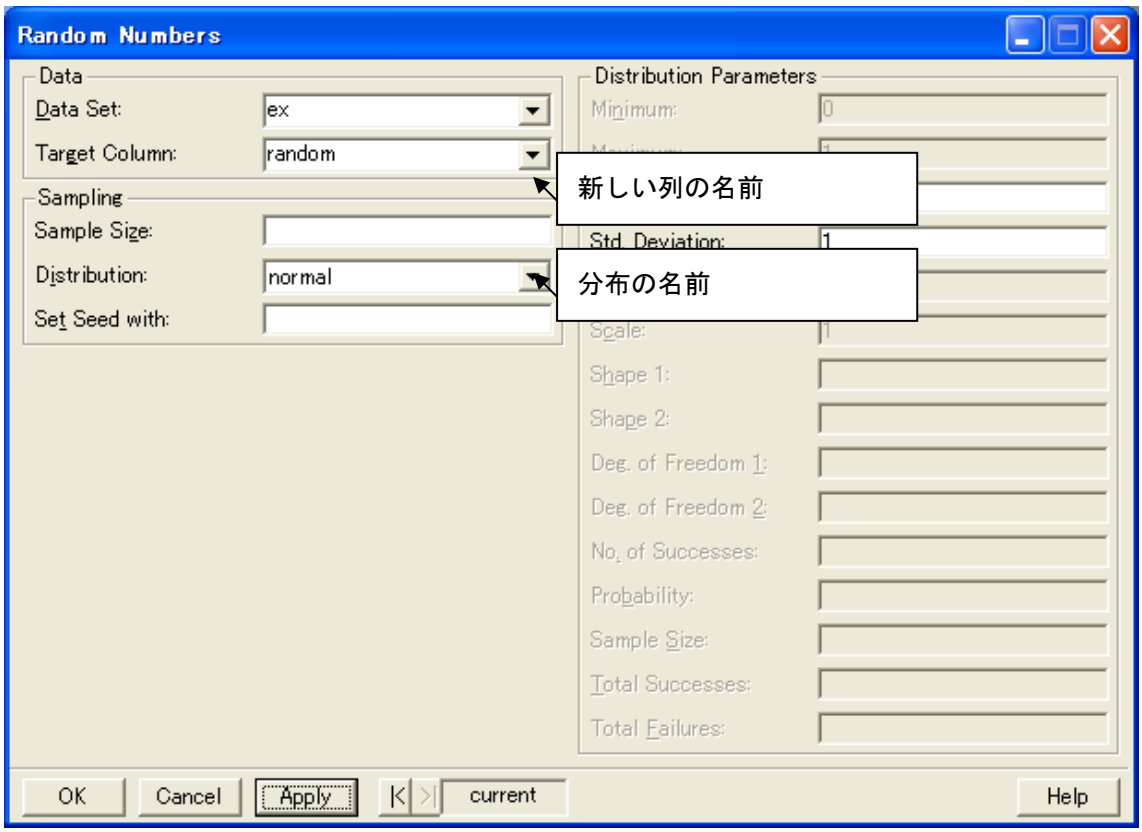
再び、同じメニューで

- **Data Set** で `exenvirn` を、
- **Variables** で `windTF` と `ozone.cat` を、
- **Method** で **Data Values** を選択
- **Save In** フィールドに `ex` と入力、**OK** をクリック

`ex` という名前で、`windTF` と `ozone.cat` の組み合わせからなるデータフレームが作られます。ここに列を追加しますが、乱数を埋めてみましょう。乱数で 1 つの列を埋めるには

メニュー **データ / 乱数** を選択

図 2-21 乱数を選択したダイアログ



- **Target Column** に `random` と指定（この名前で列が作られます）
- **Distribution** に **normal** を指定（分布関数の名前を指定します）

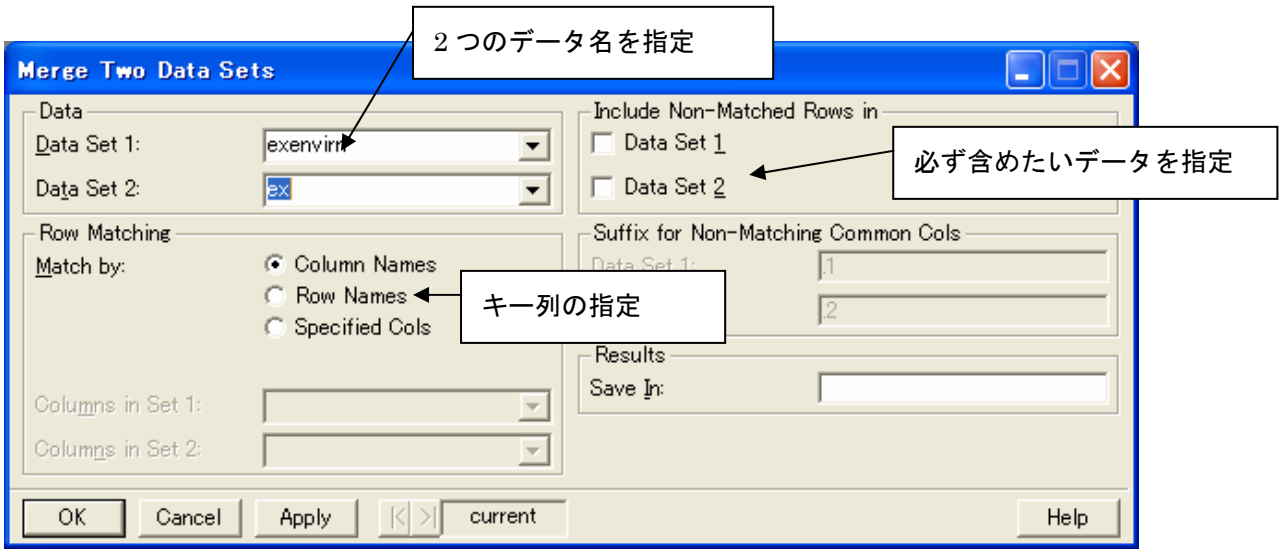
Distribution で何を指定したかによって、右側で指定可能なパラメータが変わります。デフォルトは正規分布ですが、特に平均ゼロ、標準偏差 1 の標準正規分布に従うような乱数が挿入されます（平均、標準偏差はメニューの左側から変更可能です）。
Set Seed には数値を指定することができます。後から乱数を再現させたいときに便利です。

マージ

2 つのデータを指定したキーの値が一致したところで結合させます。例として、`exenvirn` と、今、乱数を 1 列挿入した `ex` を、 列 `windTF` と 列 `ozone.cat` が等しいもの同士で結合させます。

メニュー **データ / マージ** を選択

2-22 マージを選択したダイアログ



Data Set 1: `exenvirn`
Data Set 2: `ex`

を指定し、**Apply** をクリックします。

分かったこと

新しいデータが作られ、その内容は `exenvirn` に列 `random` が追加されたもので、`random` の数字が繰り返し存在する。また、データの順番が変わっている。

キー列の指定

ダイアログでは特にキー列の指定を行いませんでしたが、自動的に `windTF` と `ozone.cat` をキー列としていることが分かりました。これはダイアログの **Row Matching** グループで規定します。

Column Names:

2 つのデータに共通の列名があった場合、その列がキーに用いられます。共通の列名が 2 列分あれば、2 列両方が等しい場合のみ、結合が行われます。デフォルトでは、この方法が用いられます。

Row Names:
行名が等しいデータ同士を結合します。

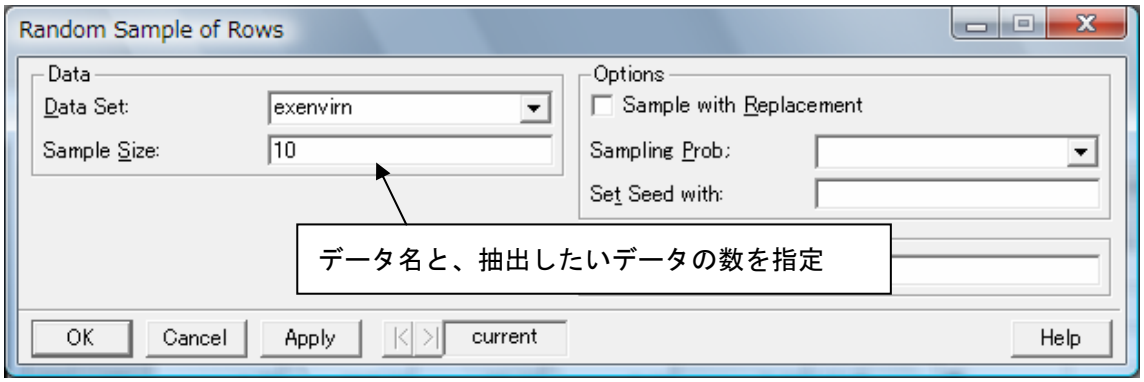
Specified Cols:
キー列を指定することができます。キー列の名前がデータ間で異なる場合、あるいは、同じ名前の列が 2 つのデータ両方に存在するが、キー列にしたくない場合に有効です。**Specified Cols** にチェックをすると、それぞれのデータから列を選択するためのフィールドが有効になりますので、キー列にしたい列を選択します。2 列以上を、CTRL キーを押しながら選択することも可能です。

データはキー列の ASCII コード順に並べ替えられ、結合されますが、元データそのもののソートを行うわけではありません。

ランダムサンプリング

元のデータから一部を乱数を用いて抽出したい場合、あるいは元データの並べ替えを行いたい場合に便利です。

メニュー **データ / ランダムサンプリング** を選択



Data Set: `exenvirn`
Sample Size: `10`

OK をクリックすると、`exenvirn` の中から、ランダムに 10 個のデータが抽出されます。**Sample Size** にデータの全個数（行数）を指定すると、並べ替えになります。さらに複雑なサンプリングはダイアログの右側の **Options** グループで指定を行うことにより、可能です。

Sample with Replacement: チェックボックスにチェックがあると、復元抽出を行います。（つぼに球を戻す場合）

Sampling Prob: 列の指定をすると、その列を確率と見立て、その大小に応じた抽出確率で抽出を行います。このフィールドではたとえば、人数、割合などが含まれている列を選択すると、人数が多い、あるいは割合の大きいデータほど、抽出される確率が高くなります。

Set Seed with: 乱数のシード（種）を固定するので、サンプリングの再現が可能になります。整数を指定します。

スタック

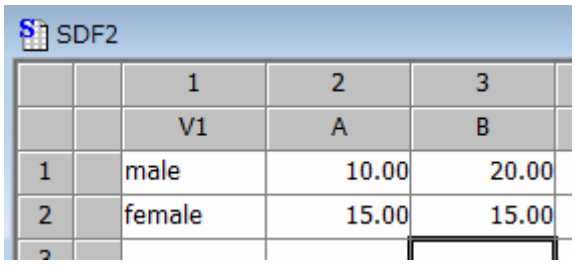
27 ページで、**メニュー データ / テーブル** を利用して、クロス表集計を行いました。その際に、**Save In** フィールドにデータ名を指定すると、クロス表の縦横の属性 2 列と、その組み合わせの場合のデータからなる 3 列のデータにして保存され、このデータ形式が S-PLUS の統計処理の基本になると説明しましたが、Excel で作られたクロス集計結果データ（ピボットテーブルなど）を用いて、統計処理をしなければならないこともあります。このようなクロス表データを、S-PLUS で統計処理に向いたデータにするには、スタックという処理を行います。

まず、クロス表として、次のようなデータを作ります。

- **ツールバー□ボタン**をクリック
- 1 列目の 1 つ目のセルから順に、下に向かって、`male, female` と入力
- 2 列目の 1 つ目のセルから順に、下に向かって、`10,15` と入力
- 3 列目の 1 つ目のセルから順に、下に向かって、`20,15` と入力
- 2 列目、3 列目の列名をそれぞれ、`A,B` と指定

データの解釈としては、ある質問に対して、A を選んだ男性が 10 人、女性が 15 人、B を選んだ男性が 20 人、女性が 15 人ということになります。

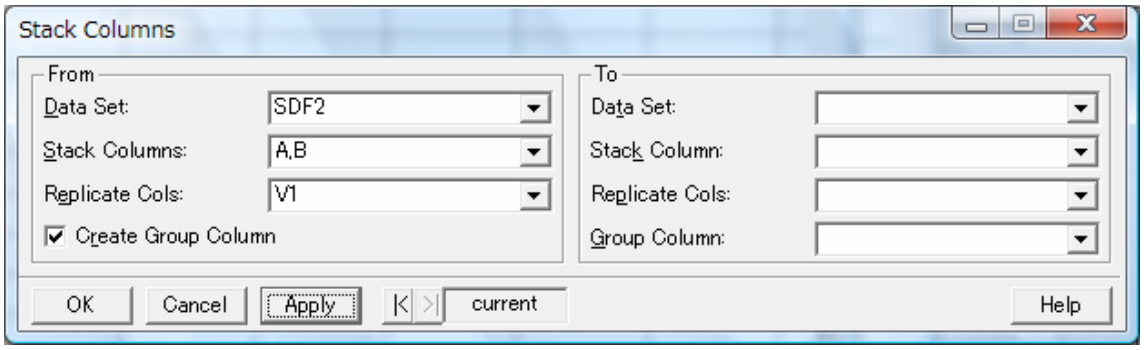
2-23 クロス表データを入力



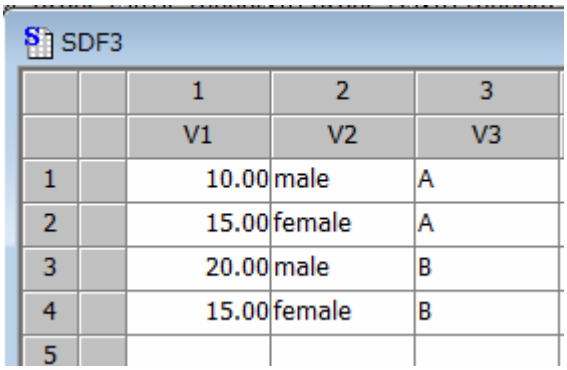
		1	2	3
		V1	A	B
1		male	10.00	20.00
2		female	15.00	15.00

- メニュー **データ / 構築 / スタック** を選択

2-24 スタックのダイアログ



Stack Columns: 値が保存されている列 A, B を選択
Replicate Cols: 繰り返したい列（属性など）を指定



SDF3

		1	2	3
		V1	V2	V3
1		10.00	male	A
2		15.00	female	A
3		20.00	male	B
4		15.00	female	B
5				

できたデータは上のようになります。この形のデータにすると、分散分析など、いろいろな統計解析をあてはめることができます（カイ二乗検定などはクロス表のままでも解析可能です)。

スタックからの回復 メニューを用いると、上記のような縦長のデータを、クロス表形式にすることができます。Excel に結果をエクスポートしたい時などに便利でしょう。

3 グラフ

3.1 S-PLUS グラフ

Excel ファイルなどからインポートしたデータを S-PLUS 上でグラフにすると、データをデータウィンドウ上で眺めていた時には気がつかなかった特徴が明らかになることがあります。S-PLUS には、Excel にはない種類のグラフが多くあります。

グラフを描画するのに適したデータ

Excel ファイルなどはあらかじめインポートをして、S-PLUS データセットにしてからグラフにします。Excel Link という機能（Excel ファイルのままグラフの作成が可能です）もありますが、制限が多いので、ここでは一旦 S-PLUS データセットにしたものを用いたグラフ作成を説明します。グラフを描画させたいデータを、14 ページの方法を参考にして、データウィンドウ上に開いてください。ここでは、`exenvirn` を用いて説明しています。

`exenvirn` でグラフにできそうなのはどの列でしょうか？`windTF` や、`ozone.cat` などは一見文字のデータ（S-PLUS では、`factor`、因子という種類のデータになっています）なので、グラフにはできないように思います。このような因子（`factor`）データがグラフの列に指定された場合、因子の順番ごとに 1,2,3…の整数が用いられ、グラフにすることが可能となります。たとえば、A,B,C という、一見文字型の 3 水準の因子の場合、A=1, B=2,C=3 としてグラフ化されます。ですから、数値、因子、時系列いずれかのデータであれば、グラフ化は可能です。Excel データがインポートされると、特に設定を変えていない場合、この 3 つの種類のいずれかになりますので、常にグラフ化が可能です。

3.2 グラフの簡単な描画

グラフは

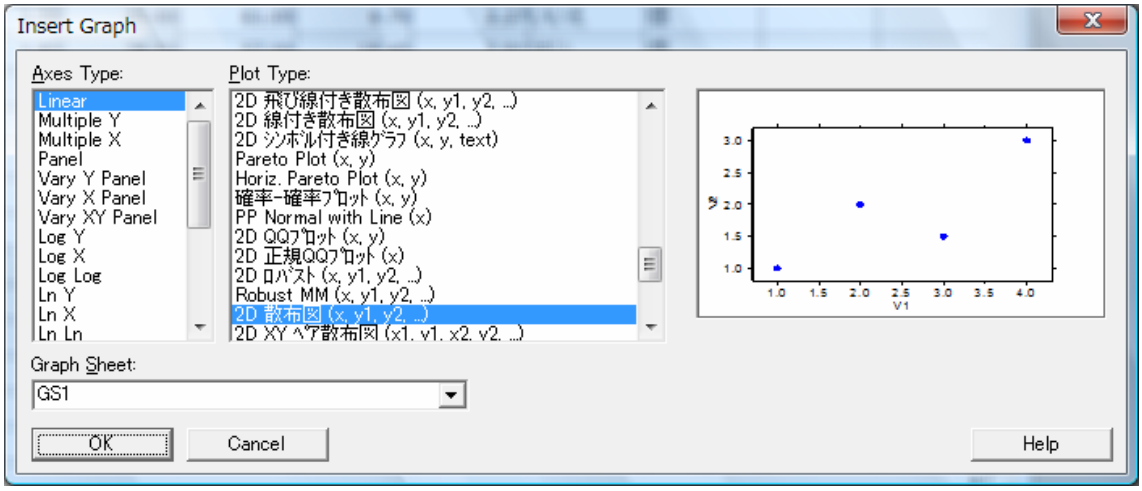
- ツールバーのグラフボタンをクリックして開いた、グラフパレット上のボタンをクリック
- メニュー **グラフ** の利用

いずれでも描画可能です。ためしに、メニュー **グラフ** を用いてグラフを作ってみましょ

う。

- メニュー **グラフ / 2D プロット・グラフ** の選択

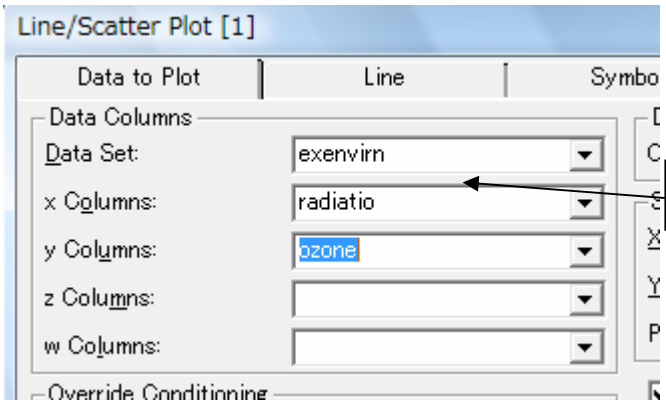
3-1 グラフのダイアログ



- Axes Type は軸の種類を指定します。詳しくは後で説明します。ここでは、Linear にします。
- Plot Type はグラフの種類です。ここでは、2D 散布図とします。右側にはそのサンプルが表示されます。
- Graph Sheet はグラフが描画されるウィンドウの名前を指定します。

OK をクリックします。続いて、X,Y 軸それぞれの列を指定するためのダイアログが現れます。

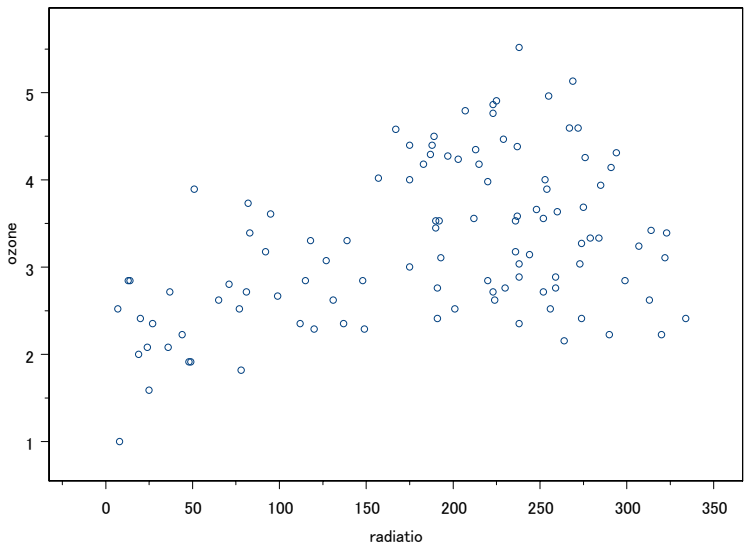
3-2 データと、X,Y 列を選ぶダイアログ



x Columns:radiatio
y Columns:ozone

を選択し、OK をクリックします。X 軸を radiatio、Y 軸を ozone とした散布図が作られます。

3-3 出来上がった散布図



分かったこと

ウィンドウ **GS1** が開き、グラフが描画された。

解説

GUI によって作成したグラフは新しいウィンドウ上に描画されます。このウィンドウを**グラフ・ウィンドウ**と呼びます。

3.3 グラフの保存

グラフはデータと異なり、ドキュメントなので、閉じる前に保存をしないと失われてしまいます。保存の方法には 2 通りあります。

- S-PLUS グラフシート形式で保存
- イメージファイル形式で保存

S-PLUS グラフシート形式で保存

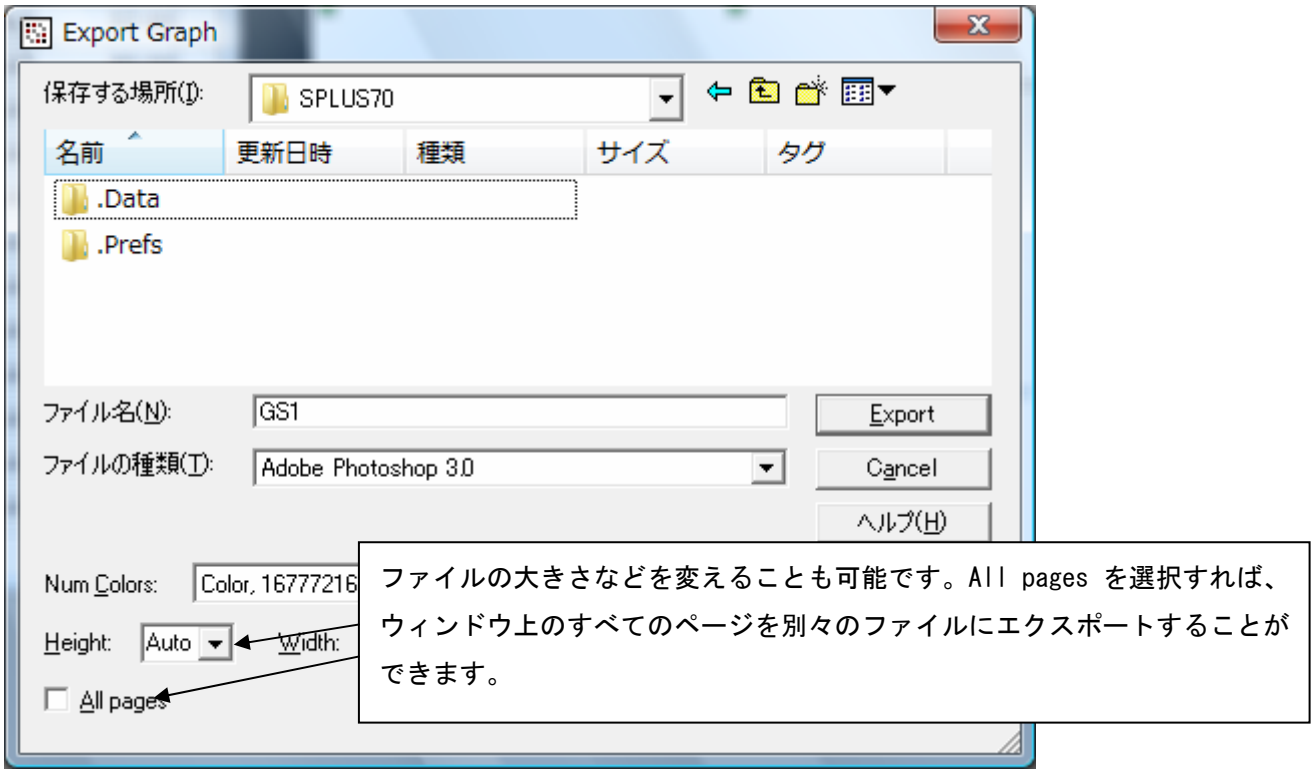
グラフウィンドウをアクティブにして、メニュー **ファイル / 名前を付けて保存** として、保存する場所とファイル名を指定します。

このファイルを開くには、S-PLUS が必要です。また、再度描画させるためには、元になっているデータ（このグラフでは、`exenvirn` というデータ）をユーザーがデータベース上に持っていないと描画されません。ファイルに保存されるのは、どのデータのどの列をどういうグラフにするかという情報だけだからです。逆に、データを更新すれば、その更新はすぐにグラフに反映されます。

イメージファイル形式で保存

グラフウィンドウをアクティブにして、メニュー **ファイル / グラフのエクスポート** として、保存する場所とファイル名、ファイルの種類を選びます。ファイルの種類で選んだ形式によって、イメージファイル形式が変わります（jpeg, wmf など）。**Export** ボタンをクリックします。

3-4 グラフエクスポートのためのダイアログ



出来上がったグラフはブラウザや Word ファイル上で見ることができます。ただ、データの値などを変えても、画像ファイルには反映されないので注意します。

まだ編集途中のグラフや、データが変わるかもしれないグラフの場合は、 **名前を付けて保存** が便利でしょう。もうこれ以上変更する予定がなく、多くの人に見てもらいたいグラフの場合は、**グラフのエクスポート** を選択します。

3.4 プロットパレット

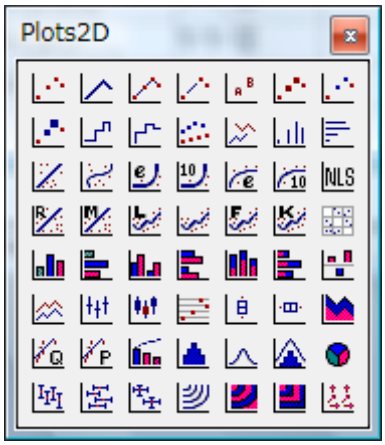
アクティブになっているデータウィンドウとプロットパレットを用いても、グラフを作ることが可能です。一般的に、描画のためのクリック回数はプロットパレットによる方法の方が少なくなります。

パレットを開く

ツールバーの **2D プロットボタン** をクリックします。**2D プロットパレット**が開きま

す。

3-5 2D プロットパレット



　いろいろなアイコンが 1 つのパレットに集められています。アイコンはグラフの種類を表現しています。

パレット上のボタンをクリックして、グラフを作る

exenvirn データウィンドウから、

- temperat 列を選択、
- CTRL キーを押しながら、ozone 列を選択
- パレット上の **線グラフ** **ボタン** をクリック

分かったこと

　temperat が X 軸、ozone が Y 軸の、各点を線で結んだグラフが新しいウィンドウに描画された。
　線がジグザグに引かれている。

解説

グラフの X,Y は、列の選択順番で決まります。最初に選択した列が X 軸、次が Y 軸になります。

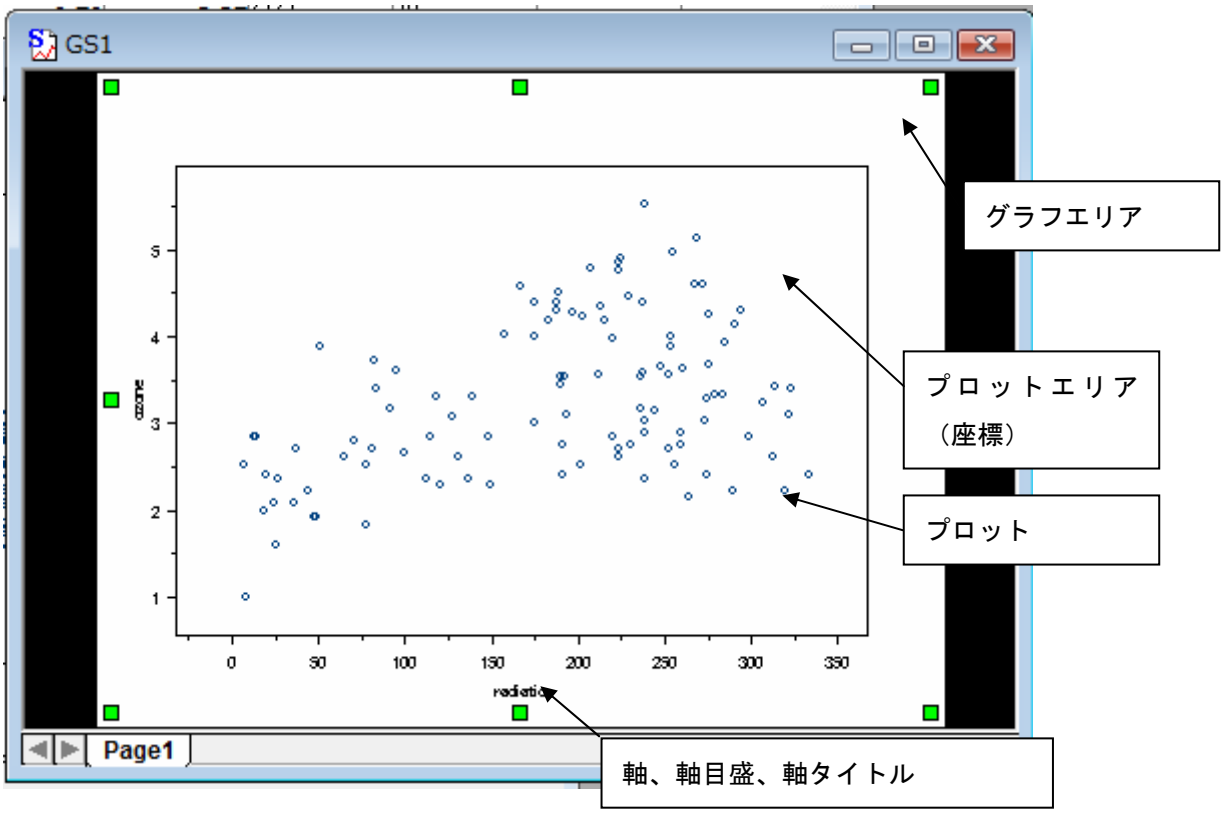
線がジグザグにひかれているように見えるのは、値の大小ではなく、線がデータの順番に（データの 1 番目の X,Y 組）（データの 2 番目の X,Y 組）…と結ばれているからです。これを、X 軸の値の小さい順に並べ替えるのは、すぐ後で説明します。

3.5 グラフの編集

GUI のメニューやボタンを利用して作成したグラフは、多くのオブジェクトの集まりで構成されています。グラフのオブジェクトには次のようなものがあります。

グラフオブジェクトの種類と内容

オブジェクトの名前	領域	変更可能な内容
グラフシート	ウィンドウ上の白い、グラフが描画される領域	グラフの向き（縦横）、余白の大きさ、利用可能なフォント、背景色
グラフエリア	グラフ軸の外側でかつ、シートの内側	グラフ全般に関する情報
プロットエリア	実際に作図される領域、座標軸の内側で点など以外	グラフ全般に関する情報
プロット	グラフの点やシンボル	点や線の形や大きさ、太さ、色など
軸	X 軸、Y 軸	軸の範囲や目盛の位置
軸目盛	それぞれの軸の目盛	目盛のフォントや表示形式
軸タイトル	それぞれの軸についているタイトル	タイトル文字、大きさ、フォント

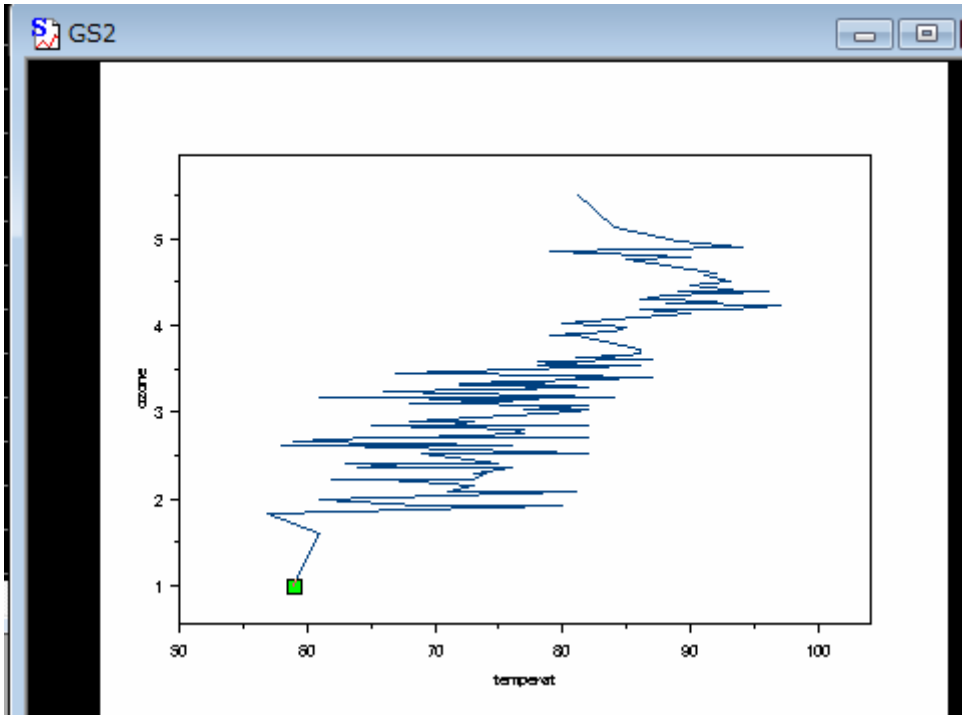


グラフの編集には、対応するオブジェクトを選択して、ダブルクリック、あるいは右ボタンクリックでメニューを表示させて、選択します。

点の並べ替え

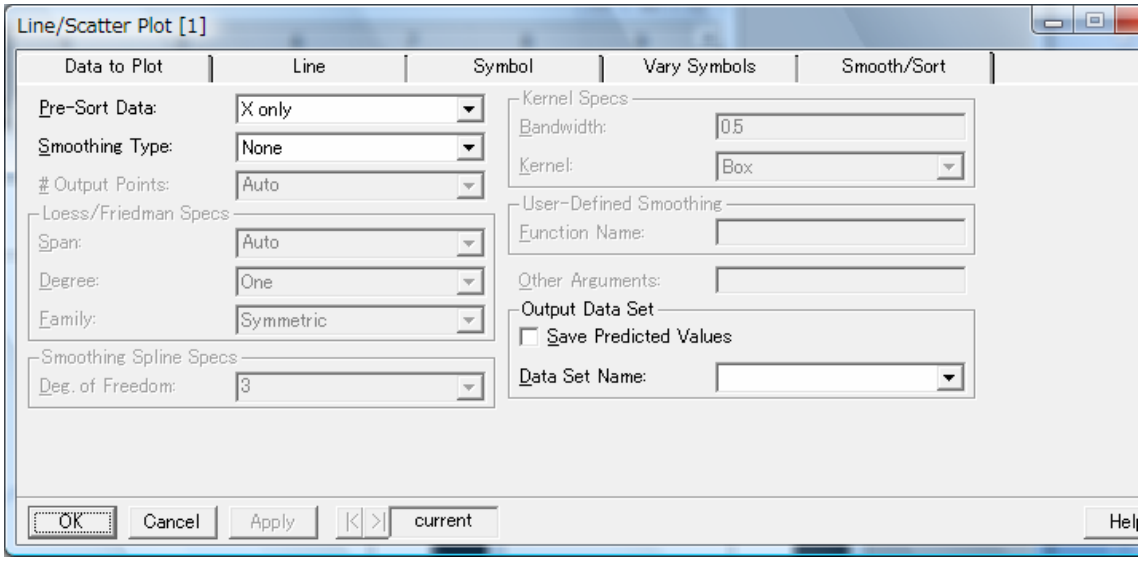
先ほどの線グラフで、線を引く順番を X 軸の小さい順に変えます。

3-6 並べ替える前のグラフ



- 線に関する変更なので、プロットオブジェクト（この場合は、線グラフの線）を選択、
- 線に１つだけ、四角い緑のマークが表示されたら**右ボタンクリック**
- メニューから **Smooth / Sort** を選択

3-7 メニュー **Smooth / Sort** を選択したダイアログ



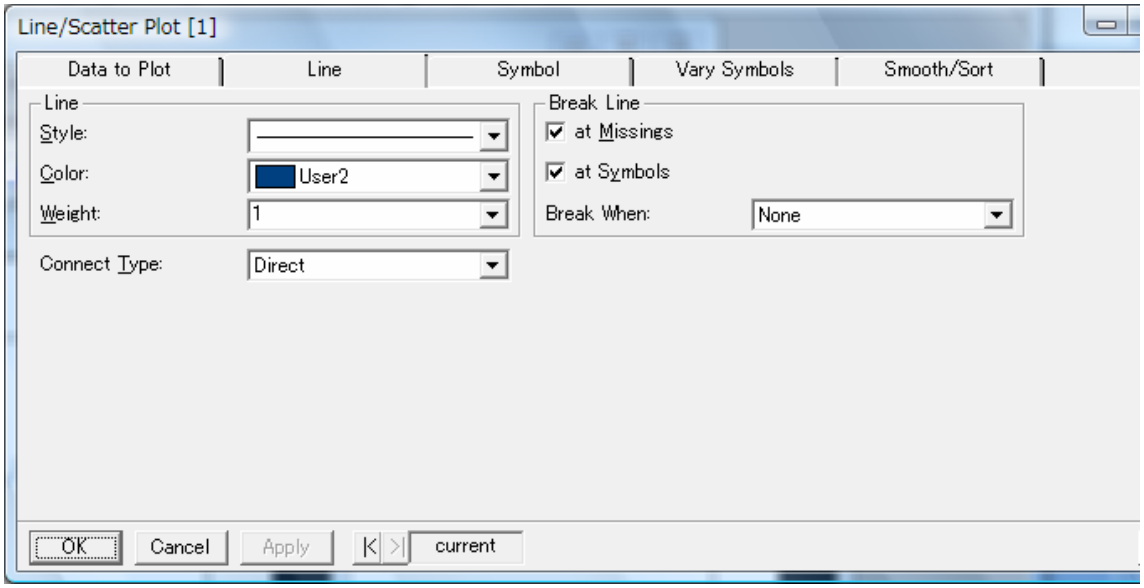
- ダイアログの **Pre-Sort Data** フィールドから **X only** を選択し、**Apply** ボタンをクリック

線は X 軸の値順に並べ替えられました。

続いて、同じダイアログの **Line** タブを用いて、線の色や太さなどを変更します。

- **Line** タブを選択

3-8 Line タブを選択したダイアログ



Line タブのフィールド名とその説明

フィールド名	説明
Style	線の種類
Color	色
Weight	太さ
Connect Type	どのように線を結ぶかの指定 **Direct:** 直結 **Isolated Points:** 細かい点で結ぶ **Vert, Horiz:** 点から次の点まで、垂直、あるいは水平に移動した

	距離を表す線だけを引く **To:** 点から指定場所まで（ゼロ、最小値）垂直、あるいは水平な線を点ごとに引く
Break Line	線を結ばない場所の指定 **at Missing:** 欠損値があれば、その前後は線を結ばない **at symbols:** シンボルのある場所は線を結ばない **Break when:** 一旦、線を引くのを中断する、たとえば、**X Decrease** は、X の値が前の X の値より小さければ、大きい方から小さい方への線を結ばない

必要に応じて、フィールドの値を変更し、グラフを編集します。

点グラフやそれ以外のグラフを作成した時も、同様に対応するオブジェクトを選び、ダブルクリックや右ボタンメニューを利用して、編集を行います。

3.6 プロットの追加

いままで作成したグラフでは、必ずグラフが新しいウィンドウに描画されました。既存のグラフにその座標を利用して、点などを追加する方法を説明します。

線グラフに、同じデータを用いた点グラフ（散布図）を追加します。

プロットエリアの選択、座標軸の内側で、点のない場所をクリックすると、座標軸の周りに 8 個の四角いマークが現れます。これが、**プロットエリアが選択された**状態です。

- `exenvirn` データウィンドウの `temparat` と `ozone` を、この順番で CTRL キーをクリックしながら選択
- SHIFT キーを押しながら、Plots2D パレットの **散布図ボタン** をクリック（SHIFT キーを押さないと、新しいグラフウィンドウが開き、そのウィンドウに散布図が描画されます）

3-9 既にあるグラフに点を追加する

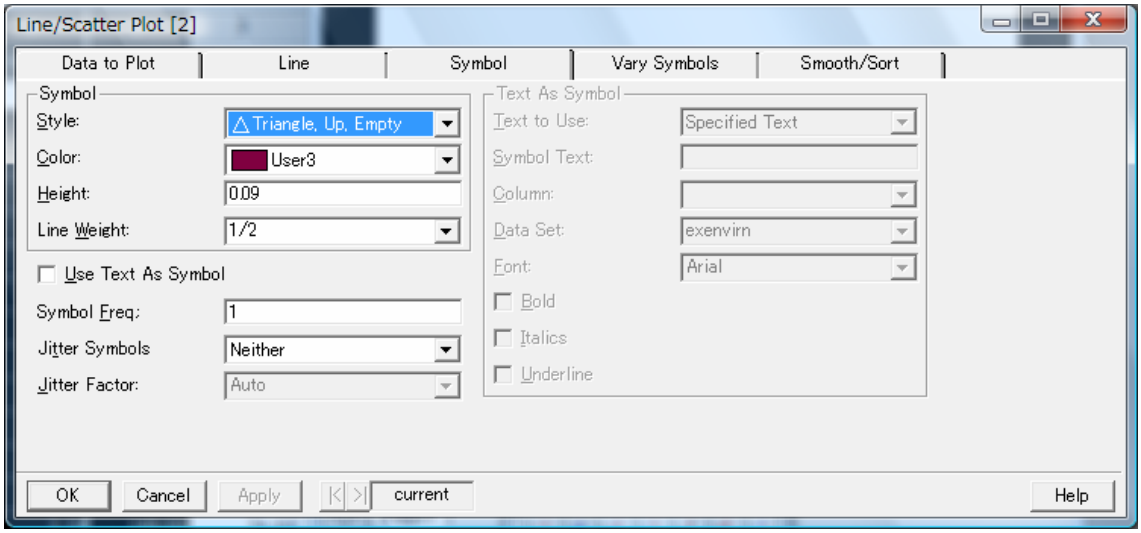
今ある座標軸に、点が追加されています。しかしこの点は線グラフに重なりません。これは、線グラフが X の値順に並べ替えられていたためです。先ほど同様、プロットオブジェクトを選択し、編集を行います。このグラフでの**プロットオブジェクト**は点です。

- 点をどれでも 1 個選択し、クリック
- 右ボタンクリック、メニューから **Smooth/Sort** を選択
- ダイアログの **Pre-Sort Data** フィールドから **X only** を選択し、**Apply** ボタンをクリック

これで、点と線が重なりました。続いて、点の編集を行います。

- **Symbol** タブを選択

3-10 Symbol タブを選択したダイアログ



Symbol タブのフィールド名とその説明

フィールド名	説明
Style	点の種類
Color	色
Height	直径
Line Weight	（線の）太さ
Use Text as Symbol	点の代わりに、文字をプロットする場合はチェック
Symbol Freq	何個のデータにつき、1 個の点をプロットするか、たとえば 2 とすれば、2 個のデータにつき、1 個の点をプロットする
Jitter Symbols	特定した座標軸の方向の座標に、小さい乱数を足す
Jitter Factor	数値を指定、上記乱数は標準でデータ範囲の 50 分の 1 だが、その値の Jitter Factor 倍する
Text to Use	文字を点の代わりに使う場合、どの文字を使うか、Specified Text ならば、固定文字、それ以外は文字をもつ列の指定
Symbol Text	固定文字の場合の文字指定
Column, Data Set	指定文字のデータ名、列名

Font	文字のフォント

3.7 グラフの追加

SHIFT キーを押しながら、プロットパレットのボタンをクリックすると、プロットが既存のグラフに追加されました。これは、**プロットエリア（座標）**が選択されていたときに、SHIFT キーとグラフボタンをクリックしたためです。

では、今度はそれ以外のオブジェクトを選択して、同じ操作を試します。たとえば、点（プロットオブジェクト）あるいは、軸タイトルなどのオブジェクトを選択します。

- `Exenvirn` データウィンドウから、列 `temperat` と、`ozone` を選択
- SHIFT キーを押しながら、Plots2D パレットの**散布図ボタン** をクリック

分かったこと

グラフが画面に 2 つ並べられて表示された。

解説

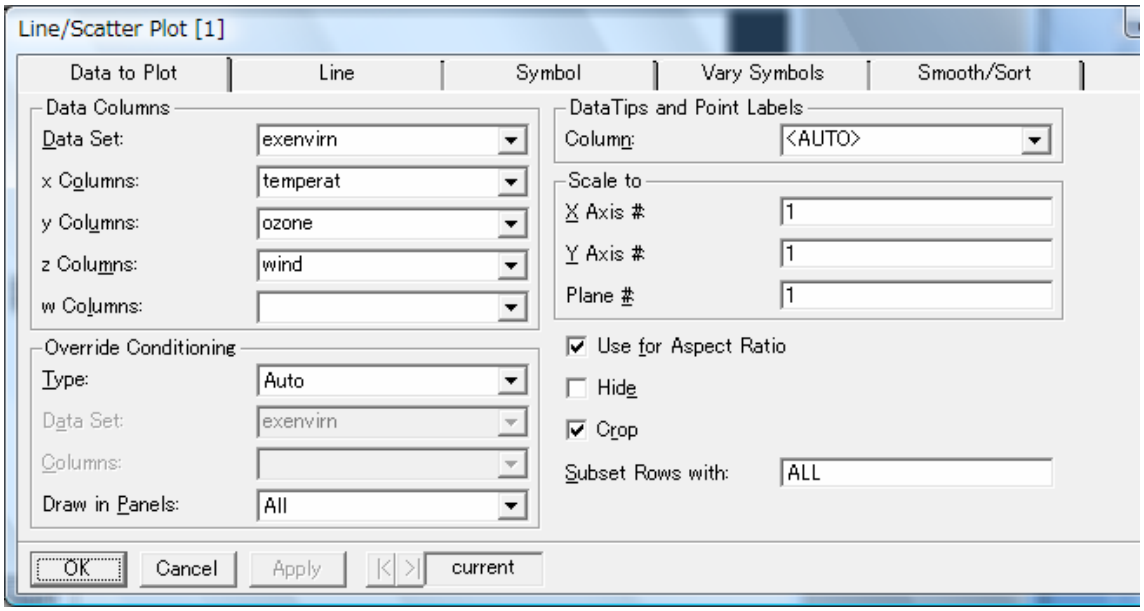
この操作により、プロットはグラフではなく、グラフウィンドウに追加されます。
つまり、オブジェクトのうち、プロットエリアを選択した状態で SHIFT キーとグラフボタンをクリックしたときだけ、点や線が座標軸内に追加されます。それ以外のオブジェクトが選ばれた時には、異なるプロットエリア（座標軸）を持つグラフがウィンドウに追加されます。

このようにグラフでは、どのグラフオブジェクトを選択したかによって、操作の内容が変わります（42 ページも参照）

3.8 さらにプロットの編集

続いて、プロットオブジェクトを選択して、どのような編集が可能か、説明します。ウィンドウに追加した `temperat` と `ozone` の散布図の点を 1 つ、クリックします（オブジェクトの選択）。右クリックし、メニューから **Data to Plot** を選択します。

3-11 Data to Plot タブを選択したダイアログ



Data to Plot タブのフィールド名とその説明

フィールド	説明
Data Set, x Column, y Column, z Column, w Column	データと列の選択
Override Conditioning グループ	後で説明する、マルチパネルプロットを描画する際に有効。62 ページ マルチパネルグラフを参照
Data Tips and Point Labels グループ	点をポイントしたときに表示されるデータティップ、あるいは点ラベルにどの列の値を表示させるか、「例」を参照
Scale グループ	座標軸の数、たとえば、**X Axis #** が **2** ならば、X 軸が上下 2 軸になる。**Plane** は 2 次元グラフを後で説明する 3 次元グラフに投影するときに便利
User for Aspect Ratio	縦横比を自動的に計算する場合はチェック
Hide	隠す
Crop	チェックがあれば、座標軸の範囲にはいりきらない点

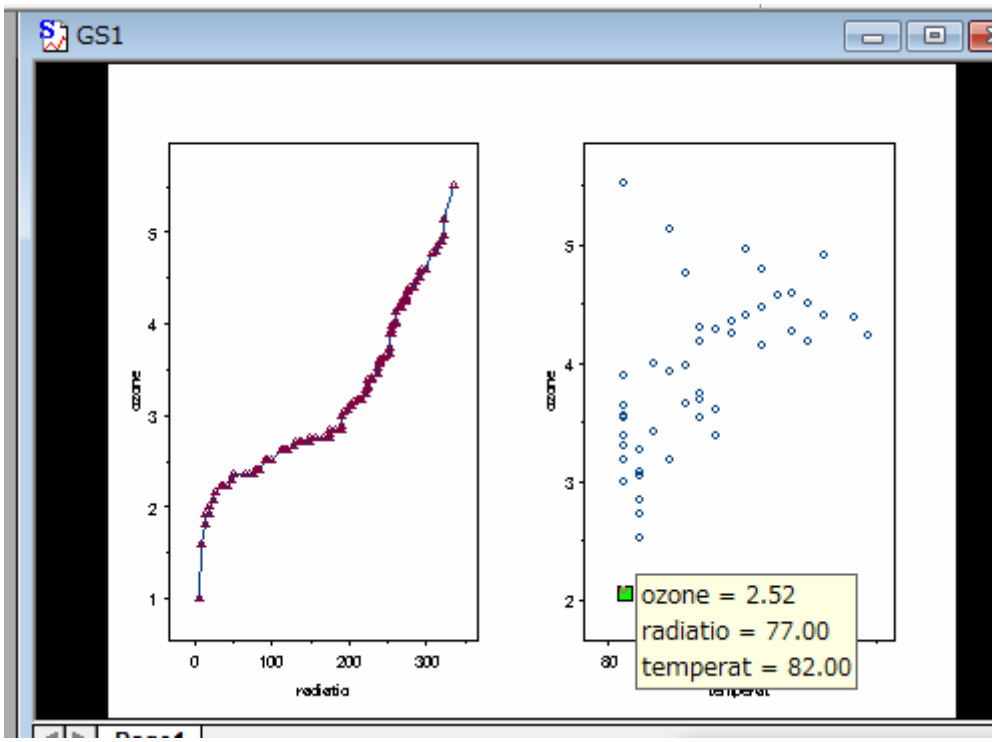
	を描画しない
Subset Rows with	描画したい点をある条件に従って特定したい場合、その条件式を指定、たとえば `exenvirn` データで `temperat` が 80 以上のデータだけを描画するには **temperat>80** と入力

Data Tips and Point Labels グループの **Column** フィールドを指定すると何が起こるか、試します。

- 右上の **Column** フィールドのリストボックスから、CTRL キーを押しながら、`ozone`, `radiatio`, `temperat` の 3 つを選択、**OK** をクリック
- グラフウィンドウの右側（今、プロットオブジェクトをメニューから編集したオブジェクト）のグラフ上で、点の 1 つにマウスを移動（クリックではありません、点の上にマウスがのるように移動）

点に関する情報が現れます。

3-12 点の上にマウスポインタを移動させ、データティップを表示する

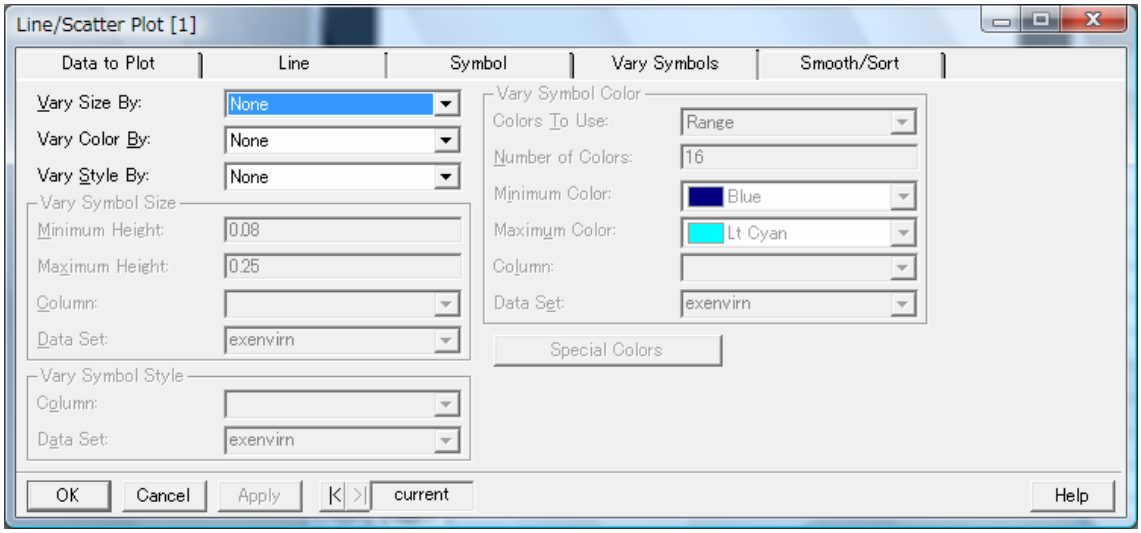


この情報をグラフ上にテキストとして貼り付ける方法を 55 ページで説明します。

グラフのシンボルを条件によって変える

プロットオブジェクト（点など）を 1 つ選択して右ボタンクリック、メニューの **Vary Symbols** を選択

3-13 Vary Symbols タブを選択したダイアログ



Vary Size By, Vary Color By, Vary Style By のリストボックスで、それぞれ列を選びます。選択した列の値の大小により、プロットのサイズ、色、形が変わります。それぞれの **Vary** リストボックスで **Other Column** を選択すると、対応する **Vary** グループで、データや列を選択できるようになります。**Size** ならばシンボルの直径に対する最大値、最小値、**Color** ならばグラデーションで表現する場合の開始、終了色、その段階数を指定することができます。

たとえば、**Vary Color By** で **Other Column** を選択し、右側の **Vary Symbol Color** グループで **Column** に `wind` を選択して、グラフがどのようになるか、確認します。

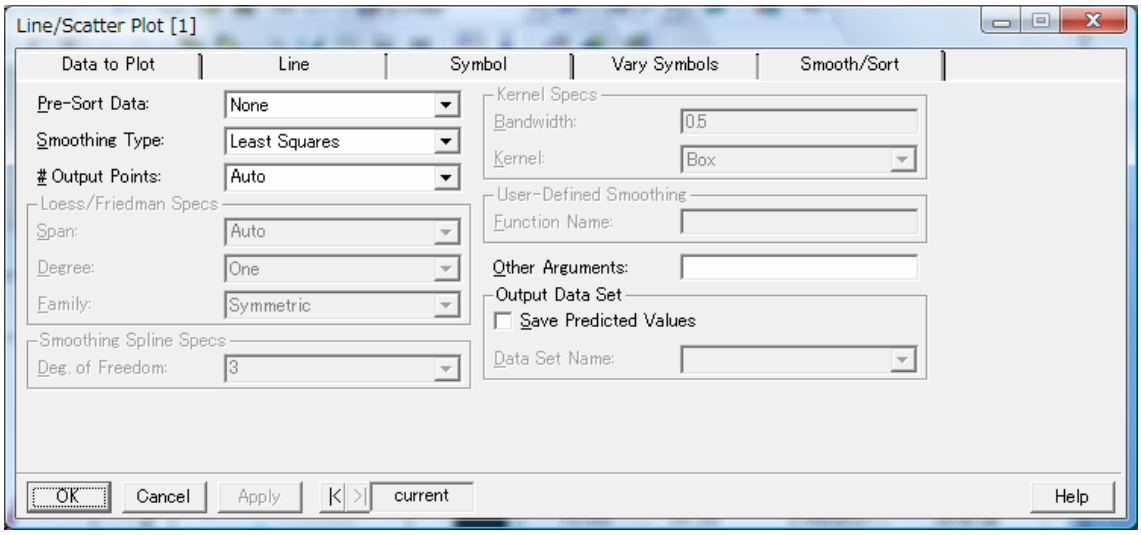
回帰直線をグラフに挿入

現在の、`temperat` と `ozone` のグラフに、この 2 列のデータの関係を表す回帰直線を

挿入します。

- **プロットオブジェクト**（点など）を 1 つ選択して右ボタンクリック、メニューの **Smooth/Sort** を選択

3-14 メニュー Smooth/Sort を選択したダイアログ



Pre-Sort Data は 43 ページにもあるように、グラフ描画の際に並べ替えを行う場合、有効です。

Smoothing Type リストボックスから、手法を選びます。

Smoothing Type リストボックスで選択可能な手法

選択可能な手法	説明
Least Squares	最小二乗法による回帰
Robust	ロバスト（頑健）法
Robust-MM	ロバスト MM 法（lmRobMM）
Kernel	カーネル（核）平滑化法
Loess	局所重み付き回帰
Smoothing Spline	スムージングスプライン
Friedman's Super	フリードマンスーパースムーザー
User	ユーザー定義関数

このうち、パラメトリック（式にできる）回帰法は **Least Squares** だけです。また、選択した手法により、入力可能なパラメータフィールドが変わります。まずはデフォルトの

パラメータで試してみるとよいでしょう。

例として、**Least Squares** を選びます。続いて、**Line** タブをクリック、**Style,Color** フィールドから、適当な種類の線と色を選択します。

3.9 グラフ編集のためのツールバーボタン

グラフウィンドウをアクティブにしたとき表示されるツールバーボタンで、グラフオブジェクトの編集に有効なものを説明します。

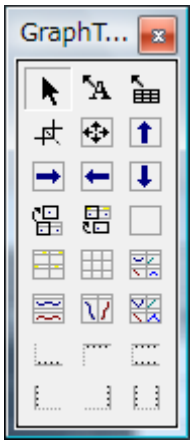
グラフ編集のためのツールバーボタン

	Send to Back, Bring to Front	複数のグラフを重ねている場合、選択したグラフを後ろ、前に移動させる
	Fill Color, Line Color, Pattern, Line Style, Line Weight	プロットエリアの塗りつぶし色、囲み線の色、塗りつぶしパターン、線の種類、線の太さ
	Auto Legend, Color Scale Legend	凡例の作成、色凡例の作成
	Graph Tools, Annotation	グラフツールパレット、図形描画パレットを開く
	Send Graph to Other App	グラフをクリップボードにコピーして他のアプリケーションで使えるようにする

3.10グラフツールパレット

ツールバーの **グラフツール** ボタンをクリックします。

3-15 グラフツールパレット



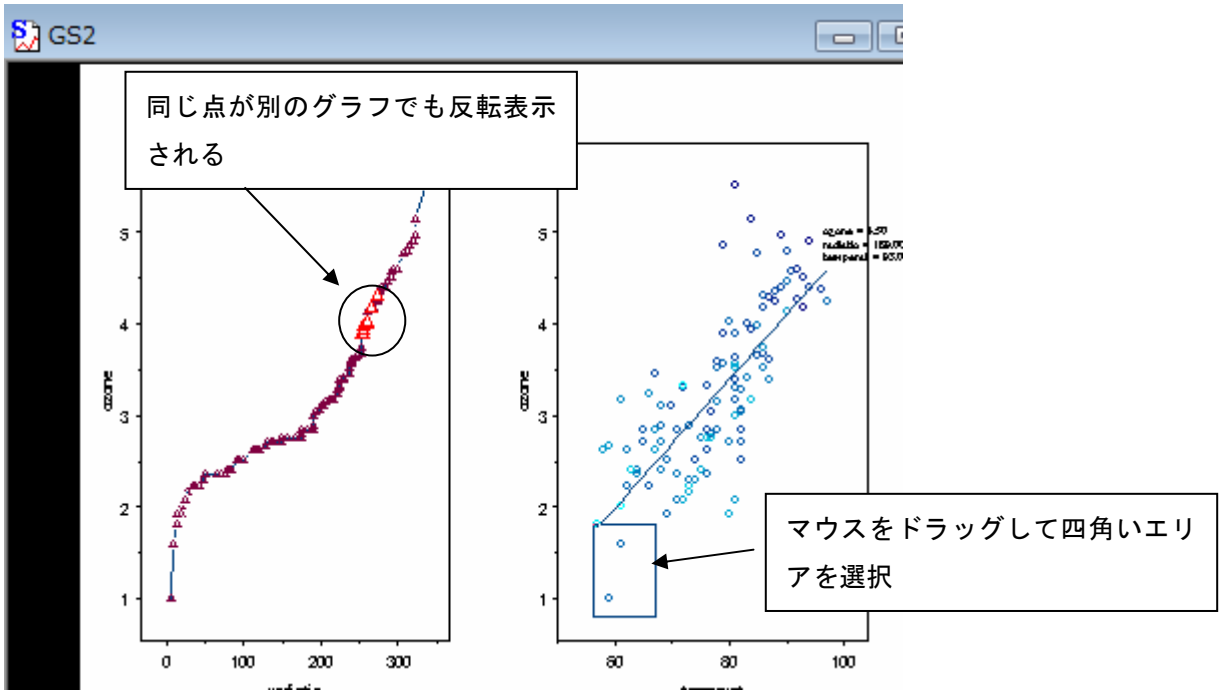
パレット上のボタンの説明（上のパレット上の位置と表の位置が対応）

オブジェクトの選択	ラベルを表示	データの選択
グラフの拡大	グラフの大きさを元に戻す	拡大したグラフで上に移動
拡大したグラフで右に移動	拡大したグラフで左に移動	拡大したグラフで下に移動
マルチパネルグラフの 1 つを取り出す	元のマルチパネルグラフに戻す	マルチパネルグラフを通常のグラフに戻す
4 マルチパネルの作成	9 マルチパネルの作成	プロットを異なるパネルに描画
Y 軸のレンジが異なるパネル	X 軸のレンジが異なるパネル	X,Y 軸のレンジが異なるパネル
X 軸目盛を下方に追加	X 軸目盛を上方に追加	X 軸目盛を上下方に追加
Y 軸目盛を左に追加	Y 軸目盛を右に追加	Y 軸目盛を左右に追加

ラベルを表示する ボタンをクリックします。マウスをグラフに移動させます。マウスポインタが+と x,y の組み合わせになっていることを確認したら、適当な点をクリックします。グラフにラベル（51 ページで表示された内容と同じです）がプリントされます。さらに異なる点をクリックすると、先ほどのラベルが消え、新しいラベルがプリントされます。同時にいくつかのラベルを表示させるには、SHIFT キーを押しながら、次の点をクリックします。

その間、マウスポインタは+と x,y のままです。作業を終えるには、ツールバーの **オブジェクトの選択** をクリックします。

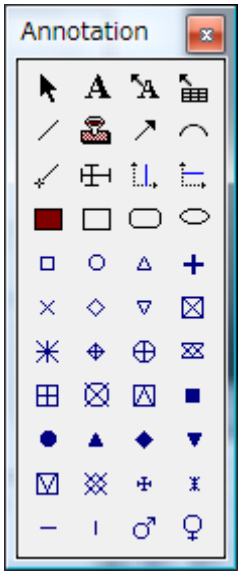
隣にある をクリックして、いくつかの点が含まれるよう、ドラッグします。つまりマウスの左ボタンを離さず移動させ、画面に四角が描かれるので、適当な場所でマウスの左ボタンを離します。



右側のグラフは色が変化するグラフのため、反転しませんが、同じデータを用いて描画されたグラフでは、その対応する点が反転表示されます。他のグラフウィンドウにそのデータを用いたグラフがあれば、同様です。また、データウィンドウ上でも、対応する行が反転しています。マルチパネルについては、62 ページの マルチパネルグラフで説明しています。

最後にツールバーの **オブジェクトの選択** をクリックしてマウスポインタを元の矢印に戻します。

ツールバーの **図形描画** ボタンをクリックします。



図形描画のためのパレットが開かれます。

図形描画パレット（上のパレット上の位置と表の位置が対応）

オブジェクトの選択	テキスト	ラベルの表示	データの選択
線	スタンプ	矢印	曲線
星付きの線	エラーバー	垂直線	水平線
塗りつぶし四角	四角	角が丸い四角	楕円

以下、同様に、既存のグラフに図形を描くためのボタンです。

ボタンをクリックしたら、グラフ上にマウスカーソルを移動し、図形（あるいは線）を描きたい場所でクリックします。図形を選択して右ボタンクリックすることにより、その図形のプロパティ（大きさなど）を変えるメニューを表示させ、変更すること、あるいは図形をマウスでドラッグして、表示する場所を変えたり、大きさを変えることも可能です。

3.11 いろいろな 2 次元グラフ

2D プロットパレットには多くのグラフボタンがあります。このうちのいくつかを用いて、2 次元グラフを作成します。作成方法は基本的に同じです。

1. データウィンドウから列を選び、2D プロットパレット上のグラフボタンをクリック
2. データを選択せず、2D プロットパレット上のグラフボタンをクリック

1.の場合はグラフがすぐに現れます。2.の場合は白紙のグラフウィンドウが開いて、グラフのダイアログが開かれます。必要な設定（データ名や列名など）を行うと、グラフが描画されます。

たとえば、2 次元グラフを作るのに、列を 1 列しか選ばなかったり、3 列選んだ場合はどうなるでしょうか？グラフによって異なりますが、概ね次のようにグラフが作られます。

データが 1 列足りない場合

たとえば、2 次元散布図を作るのに、1 列しか指定しなかった場合は、X 軸にデータの並び順を表す番号、Y 軸に指定されたデータが用いられます。

データが 2 列以上足りない場合

グラフは描画されず、グラフシート上にグラフのボタンが置かれます。このボタンをダブルクリックし、必要なデータ列を指定してください。

データが 1 列以上多い場合

別系列のグラフになります（色やシンボルで区別できます。ツールバーボタンを利用して凡例をつけると、さらにわかりやすくなります）。**等高線グラフボタン**グループ では、4 列以上のデータは**グリッドデータ**（m 行 n 列のデータに対して、何行目かという情報が Y 座標に、何列目かという情報が X 座標になる）とみなされます。

グラフ作成が可能なデータは、**数値、因子（factor）、日付時間(timeData)** です。
また、**列の選択順**に、x 軸データ、y 軸データ、z 軸データ、…となります。

では、次の操作をしてどのようなグラフが作られるか、確認します。
`exenvirn` データウィンドウを開いて、`ozone, raiatio, temperat` を、CTRL キーを押しながらクリックし、3 列選択します。以下のボタンをクリックします。

1. **回帰直線** をクリック
2. **X,Y ペア線グラフ** をクリック
3. **カラープロット** をクリック
4. **バブルカラー** をクリック
5. **対散布図** をクリック

気づいたこと

1.から 3.では ozone と radiatio, ozone と temperat の組が別系列でグラフ化されていることがわかります。

4.はエラーになります。4.は 4 列以上のデータが必要です（2 列が X,Y 座標、1 列が色、1 列がシンボルの大きさを決めるのに必要です)。

5.はもともと複数列のデータをグラフ化する際に使われるグラフです。

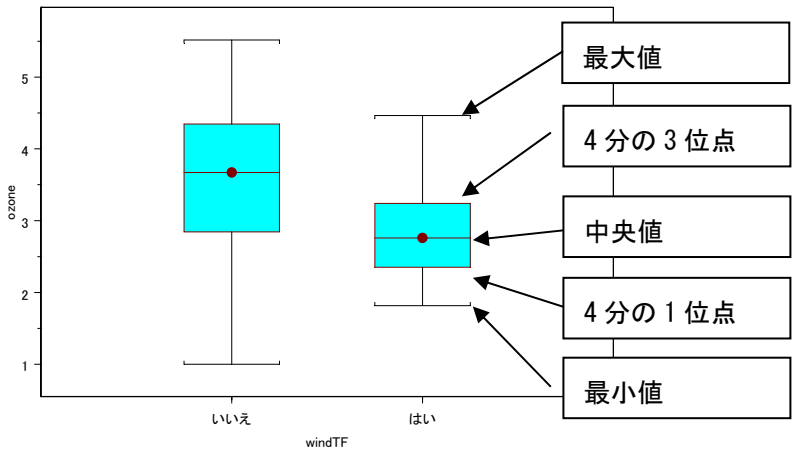
これらのグラフは数値データを直接データ 1 個につき、点を 1 個プロットするグラフです。点の数はデータの件数によりますから、データが多いと描画に時間がかかったり、画面上が点で覆われて意味をなさないことがあります。

次のグラフは集約を行うグラフです。グラフの見やすさはデータ件数にあまり依存しません。

- exenvirn の列、windTF, ozone を、CTRL キーを押しながらクリック
- **箱ひげ図** をクリック

箱ひげ図はデータの四分位範囲をグラフで表したもので、一番下の線がデータの最小値、箱の下底が 4 分の 1 位点、中央の線が中央値、箱の上底が 4 分の 3 位点、一番上の線が最大値を表します。

3-16 箱ひげ図



因子データを１列だけ指定する集約グラフ

- exenvirn の列 windTF を選択
- **円グラフ** をクリック
- **パレート図** をクリック
- **棒グラフ** をクリック

これらも、集約するタイプのグラフなので、データの数にあまり依存しないで描画ができます。

数値データを１列だけ指定する集約グラフ

exenvirn の列 ozone を選択し

- **ヒストグラム、密度曲線、密度曲線付きヒストグラム** をクリック

統計的なグラフ

箱ひげ図、パレート図、ヒストグラムなどもそうですが、統計的な判断をする材料にするグラフを作ります。たとえば、ozone は正規分布をしているでしょうか？

- exenvirn の列 ozone を選択
- **QQ プロット、PP プロット** をクリック

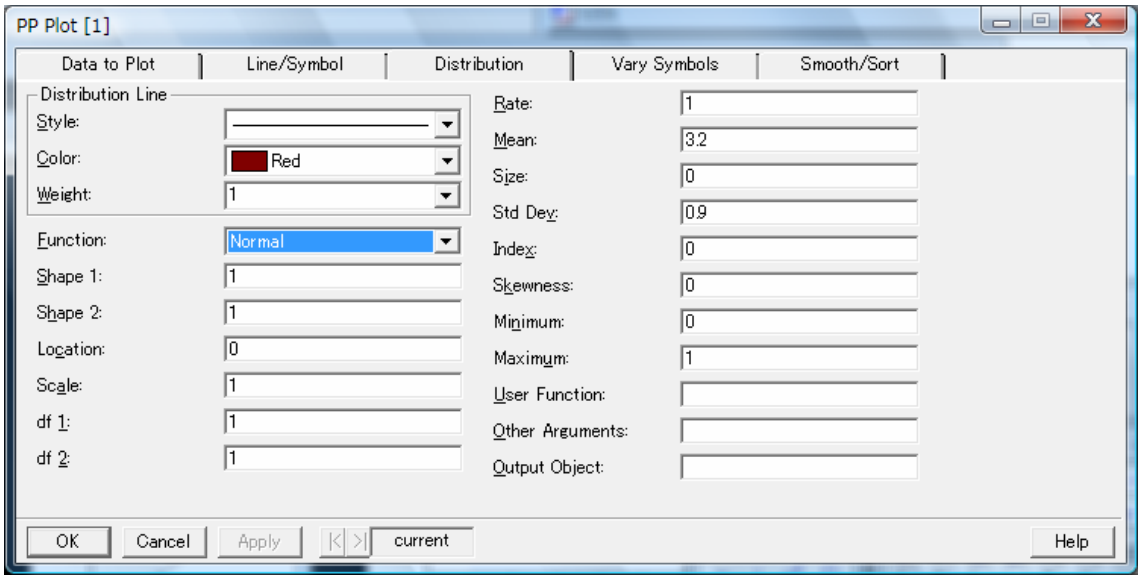
QQ プロットは指定されたデータが正規分布をしているかを確認するのに便利です。データが正規分布をしていれば、点は斜め線に沿ってプロットされます。

PP プロットでは、任意の確率密度関数の指定が可能です。

- 点を選択して、右ボタンクリック、メニューから **Distribution** を選択します。
- **Function** リストボックスから、確率密度関数の名前を選択、**Normal** は正規分布
- **Mean** に平均値を指定、ozone の平均値はおよそ 3.2 なので、3.2 と指定
- **Std Dev** に標準偏差を指定、ozone の標準偏差はおよそ 0.9 なので、0.9 と指定

指定可能なパラメータは確率密度関数の種類によって変わります。

3-17 PP プロットのダイアログ



3 次元の関係を表すグラフ

前述したカラープロット、バブルカラーなども 3 次元以上のデータに対応したグラフですが、等高線や色のついた等高線グラフをつくることもできます。

- `exenvirn` の列、`radiatio, temperat, ozone` を CTRL キーを押しながらクリック、
- **等高線、カラー等高線、イメージグラフ** をクリック

X 軸が `radiatio`, Y 軸が `temperat`, その場所における線の示す高さ、あるいは色が `ozone` の高低を表しています。

特殊な形式のデータが必要なグラフ

ハイロー、ロウソク足グラフ

エラーバーグラフ

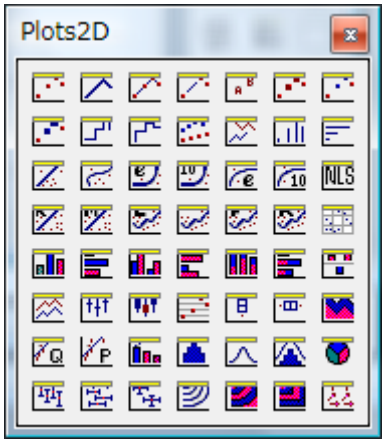
の作成には、グラフに合わせたデータが必要となります。詳しくはデータ列を選択せず、ボタンをクリック、表示されるダイアログの **Help** ボタンをクリックして、オンラインヘルプを参照してください。

3.12 マルチパネルグラフ

グラフをある条件によって、場合分けして描画することができます。

グラフを作る前に、ツールバーの **条件モードのオンオフ** **ボタン** をクリックします。2D グラフパレットを見ると、各グラフボタンの上に黄色い線が引かれていることがわかります。

3-18 パレットの上部に黄色い線が引かれているのは条件モードになっている場合



この状態で、

- `exenvirn` の列、`radiatio, ozone, windTF` をこの順番で CTRL キーを押しながら選択
- **線形回帰** をクリック

分かったこと

`windTF` が、はい、いいえの場合ごとに異なるパネルに、日射量とオゾンの量の関係を表す線形回帰の線が引かれたグラフができた

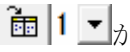
解説

最後に指定した列を場合分けの条件として、それぞれの場合ごとに異なるパネルへグラフが描画されます。

場合分けの条件として指定できるのは因子、数値いずれかです。

因子の場合は、因子水準ごとにグラフが作られます。

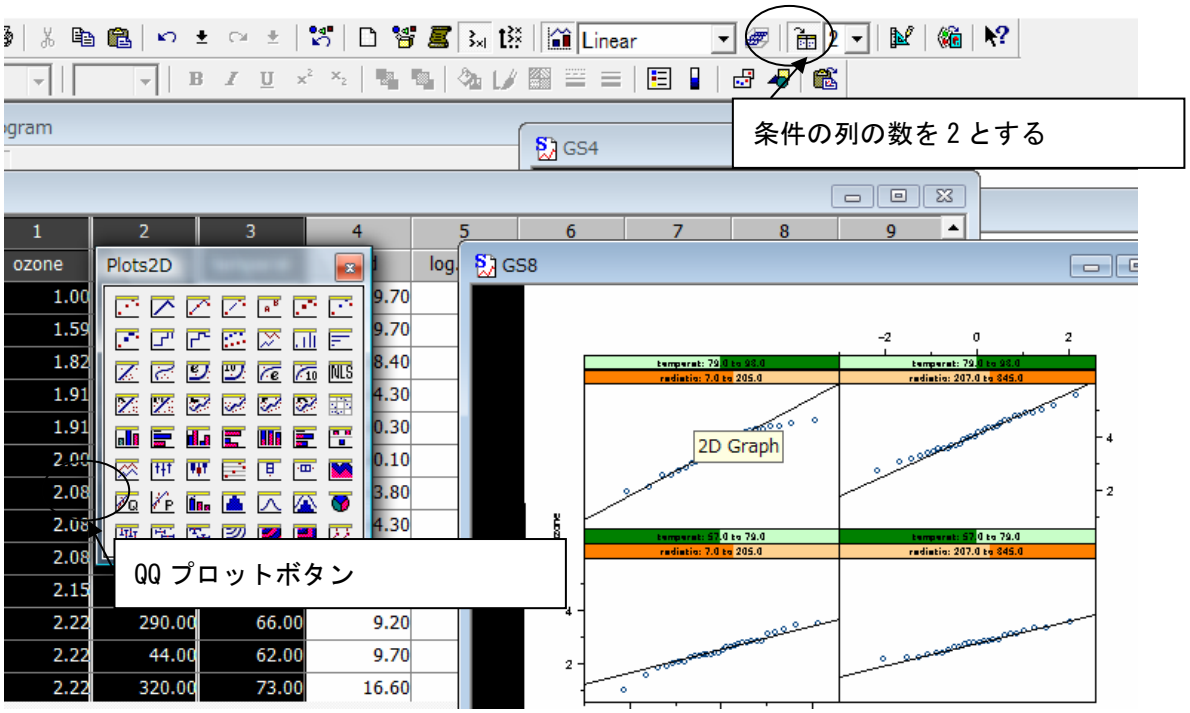
数値の場合は、特に指定がないと 4 分割でかつ、データの個数が大体同じになる場所で場合分けされます。

条件にする列の数は最大 5 列まで指定可能です。変更するには、ツールバーの **条件の列の数**  から、列数を選択します。

ここで指定した列数は列選択で優先的に用いられます。たとえば、条件の列の数を 2 とし、列を 3 列指定した場合、最初に指定された 1 列が Y 軸、あとの 2 列が条件列となります。1 列不足しますが、散布図などではデータの順番を示す番号が X 軸の値となります。

マルチパネルにできるグラフは、パレットで黄色い線が引かれているグラフ全てです。例として、条件の列の数を 1 から 2 にかえ、CTRL キーを押しながら、`ozone, radiatio, temperat` を選択し、**QQ プロットボタン** をクリックし、描画します。

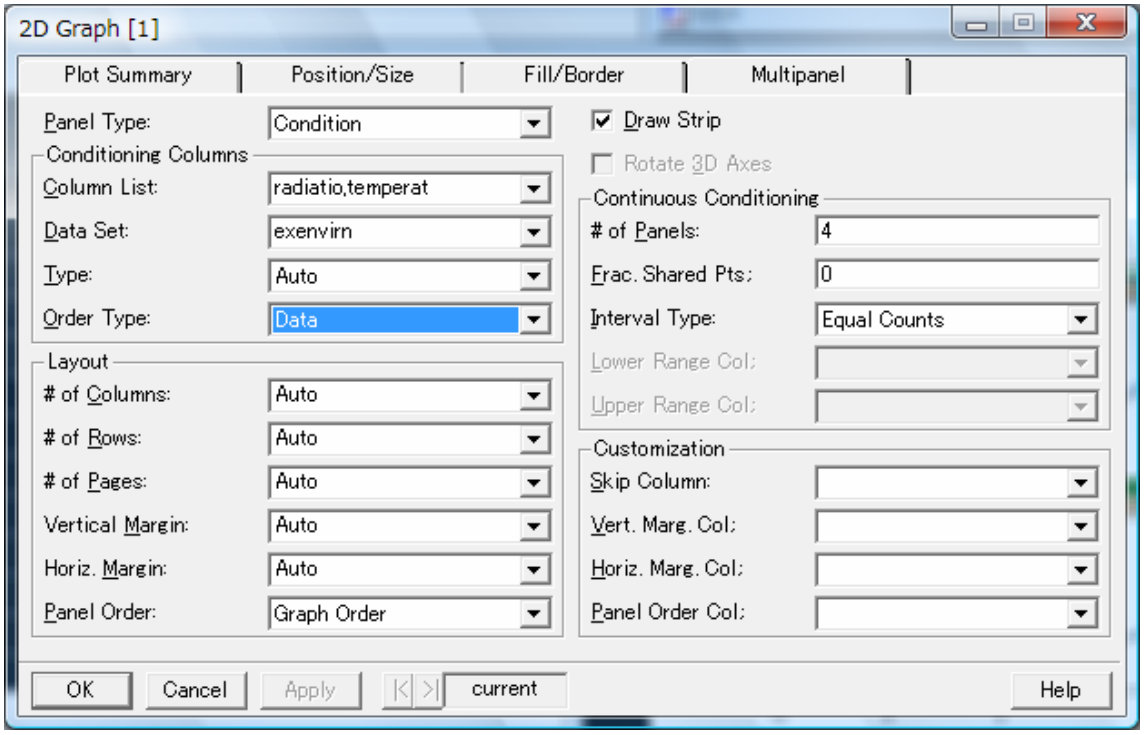
3-19 条件の列を指定する



3.13 マルチパネルの編集

分割数や、1 ページに並べるパネルの数などを変えることができます。**プロットエリア**（グラフの内側で、点などのない場所をクリック）を選択し、右ボタンクリック、メニューから **Multipanel** を選択します。

3-20 マルチパネルのダイアログ



マルチパネルダイアログの各フィールド

Column List	条件列の指定
Type	連続量（**continuous**)か離散量（**discrete**）の指定、Auto でデータから判断される
Order Type	パネルの並び順の定義
# of Columns	1 ページあたりグラフ列数、半角数字の入力
# of Rows	1 ページあたりグラフ行数、半角数字の入力
# of Pages	ページ数、半角数字の入力
Vertical Margin/Horiz. Margin	グラフとグラフの間のスペースを指定
Panel Order	**Graph Order:**左下から右上へ並べる **Table Order**:左上から右下へ並べる
Draw Strip	条件を示す各グラフタイトルを表示するかどうか
# of Panels	**Type** が **discrete** の場合は 1 種類の値につき、1 つのグラフが作られるが、**continuous** の場合、数値をいくつにグループ分けするかの指定
Frac. Shared Pts.	数値のグループ分けの際、条件近くのデータを 2 つのいずれのグループにも含まれるようにする

	ことが可能。0 から 1 より小さい数を指定すると、分割条件がその割合だけデータが重複するように計算される
Interval Type	分割条件決定方法、 **Equal Counts:** 1 つのパネルに約同数のデータ **Equal Ranges:** データの最小値最大値から等分割 **Columns of Ranges:** 指定した列にある、分割の端点を表す数値を用いる。列は **Lower Range Col.** と **Upper Range Col.** で指定
Skip Column, Vert. Marg. Col, Horiz. Marg. Col, Panel Order Col.	スキップして空白にしたいパネルがある場合を 1,表示させる列を 0 で表す数値の指定、余白の大きさを数値化した列の指定、パネルの表示順を指定した列の指定

マルチパネルグラフを作成後には、必ず **条件モードのオンオフ ボタン** をクリックし、通常のグラフモードに戻します。

3.14 軸タイプリストボックス

ツールバーの **軸タイプ** Linear **リストボックス** を指定すると、通常の線形軸だけでなく、対数軸や複数軸などに変更することができます。

軸タイプリストボックスで選択可能な軸

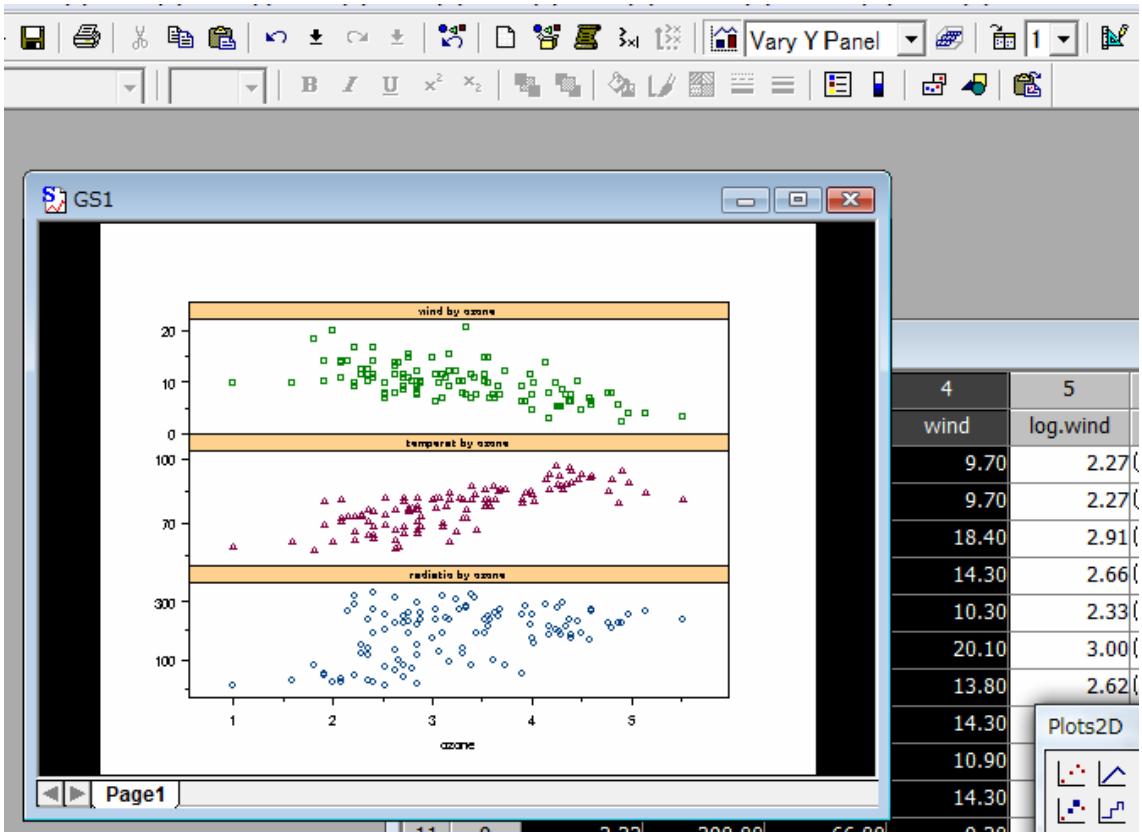
Linear	通常の線形軸
Multiple Y	複数の Y 軸を左右に表示
Multiple X	複数の X 軸を上下に表示
Panel	最初に指定した列に対して、後で指定した 2 列以上の列をそれぞれ別のグラフとし、パネル状に表示、座標は共通
Vary Y Panel	最初に指定した列に対して、後で指定した 2 列以上の列をそれぞれ別のグラフとし、パネル状に表示、Y 軸がグラフごとに異なる
Vary X Panel	最初に指定した列に対して、後で指定した 2

	列以上の列をそれぞれ別のグラフとし、パネル状に表示、X 軸がグラフごとに異なる
Vary XY Panel	最初に指定した列に対して、後で指定した 2 列以上の列をそれぞれ別のグラフとし、パネル状に表示、XY 軸がグラフごとに異なる
Log Y	Y 軸を常用対数
Log X	X 軸を常用対数
Log Log	両軸とも常用対数
Ln Y	Y 軸を自然対数
Ln X	X 軸を自然対数
Ln Ln	両軸とも自然対数
Prob Y	Y 軸を確率密度軸
Prob X	X 軸を確率密度軸
Prob Prob	両軸とも確率密度軸

軸リストボックスを指定してから、グラフボタンをクリックします。たとえば、

- `exenvirn` の `ozone, radiatio, temperat, wind` を、CTRL キーをクリックしながら選択
- 軸リストボックスで、**Vary Y Panel** を選択
- **散布図ボタン** をクリック

3-21 Vary Y Panel を選択したグラフ



この軸は、生成済みのグラフから変更することも可能です。**プロットエリア**を選択（座標軸の内側で白い部分をクリック）し、**軸タイプリストボックス** から、他の軸を選びます。

軸タイプリストボックス を変更したら、最後に **Linear** に戻します。

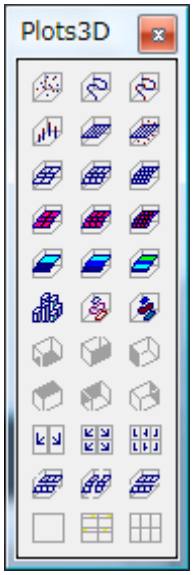
3.15 3 次元グラフ

3 次元グラフを作成します。

作成手順は 2 次元と同じですが、基本的に 3 列以上の指定が必要となります。

ツールバーの **3D プロット** **ボタン** をクリック、**3D プロットパレット** を開きます。

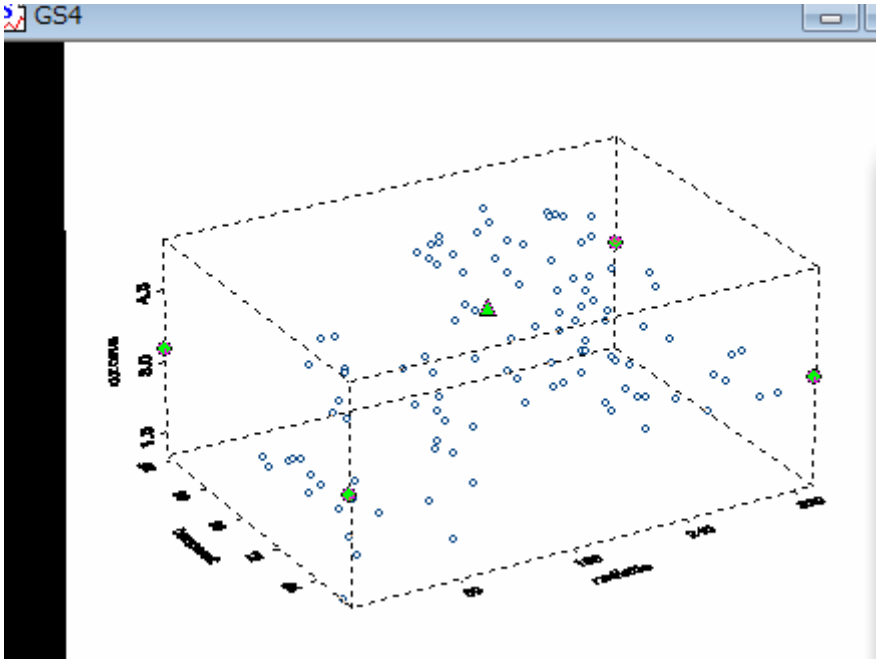
3-22 3D プロットパレット



- `Exenvirn` の列、`radiatio, temperat, ozone` を、CTRL キーを押しながら、選択
- **3D プロットパレット**から、**散布図** **ボタン** をクリック

プロットを選択して、右ボタンメニューで点や線あるいは座標軸などの変更が可能なのは 2 次元グラフと同じですが、3 次元グラフ固有の操作があります。**プロットエリア**（座標の中の、点などがない場所をクリック）を選択します。**プロットエリア** を選択すると、丸が 4 個と三角が 1 個のマークが現れます。

3-23 3D グラフのプロットエリアを選択



三角のマークを画面上、上下にドラッグします。グラフが垂直方向に回転します。マウスの左ボタンから手を離すと、その場所で固定されます。

丸のマークは画面上、左右にドラッグすることで、グラフは水平方向に回転します。

さらに、3D プロットパレットの **4 方向に回転角で分割ボタン** をクリックすると、3 次元グラフを 4 方向から見た 4 つのグラフになります。パレットの **条件なしボタン** をクリックすると、グラフを 1 画面に戻すことができます。さらに、**X,Y,Z データで条件づけボタン** のいずれかをクリックすると、グラフは指定座標のデータの大小で分割されます。

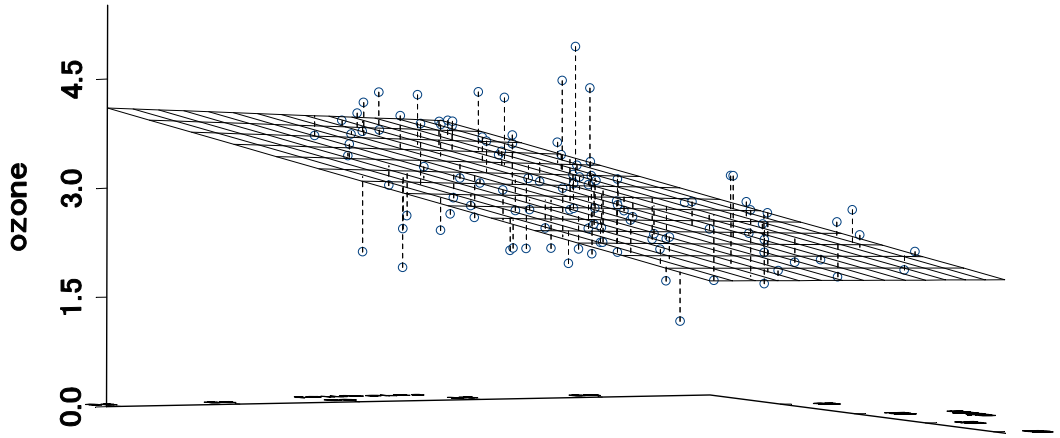
3.16いろいろな 3 次元グラフ

回帰平面グラフ

- `Exenvirn` の列、`radiatio, temperat, ozone` を CTRL キーを押しながら、選択
- 3D プロットパレットから、**垂直線付き散布図ボタン** をクリック

グラフは各点から、z=0 の平面までの垂直線が引かれたグラフになります。**プロットオブジェクト**（点や線をクリック）を選択し、右ボタンクリック、**Regression** メニューを選択します。ダイアログボックスの **Draw Plane** チェックボックスにチェックします。回帰平面が引かれ、垂直線は各点から、平面への垂線、いいかえると残差を表します。

3-24 各点から、回帰平面へ線が引かれた



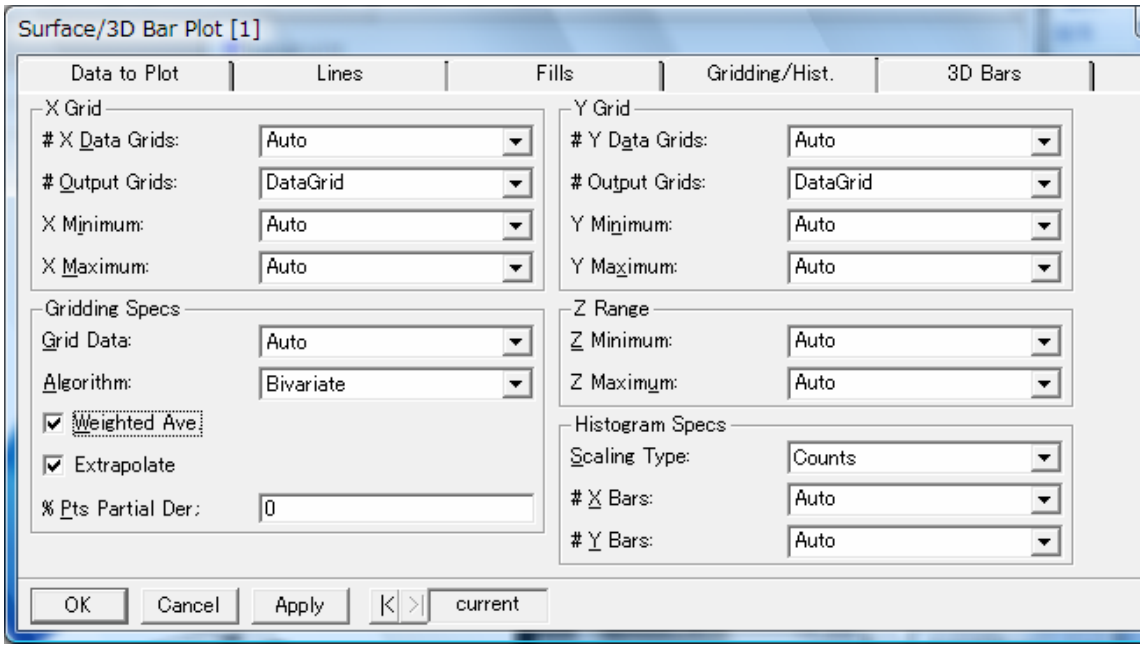
鳥瞰図

点を補間してなめらかな曲面からなる鳥瞰図にすることができます。作図の操作方法は同じです。3 列を選んで、**鳥瞰図** のボタンをクリックします。

ここで注意しなければいけないことは、これらのグラフではただ単純に点と点を結ぶ線を引くのではなく、点を覆うようななめらかな面を描画していることです。そのために、3 列の選択では、データの補間をしているため、実際には点がない場所に面が描画されることがあります。また、特に、端は外生補間という、このまま平面を伸ばしていくとどのように予測されるかを計算することにより、描画されますので、実際には点がない場所の高

さ方向の座標が思ってもみない値になることがありますので、注意してください。外生補間などのプロパティは**プロットオブジェクト**（面や線、点などをクリック）を選択し、右ボタンクリック、メニューから **Gridding/Hist.** を選択します。

3-25 外生補間を変更する Gridding/Hist. ダイアログ



ダイアログの各フィールド

# X Data Grids, #Output Grids	補間に用いる座標の数、細かいほどなめらかになる。データによっては指定数に応じて、描画が遅くなる
X Minimum, X Maximum	X 座標の最小、最大
Grid Data	データは 3 列の X,Y,Z 型データか、m 行 n 列で各値が Z の値を表すグリッド型かの指定、デフォルトは自動判定
Algorithm	補間計算のアルゴリズム
Weighted Ave.	重みづけ平均を計算するかどうか
Extrapolate	外生補間をするかどうか
% Pts. Partial Der	一部微分計算をするかどうか、ゼロはしない、数値を指定すると、%とみなす。大きい数を指定するとより滑らかになる

3.17 3 次元グラフと 2 次元グラフを重ね合わせる

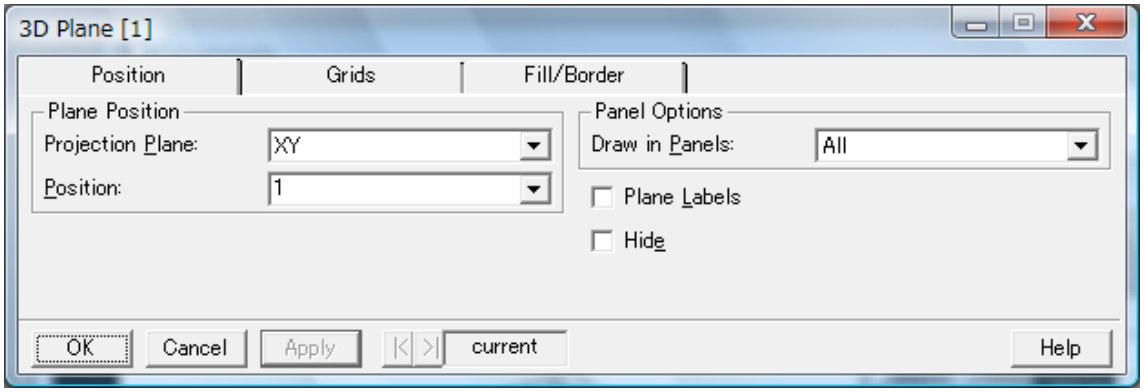
3 次元グラフのある面に、2 次元グラフを重ね合わせることができます。

- 先ほどの、`radiatio, temperat, ozone` の 3 次元グラフをアクティブにする
- **グラフエリア**（3 次元グラフの軸の内側をクリック）を選択
- `Exenvrin` データウィンドウの `radiatio, temperat` を、CTRL キーを押しながら選択
- 2D プロットパレットの **線形回帰ボタン** をクリック

3 次元グラフの下部 XY 平面に、回帰直線のある 2 次元グラフが表示されます。

3D プロットパレットの **平面ボタン** を、3D グラフにドラッグアンドドロップすることにより、いろいろな面にグラフを重ね合わせることができます。グラフを描く平面は複数でもかまいません。また、できた 2 次元グラフをクリック、右ボタンメニュー **Position** を選択し、フィールド **Position** に座標を指定することにより、任意の場所に面を移動させることができます。**Position** を変えることにより、複数のグラフをレイヤーにして表示させることも可能です。

3-26 2D グラフの表示場所を座標で変更するダイアログ



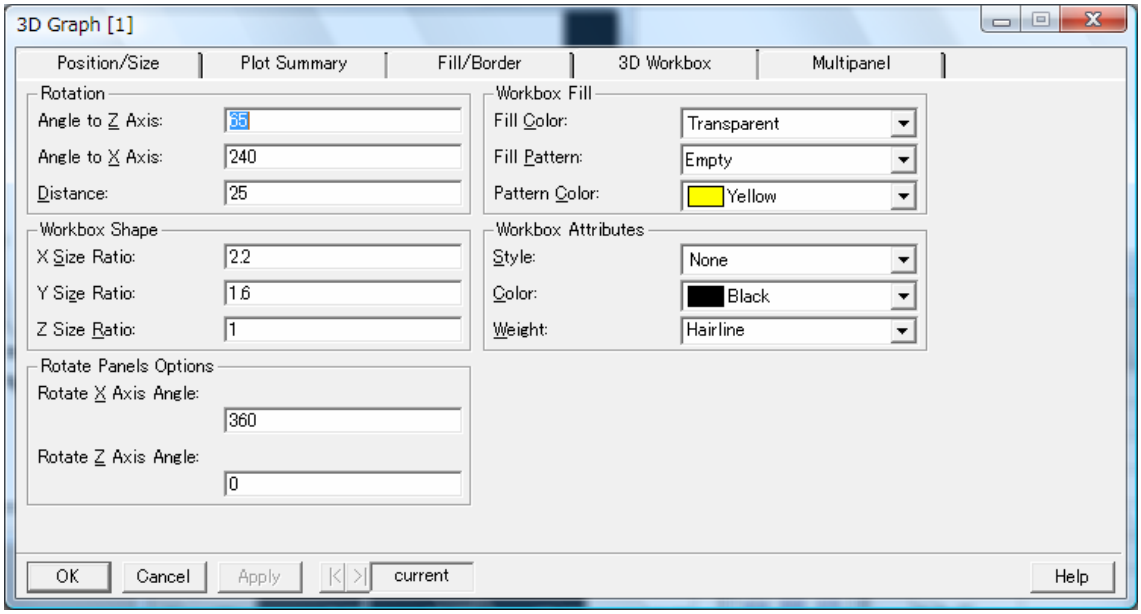
3.18 3 次元グラフの編集

3 次元グラフに関する編集は、2 次元グラフ同様です（点の形や色を変える、軸の範囲を変える、グラフをパネルに分けるなど）。いくつか、3 次元グラフ固有の内容があります。

3D Workbox

3 次元グラフの **グラフエリア**（軸の内側の点以外をクリックします）を選択、右ボタンクリックし、メニューから **3D Workbox** を選択します。

3-27 3D Workbox ダイアログ



ダイアログの各フィールド

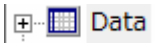
Angle to Z/X Axis	角度の指定、実際に回転させてどのような値になるか確認が可能
Distance	視点からの距離、グラフの大きさは変化しません
X,Y,Z Size Ratio,	X,Y,Z の比を指定、1,1,1 とすれば、座標が立方体になる
Rotate X Axis Angle	X 軸の回転可能な角度、たとえば、180 とすれば、**Angle to X Axis** で指定された角度から 180 度回転までを 4 分割し、45 度回転させて表示することが可能
Workbox Fill グループ	3 次元プロットエリアに色をつけたり、パターンを表示させることが可能
Workbox Attributes グループ	3 次元プロットエリアの周りに線を引く、補助線が引かれ、3 次元グラフがより見やすく

	なることがある

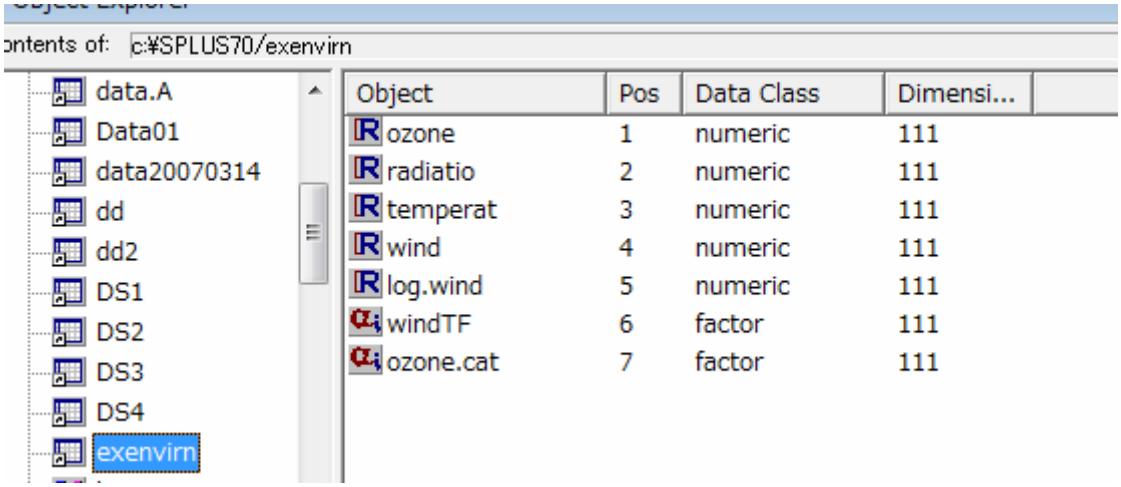
3.19 オブジェクト・エクスプローラとグラフ

データの章で、**オブジェクト・エクスプローラ**について説明しましたが、このオブジェクト・エクスプローラをグラフ作成や編集に生かすことができます。

ツールバーの **オブジェクト・エクスプローラ** **ボタン** をクリックし、オブジェクト・エクスプローラを開きます。

左側のパネルから、Data の左に表示されるプラスの  **展開ボタン** をクリックすると、**Data** の下に、データが表示されます。そのうち `exenvirn` をクリックします。

3-28 オブジェクト・エクスプローラからデータを選択し、グラフを作成

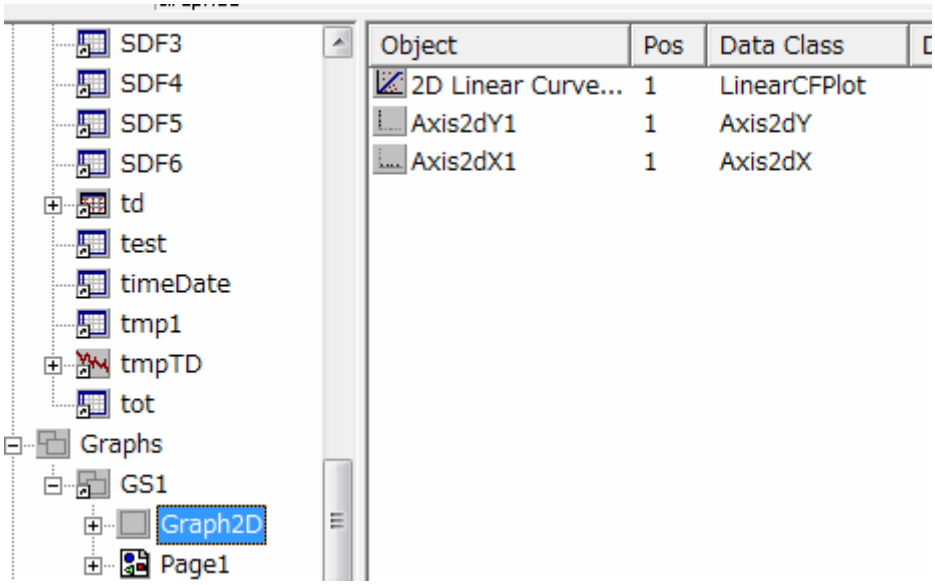


`Exenvirn` に含まれる列が右側のパネルに一覧表示されます。データ列の選択をオブジェクト・エクスプローラ上で行うことができます。データによっては列が多く、離れた X 列と Y 列を同時に選択するのが面倒な場合があります。その場合に有効です。たとえば、`radiatio, ozone` と順に列を右パネルから CTRL キーを押しながら選択、2D プロットパレットから **線形回帰** ボタンをクリックします。

続いて、オブジェクト・エクスプローラの左パネルから、**Graph アイコン** の隣にあるプラスマークの **展開ボタン** をクリックします。今作った線形回帰グラフの名前

があることを確認します。そのグラフ名の横にも、プラスマークの展開ボタンがあるので、クリックします。すると、左側に **Page1** と、**Graph2D** が現れますので、**Graph2D** を選択します。

3-29 オブジェクト・エクスプローラからグラフを選択したところ



Graph2D を選択すると、右側には、いくつかのアイコンが現れます。上記の画面では、**線形回帰プロット** と、**Y 軸**, **X 軸** の 3 つのオブジェクトがあります。**2D Linear Curve** アイコンをダブルクリックすると、回帰直線プロットのダイアログが開きます。このダイアログを用いて、グラフの変更を行います。

回帰直線を 2 次にするには

- **Curve Fitting** タブをクリック、**Poly. Order** を 1 から 2 に変更

信頼区間を線で表示するには

- **By Conf Bound** タブをクリック、**Level** に 0.95 と入力、**Style** リストボックスから線を選択

グラフが多数重なっていたり、多くのオブジェクトの組み合わせで描画されている場合など、編集したいオブジェクトの選択が容易になります。

オブジェクト・エクスプローラで表示されるグラフについて

S-PLUS のウィンドウ上に開かれているグラフのみ、オブジェクト・エクスプローラで選択が可能です。保存はされているが、開かれていないグラフは表示されませんので注意してください。この点がオブジェクト・エクスプローラにおけるデータの表示と異なります。データは開いていてもいなくても、オブジェクト・エクスプローラに表示され、選択が可能です。

3.20 グラフへデータを埋め込む

39 ページで、**S-PLUS グラフ形式（*.sgr）**で保存した場合は、グラフを表示するためには、S-PLUS のデータベース上にそのグラフの元となるデータがないと表示できないと説明しました。たとえば、前の節で作った、線形回帰グラフを表示させるには、`exenvirn` が必要です。データがなくてもグラフを表示させるためには、グラフにデータを埋め込むという方法があります。

- グラフを選択
- メニュー　**グラフ / データのグラフへの埋め込み** を選択

データが埋め込まれたグラフは元のデータがなくても表示可能です。また、グラフからデータを抽出することも可能です。

- グラフを選択
- メニュー **グラフ / データのグラフからの抽出** を選択

データ量が多いと、埋め込みをしたグラフのサイズが大きくなるので注意してください。また、埋め込み可能なグラフはデータ描画に必要なデータのみです。X,Y いずれにも使われなかったデータは埋め込まれません。

3.21 メニューからグラフを選択する

よく使われるグラフはグラフパレットにボタンアイコンとして保存されています。メニューから選択しないと作成できないグラフがありますので、注意してください。

- 極座標グラフ
- スミス・チャート

などがその例です。

3.22 ブラッシングとスピン

対話型 3 次元グラフを作成します。メニュー **グラフ / ブラッシングとスピン** を選択し、表示させたいデータを選択します。対話型ウィンドウ上で、3 次元グラフの回転をさせたり、列選択を変更したり、点の認識が可能です。

4 統計

一般的な解析手法はメニューから選択して実行が可能です。

メニュー 統計の説明

メニュー	主な解析
データサマリー	平均、標準偏差、最小値、最大値などの計算 相関係数行列、分散共分散行列 クロス表のカイ二乗検定
標本比較	パラメトリックな検定 ノンパラメトリックな検定 カイ二乗適合度検定
検出力と標本数	有意な結果を得る検定のための例数設計
計画	実験計画の設計
回帰	線形、ロジスティック、非線形など
分散分析	分散分析、多重比較
混合効果モデル	線形、非線形混合効果モデル
一般化最小二乗モデル	残差に構造をもつ回帰
生存時間解析	カプラン=マイヤー、コックスの比例ハザードなど
樹形モデル	回帰樹、分類樹
クラスター分析	デンドログラム、k-means 法など
多変量解析	主成分分析、因子分析、判別分析
QC チャート	xbar などの QC チャート作成
標本からのサンプリング	ブートストラップ、ジャックナイフ
平滑化	局所回帰、スプラインなどによる平滑化曲線付きグラフ作成
時系列解析	ACF プロット,AR, ARIMA

S-PLUS には、これ以外にも多くの統計解析手法があります。メニューに現れないものについては、コマンドでの実行が可能です。

4.1 統計メニューでの操作

ほとんどの統計解析メニューに共通のダイアログ指定フィールドには、次のようなものがあります。

ダイアログの共通フィールド

Data Set	解析したいデータ名
Save As	統計結果を保存する場合、名前を指定

共通の操作について

フィールドの説明欄に「複数選択可能」と書かれている場合、特に記述がなければ、CTRL キーと変数（列）名を選択することにより、複数の変数（列）の選択が可能です。

4.2 基本統計量の計算

`exenvirn` を用いて、各変数（列）の平均などを求めます。

- メニュー　**統計 / データサマリー / 統計量** を選択

Data タブ

Variables	統計量を計算させたい変数（列）を選択、複数選択可能
Group Variables	男女ごとに統計量を計算させたい場合などに、グループとなる列を指定、複数選択可能

Statistics タブ

Mean グループ	平均、平均値の標準誤差、平均値の信頼区間
Quantiles グループ	四分位範囲（最小値、中央値、最大値と四分の 1,3 位点）
Scale グループ	分散、標準偏差
Shape グループ	尖度、歪度

Other Statistics グループ	行数、欠損数、合計

exenvirn の例では、Variables が<ALL>となっていることを確認したら、OK をクリックします。結果は Report1 と表示されるレポートウィンドウというウィンドウに書き出されます。

レポートウィンドウについて

統計解析の結果が書き出されるウィンドウでは、グラフウィンドウのグラフ同様、メニュー **ファイル / 名前** をつけて保存 としなければ保存されません。保存形式には

- Report File 形式（*.srp）

S-PLUS 固有のファイル形式で、S-PLUS 以外で読むことはできません

- Rtf(Rich Text File)形式（*.rtf）

Microsoft Word や Wordpad から開くことができます。グラフをコピーペーストすることも可能です。

- Text File 形式(*.txt)

単純な ASCII テキストファイル形式で保存されます。貼り付けたグラフの保存はできません。

があります。**Save As** フィールドに名前を指定するとその名前の S-PLUS オブジェクトとして保存されます。名前をつけたオブジェクトはデータベースに自動保存され、上記のようなファイル形式で保存されるものとは異なります。**Save As** を利用して保存したデータについては 104 ページのオブジェクトエクスプローラによるデータ管理ページも参照してください。

ここで、ozone の平均値を確認します。平均が 3.24, 中央値が 3.14 となっています。
この結果を利用して、続いて検定を行います。
ozone の平均が 3.2 あるいは 4 という仮説を検討してみましょう。

4.3 検定

平均値のパラメトリックな検定

- 1 標本 t 検定

メニュー **統計 / 標本比較 / 1 標本 / t 検定** を選択

t 検定ダイアログフィールドの説明

Variable	検定を行いたい変数（列）を選択
Mean Under Null Hyphothesis	帰無仮説に用いたい値を指定、たとえば、3.2 と入力すれば、帰無仮説は「ozone の平均値が 3.2 に等しい」という仮説を検定する。指定がなければ「平均ゼロ」が帰無仮説となる
Alternative Hyphothesis	検定として、両側検定、片側検定どちらを行うかの指定
Confidence Level	信頼区間の幅

exenvirn の例では、

- **Variable** から ozone を選択
- **Mean Under Null Hyphothesis** に、3.2 を指定した場合と、4 を指定した場合を試します

いずれも結果はレポートウィンドウに書き出され、表示される項目は同じです。それぞれの統計量が平均=3.2 と仮定した場合と、平均=4.0 と仮定した場合とで計算されます。

```
        One-sample t-Test

data:  ozone in exenvirn
t = 0.5654, df = 110, p-value = 0.573
alternative hypothesis:  mean is not equal to 3.2

        One-sample t-Test

data:  ozone in exenvirn
```

```
t = -8.9006, df = 110, p-value = 0
alternative hypothesis:  mean is not equal to 4
```

結果には **p-value(P 値)** が表示され、平均=3.2 と仮定した場合は 0.573 と計算され、平均=4 とした場合は 0 になります。これは帰無仮説の取る両側確率で、P 値がユーザーの意図した有意水準（この確率より小さい確率しかとらなければ、滅多に起こらないことなので起こらないと結論付けてもよい）値より小さければ、帰無仮説の棄却（間違っているといえる）ができます。有意水準 0.05 で、平均=3.2 は間違っていると言えませんが、平均=4 は間違っているということができます。

t 検定はこのように、1 つの変数（列）がある値と等しいという検定を行うことも(1 変量 t 検定)、ある変数とある変数の平均が等しい、あるいは 2 つのグループの平均が等しいという検定を行うことも(2 変量 t 検定)できます。

では、`exenvirn` の `ozone` が、`windTF` の値によって変わるかどうか、つまり `windTF` が「いいえ」「はい」でグループ分けして、それぞれの平均値が等しいかどうかを検定します。

2 標本 t 検定

メニュー**統計 / 標本比較 / 2 標本 / t 検定** を選択

2 変量 t 検定のダイアログフィールドの説明

Variable1	比較対象変数 1 を指定
Variable2	比較対象変数 2、あるいはグループ変数を指定
Variable2 is a Grouping Variable	**Variable2** がグループ変数の場合にはチェックをする
Test グループ	**Paired t:** 対のある（データが行方向に 1 対 1 対応している）データの場合はチェック **Two-sample t:** 単純な変数の平均値比較の場合 **Assume Equal Variances:** 2 つのグループの分散が等しいと仮定できる場合はチェックされないと Welch の方法が

	とられる

- **Varaible1** から ozone を選択
- **Variable2** から windTF を選択
- **Variables2 is a Grouping Vaiable** にチェックする

```
        Pooled-Variance Two-Sample t-Test

data:  x: ozone with windTF = いいえ , and y: ozone with windTF = はい
t = 5.4741, df = 109, p-value = 0
alternative hypothesis:  difference in means is not equal to 0
95 percent confidence interval:
 0.5256905 1.1223982
sample estimates:
 mean of x mean of y
  3.641246  2.817202
```

結果から、”いいえ”と”はい”の 2 つのグループで、ozone の平均値が等しいという帰無仮説は棄却されます（2 つのグループ内平均値は異なる）。

ノンパラメトリック検定

今試した t 検定は、データが正規分布に従うことを仮定する、2 つのグループ内分散が等しいことを仮定する必要があります。一般的に、このような仮定は満たされないことが多いのですが、その場合はノンパラメトリックな検定が有効です。解析の手順、表示される結果はパラメトリック検定と同じです。1 標本、2 標本ともに、ウィルコクソン順位検定を選んでください。

3 つ以上のグループでの検定

- 分散分析
- クラスカル=ワリス順位検定
- フリードマン順位検定

が可能です。あてはめることができるデータは 1 列に検定したい値、1 列にグループに分けたいデータがある場合です。グループごとに 1 列持つようなデータは検定することができません。その場合は、メニュー **データ / 構築 / スタック** でデータを変換してください。

正規性の検定

同じ ozone データに対して、ozone が正規分布しているかどうかを検定します。
適合度検定を行います。

- メニュー **統計 / 標本比較 / 1 標本 / カイ二乗適合度検定** を選択
- **Variable** から ozone を選択
- **Probability** グループの **Mean** を 3、**Stdev** を 1 と指定

帰無仮説は「データは平均 3, 標準偏差 1 の正規分布をしている」です。結果の P 値を見ると 0.122 なので、この仮説は棄却できません。2D パレットから、**ヒストグラム**ボタンをクリックすることにより、実際のデータの分布の様子がよくわかります。正規分布のパラメータ（平均 **Mean** と、標準偏差 **Stdev**）をいろいろに変えて、P 値を計算してみましょう。

また、データの分布形を変えると、指定可能なパラメータはその分布関数固有のものに変わります。

2 標本で適合度検定を行うと、2 つのデータが同じ分布をしているかどうかの検定を行うことができます。

カイ二乗検定

今までは数値データに対して、その平均値がある値に等しいか、あるいは、正規分布をしているかを検定しました。因子変数（列）が 2 つあったときに、その 2 つに関連があるかどうかをカイ二乗検定します。

exenvirn の例では、2 つの後から作った因子データ windTF と、ozone.cat に関連があるかどうかを試します。

- メニュー **統計 / 標本比較 / 仮説検定 / カイ二乗検定** を選択
- **Variable1** から windTF を選択
- **Variable2** から ozone.cat を選択

帰無仮説は「windTF と ozone.cat は互いに独立」です。つまり、windTF が”はい”で

も、”いいえ”でも、`ozone.cat` の値は変わらないということです。
実行すると、ワーニング（警告）が表示されます。

Expected counts < 5. Chi-square approximation may not be appropriate. in: chisq.test(x, y, correct)

1 つのグループに含まれるデータ件数が 5 個以下になるので、カイ二乗分布は適当ではないかもしれない、ということを警告しています。このワーニングを無視すれば、P 値はゼロなので、帰無仮説は棄却することができます。

ダイアログの **Data Set is a Contingency Table** は、データが既にクロス集計された結果かどうかで指定が変わります。今回のように、クロス集計する前のデータであれば、チェックしません。既に Excel などで集約済みのデータだった場合は、チェックします。

4.4 実験計画

要因計画のデザインをしたり、得られた結果のグラフ化などを行います。

直交要因計画

総当たりでない、直交計画表を用いた実験のデザインをします。

- メニュー　**統計 / 計画 / 直交要因計画** を選択
- **Levels** に 3,3,3 と入力

（3 水準 3 因子の直交計画をデザインします。L18 直交表が参照されます）

新しいデータセットができ、3 水準因子 A,B,C とその水準 1,2,3 の組み合わせが表示されます。
空の列にその組み合わせの実験を行った結果を入力することにより、以下のメニューからグラフ作成、分散分析を行うことが可能です。もちろん、S-PLUS で実験を計画した結果でなくても、グラフ作成、分析は可能です。

実験計画データのグラフ化

グラフ作成時に共通なフィールドがあります。

ダイアログの共通フィールド

Dependent	測定値など、実験結果を指定
Independent	実験の因子を指定、複数の変数を指定可能、交互作用プロットでは、2 変数指定しないとエラーになる

要約統計量プロット

- メニュー **統計 / 計画 / 各因子ごとの要約統計量プロット** を選択
- **Dependent** から `radiatio` を選択
- **Independent** から `windTF, ozone.cat` を、CTRL キーを押しながら、選択
- **Function** には、グラフ化したい統計量計算関数（`mean, median` など）を指定

各水準の平均値（**Function** の指定により、計算されます）などがグラフ化されます。水準名のない、他よりも長い横棒は全体の平均を表します。

各因子ごとの箱ひげ図

59 ページで作成したグラフと同じです。複数の因子変数を指定すると、変数の数だけグラフのページが増えます。

交互作用のチェック

Independent として、2 つの因子変数を選びます。因子間に交互作用があれば、線がクロスします。

4.5 回帰

S-PLUS では、回帰分析として多くの種類があります。

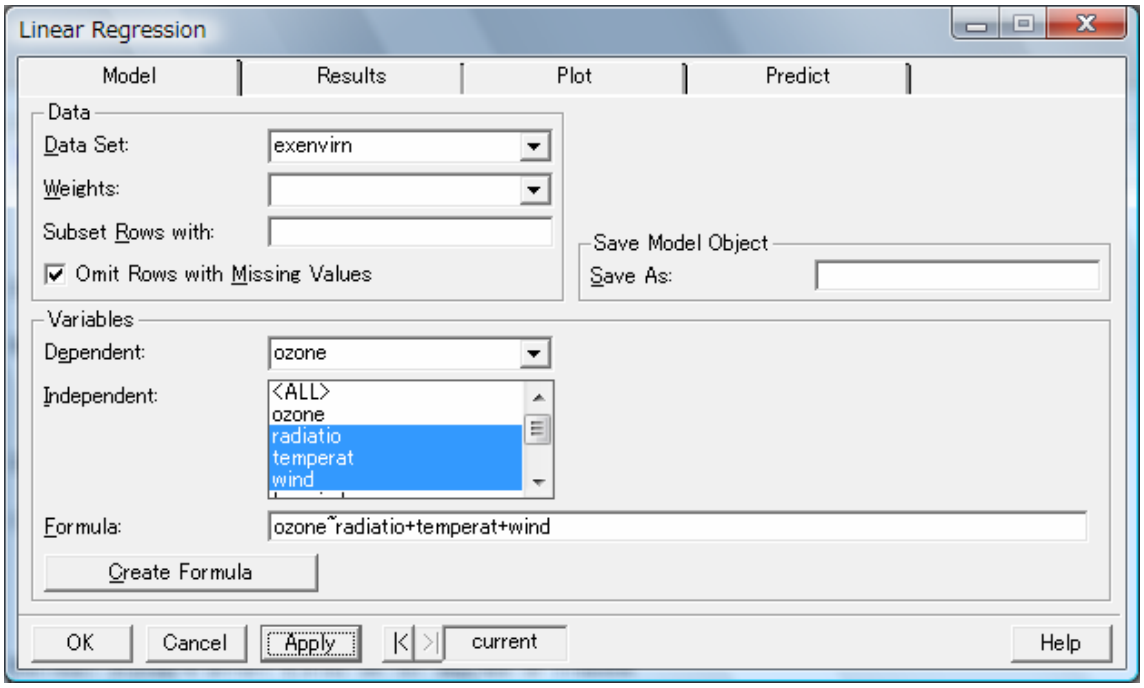
回帰に共通のダイアログ設定

Weights	重み、人数や割合の列があれば、指定可能

Subset Rows with	条件式を指定すれば、特定の条件に従うデータだけで回帰を行う
Omit Rows with Missing Values	欠損値があるデータを除いて回帰
Dependent	目的変数、通常は数値列を選択、ロジスティック回帰の場合は、2 水準因子列の指定をする
Independent	説明変数、1 個だけなら単回帰、2 個以上なら重回帰分析になる。因子変数（列）の指定も可能。複数列は CTRL キーを押しながら選択
Formula	**Dependent, Independent** の指定により、自動的に表示される。直接入力も可能
Create Formula	式ビルダー、交互作用などを追加する場合はクリックして、ダイアログから入力することができる

- メニュー **統計 / 回帰 / 線形** を選択
- **Data Set** に `exenvirn`、**Dependent** に `ozone`、**Independent** に `radiatio, temperat, wind` を CTRL キーを押しながら指定

4-1 線形回帰のダイアログ



自動的に、**Formula** フィールドに

```
ozone ~ radiation + temperat + wind
```

と表示されます。

この状態で、**Apply** をクリックすると、**レポートウィンドウ**に結果が出力されます。

```
        *** Linear Model ***

Call: lm(formula = ozone ~ radiatio + temperat + wind, data = exenvirn, na.action
=
        na.exclude)
Residuals:
    Min      1Q   Median    3Q   Max
 -1.122 -0.3764 -0.02535 0.3361 1.495

Coefficients:
            Value Std. Error t value Pr(>|t|)
(Intercept) -0.2973  0.5552    -0.5355  0.5934
   radiatio  0.0022  0.0006     3.9493  0.0001
   temperat  0.0500  0.0061     8.1957  0.0000
       wind -0.0760  0.0158    -4.8253  0.0000

Residual standard error: 0.5102 on 107 degrees of freedom
Multiple R-Squared: 0.6807
F-statistic: 76.03 on 3 and 107 degrees of freedom, the p-value is 0
```

特によく参照される結果は次のとおりです。

- 各説明変数の係数は有意です（ゼロという帰無仮説は棄却することができます)。
- R-Squared(R 二乗値、回帰式の変動の二乗和/データ変動の二乗和)は 0.6807 です。

オプションを指定した回帰

Subset

条件に合ったデータだけで回帰をしますが、指定方法は次のとおりです。

- 数値を指定

`1:50`

1 行目から 50 行目のデータだけを用いて回帰をする

`c(1,3,5,7,9,11,13,15,17,19)`

指定された行のデータだけを用いて回帰をする

- 条件を指定

`radiatio>200`

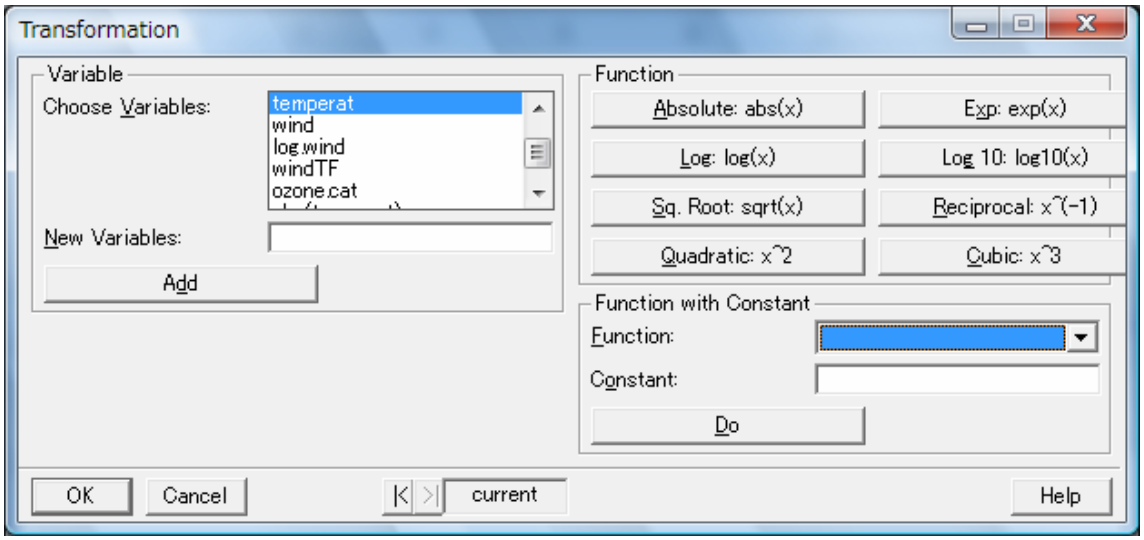
変数（列）`radiatio` が 200 より大きいデータだけで回帰をする

Create Formula ボタン

説明変数や目的変数を対数変換したり、交互作用を指定するときに用いると便利です。

- 対数変換などをする場合は、先に **Transform** ボタンをクリック

4-2 Transform ダイアログ



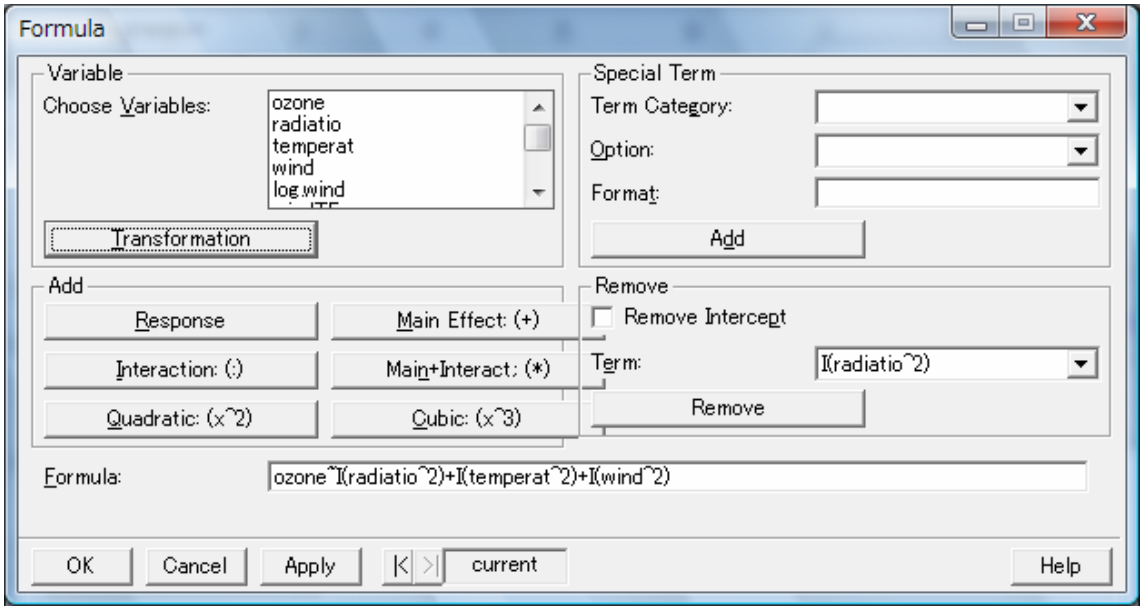
Transformation ダイアログでは、

- **Choose Variables** から変数を選択

- 右側の **Function** グループから、変換式を選択、**Add** ボタンをクリック
- **OK** をクリック

これで、**Formula** ダイアログに戻ります。

4-3 Fomula ダイアログの Fomula フィールドに指定したモデル式が表示される



- **Choose Variables** から変数を選択（**Transformation** で作成した変換式も対象になっています)、1 つ以上、複数個選択も可能です
- **Add** ボタンの 1 つをクリック

ボタンの意味は次のとおりです。

Response	目的変数
Main Effect(+)	説明変数（主効果）
Interaction	2 つ以上の説明変数の交互作用
Main + Interact(*)	2 つ以上の説明変数の主効果と交互作用
Quadratic	説明変数の 2 次式
Cubic	説明変数の 3 次式

特殊な設定は **Special Term** グループで指定します。

切片のない回帰分析（必ず原点を通る）を行う場合は、**Remove Intercept** にチェックをします。

結果を保存するには、**Save As** フィールドに適当な名前を指定します。後で再利用が可能です。

その他のタブ

- **Results** タブ

結果に何を表示させるかを指定します。標準は **Long Output** が表示されますが、**Short Output** の場合は、係数値など、一部の結果だけが表示されます。**ANOVA** にチェックをすると、分散分析表が表示され、**Correlation Matrix of Estimates** で、説明変数の予測値の相関係数が計算されます。**Save Results** グループでは、回帰結果の予測値と残差を計算させ、名前を付けて保存することができます。

- **Plots** タブ

回帰診断プロットを選ぶことができます。

診断プロットの種類

Residuals vs Fits	X 軸：当てはめ値、Y 軸：残差
Sqrt Abs Residuals vs Fits	X 軸：当てはめ値、Y 軸：残差絶対値平方根
Response vs Fit	X 軸：当てはめ値、Y 軸：観測値
Residuals Normal QQ	残差正規 QQ プロット
Residuals-Fit Speard	残差の分布と当てはめ値の分布の比較プロット
Cook's Distance	クックの距離
Partial Residuals	偏残差プロット

それぞれ、当てはめの良しあしや、残差の分散均一性、正規性、外れ値の有無などを診断します。

Include Smooth: 散布図に平滑化曲線を追加します

Include Rugplot: 散布図に X 軸上の点の密度を表示させます
Number of Extreme Points to Identify: 外れ値として認識させたい点の数を指定します

Include Smooth と、**Include Rugplot** は偏残差プロットにも追加可能です。

- Predict タブ

予測値の計算に用います。

New Data: 予測値を計算したい新しいデータの名前を指定します。指定がなければ、回帰の計算に使われたデータに対して予測値を計算させます。指定可能な新しいデータには、回帰で使った説明変数と同じ名前の列が含まれていることを確認してください。
Predictions: 予測値を計算するかどうか
Confidence Intervals: 予測値の信頼区間、パーセントは **Confidence Level** フィールドで指定します
Standard Errors: 標準誤差を計算するかどうか

4.6 その他の回帰分析

その他の回帰でも、ダイアログの指定は同様です。

線形回帰以外の回帰分析

ロジスティック、プロビット（一般化線形モデルからの実行も可能です）	目的変数が「はい」「いいえ」のように 2 水準因子の場合
ポアソン（一般化線形モデルからの実行も可能です）	目的変数が離散型数値の場合
一般化線形モデル	誤差項の確率密度関数の指定が可能
一般化加法モデル	説明変数項に平滑化関数などのあてはめが可能、また誤差項の確率密度関数の指定も可能
ロバスト	外れ値にあまり影響されない線形回帰が可能
非線形	説明変数間に非線形な関係があり、逆行列計算などではパラメータが求まらない場合
ステップワイズ法	説明変数の変数選択を行う

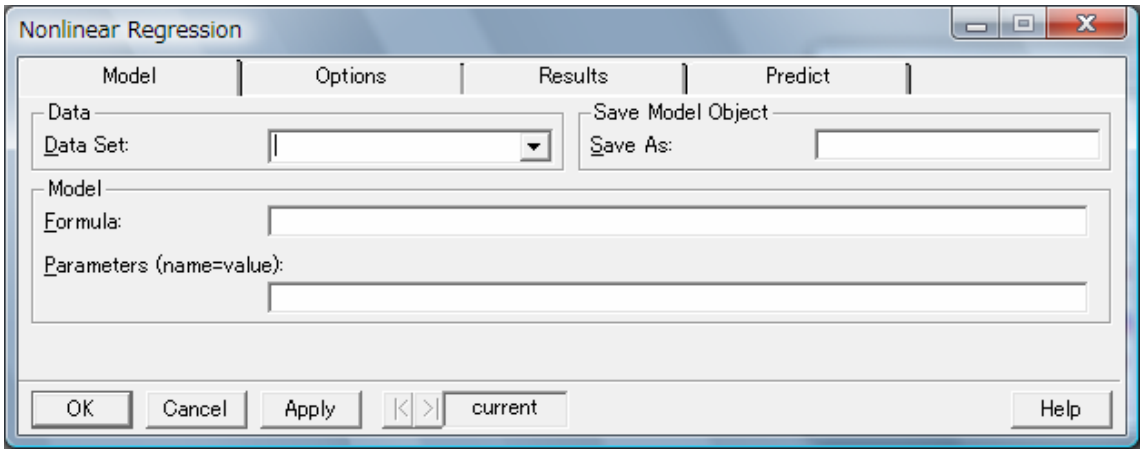
局所	局所重み付き回帰を行う。ノンパラメトリック回帰なので、平滑化で用いる

それぞれのメニューを選び、基本的には線形回帰で説明した指定を行うと、実行が可能です。メニュー固有のパラメータがいくつか存在することがありますが、詳しくはダイアログのヘルプボタンをクリックし、オンラインヘルプを参照してください。

非線形回帰とステップワイズ法には、固有のダイアログ指定がありますので、簡単に説明します。

非線形回帰

4-4 非線形回帰のダイアログ



Formula フィールドには、説明変数名とパラメータ名（求めたい係数）からなる非線形の式を入力します。たとえば、目的変数の列 y を説明変数の列 x と、パラメータ a, b, c からなる非線形の式で表すならば、

```
y ~ a*(1+exp(x-b)/c)
```

などのように入力します。

Parametres(name=value)フィールドには、**Formula** で指定したどの文字が求めたいパ

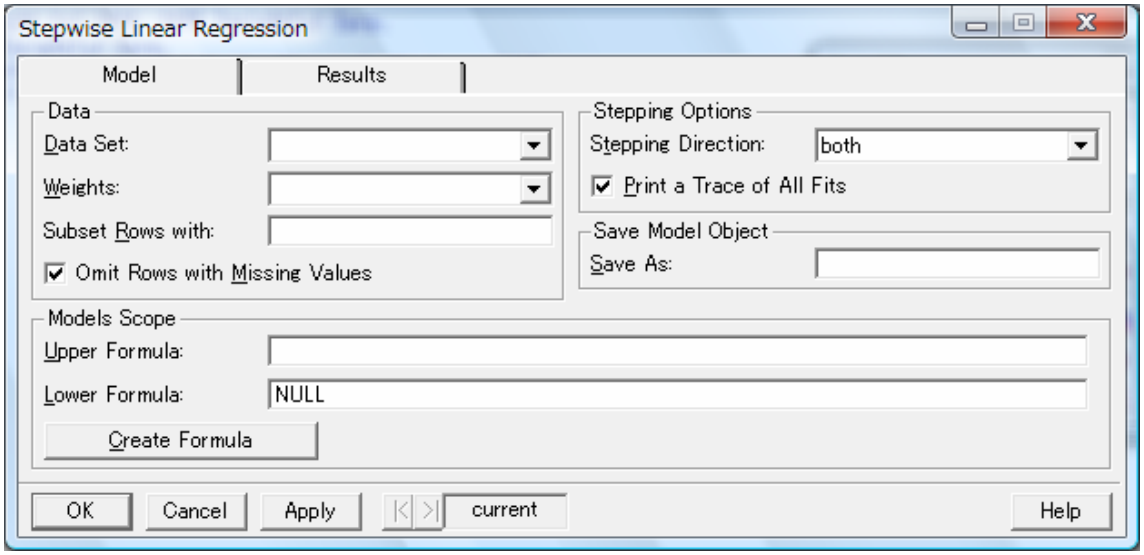
ラメータか、また、その初期値を指定します。上の例であれば、

```
a=1, b=1,c=1
```

と、パラメータ名=初期値（数値）をカンマで区切って入力します。

ステップワイズ法

4-5 ステップワイズ法のダイアログ



AIC(実際は Mallow の Cp)を用いて変数選択を行います。基本的なダイアログ指定は線形回帰に同じですが、モデル式は **Upper Formula, Lower Formula** の 2 式を指定します。

Upper Formula： 最大限度になるモデル式を指定（2 次の交互作用などを含めた式ももちろん可能です)。指定は必須です。
Lower Formula： 最小限取り入れいたい説明変数の指定、指定がなければ、**NULL**(回帰をしない)というモデルが用いられます。

4.7 その他の統計モデル

ダイアログへの指定は、概ね回帰と同じです。

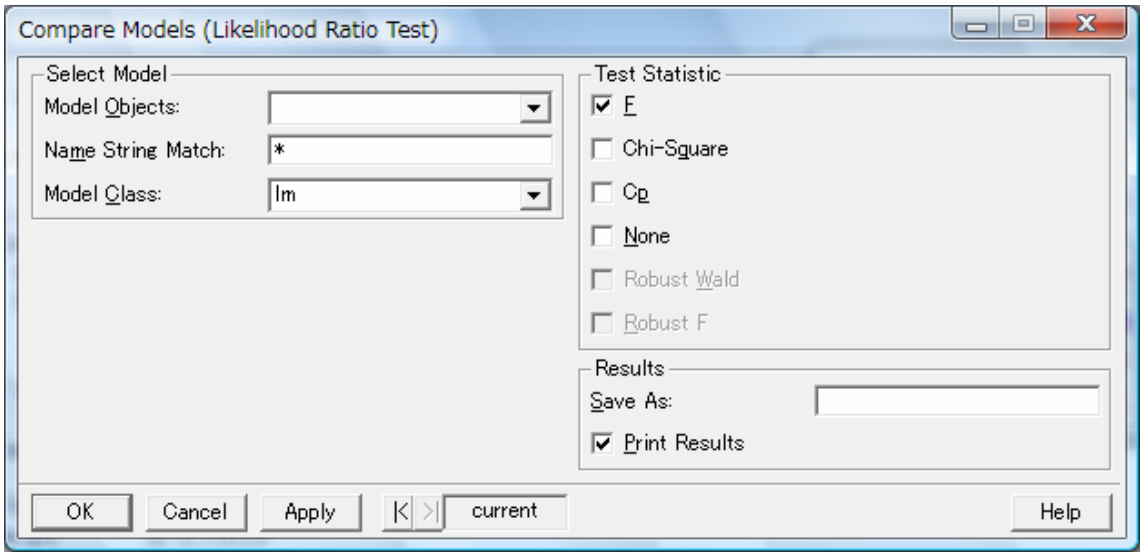
目的変数がないもの：クラスター分析、主成分分析
目的変数に因子を指定することができるもの：樹形モデル、判別分析

また、メニュー固有の指定については、ダイアログのヘルプボタンをクリックし、オンラインヘルプを参照してください。

ダイアログの指定が回帰と大きく異なるメニュー

モデルの比較

4-6 モデルの比較を選んだダイアログ



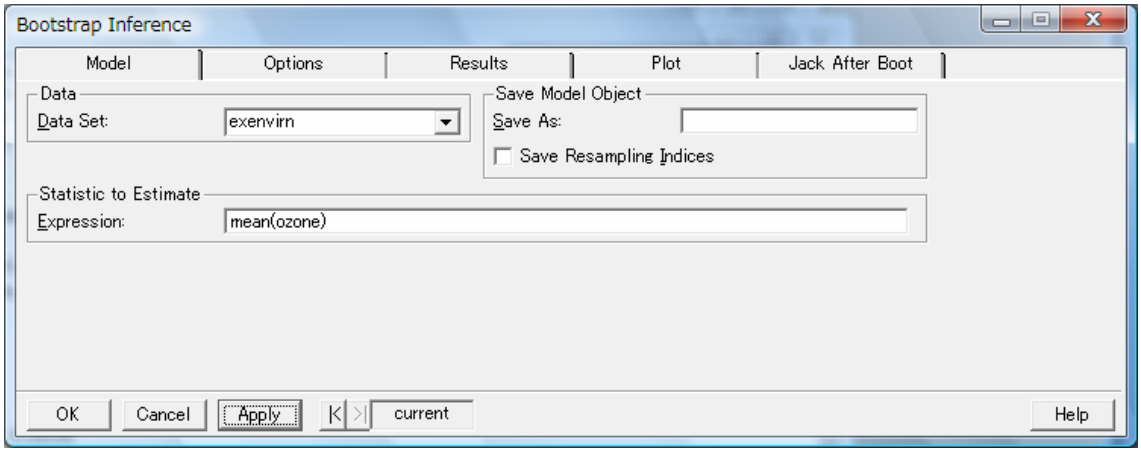
Model Objects：比較対象となるモデルを 1 個以上選択します。
1 個選択した場合は、そのモデル内での分散分析表を出力します（説明変数間の比較になります）
2 個以上選択した場合は、複数モデルでの分散分析表を出力しますので、モデル間での比較が可能です。ただし、モデル間で目的変数が同じになるよう、注意してください。また、**Model Class** が異なるモデル間での比較はできません。

標本からのサンプリング

ブートストラップでも、ジャックナイフでも、指定や出力される結果の種類などは同様

です。

4-7 標本からのサンプリングでブートストラップを選んだダイアログ

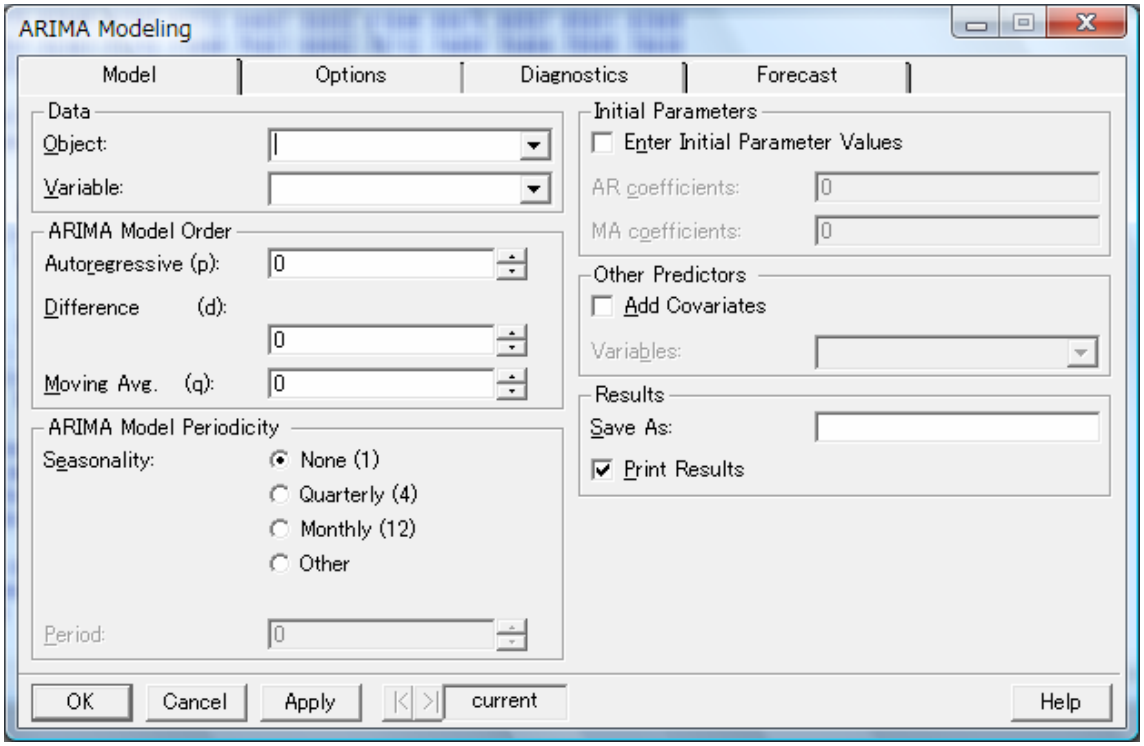


Expression: データからのサンプリングによる、ブートストラップ法により求めたい値を指定します。列 `ozone` の平均値をブートストラップで求めたい場合は、**Expression** に `mean(ozone)` と指定します。指定は S 言語の式で行います。たとえば、`exenvirn` データの `ozone` を目的変数に、`radiatio` を説明変数とした回帰係数を求めるには、`coef(lm(ozone~radiatio))` と指定します。求めたいパラメータについて、復元抽出をしたサンプルから計算するブートストラップ法により求めた値、あるいはその値の統計量、ヒストグラムなどがデフォルトで出力されます。

時系列

ARIMA モデルのダイアログを説明します。

4-8 時系列 / ARIMA を選んだダイアログ



Object:でデータ名、**Variable:**で列名を選択します。
ARIMA Model Order グループでは、ARIMA(p,d,q)の次数を順に指定します。**ARIMA Model Periodicity** では、時系列の何期前を参照するかを指定します。特に指定がなければ、**None(1)** で、単純に 1 期前を参照します。たとえば、ARIMA(1,0,1)などでは、直前のデータに対して、自己回帰係数を求めます。

4.8 オブジェクトエクスプローラと統計解析

データの章では、オブジェクトエクスプローラを用いてデータの選択や検索などを行いました。また、グラフの章では、オブジェクトエクスプローラからグラフ描画したい列を選んだり、既に出来上がったグラフをオブジェクトエクスプローラの利用により編集しました。ここでは、オブジェクトエクスプローラによる、統計解析の結果出力、あるいは応用に関して説明します。

オブジェクトエクスプローラに統計解析結果用フォルダを作成する

オブジェクトエクスプローラには多くのデータなどが表示されているので、表示させたいオブジェクトを探しにくいことがあるかもしれません。オブジェクトエクスプローラには、オブジェクトを種類ごとにまとめて表示できるフォルダを作ることができます。

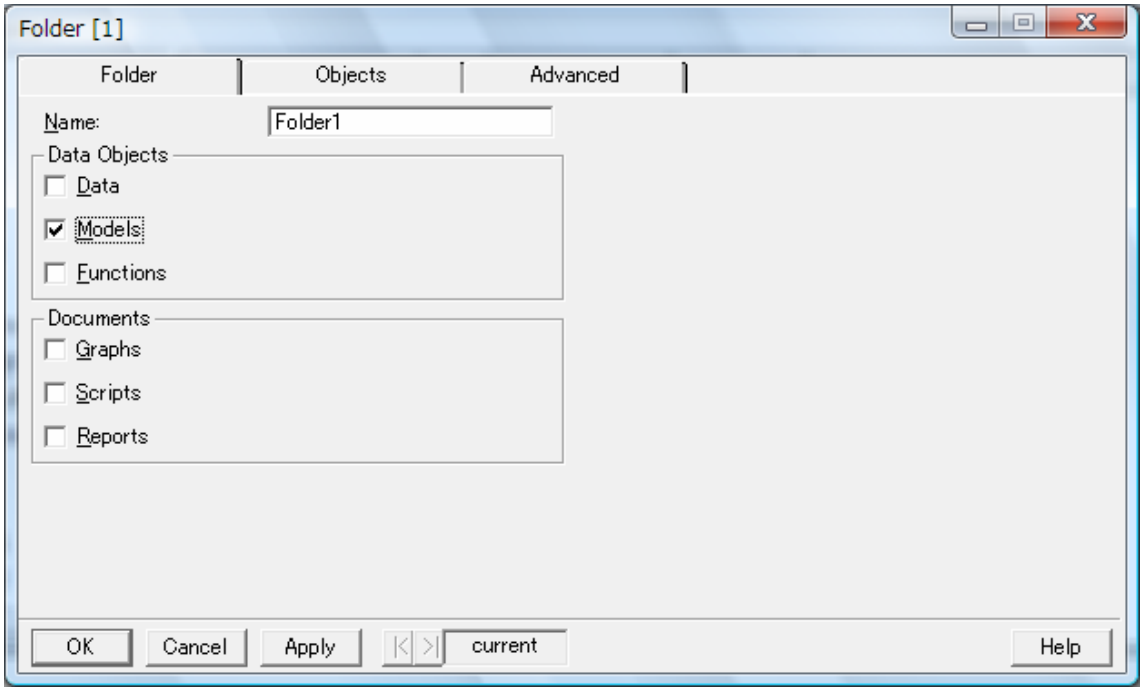
オブジェクトエクスプローラを開く

- メニュー **挿入 / フォルダ** を選択

フォルダの名前は、ダブルクリックで変更することができます

- フォルダを選択して右ボタンクリック、メニューから **Folder** を選択
- **Models** にチェック（**Data** にチェックすると、データだけが、**Functions** にチェックすると関数だけがフォルダに表示されるようになります）

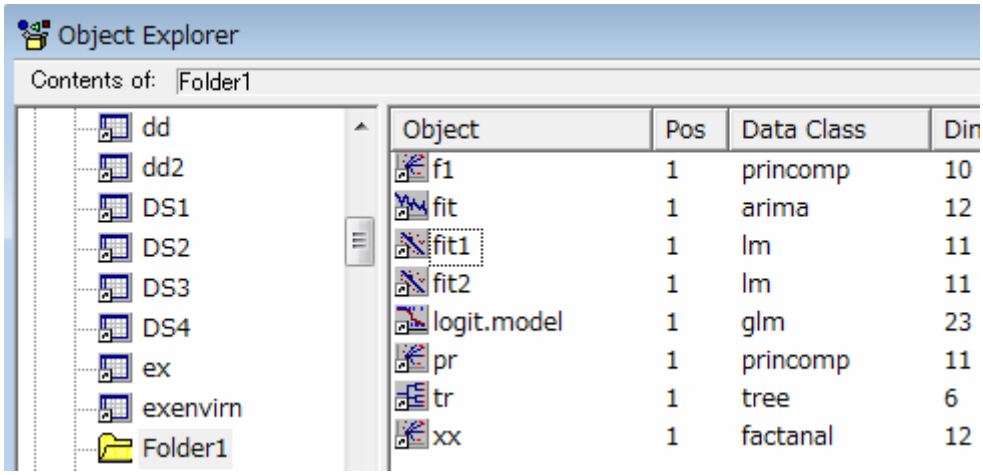
4-9　Folder ダイアログ



できたフォルダはオブジェクトへのショートカットが保存されますが、選択して delete キーを押すとオブジェクトも削除されてしまうので注意してください。フォルダごと削除しても、その中身は削除されません。

線形回帰などを行った結果をダイアログの Save As に名前を指定することにより保存していれば、フォルダに表示されます。行った解析によって、異なるアイコンで表示されています。

4-10 統計解析結果だけを表示するフォルダ



オブジェクトエクスプローラの **Data Class** には、解析の種類が表示されます。

オブジェクトの 1 つを選択し、ダブルクリックすると、そのオブジェクトのサマリーがレポートウィンドウに表示されます。右ボタンクリックすると、その解析の種類に応じて、メニューが現れます。これら選択可能なメニューは、統計から選択して実行する際のダイアログのタブと対応しています。

5　スクリプトの利用

S-PLUS で操作した作業はすべて､スクリプトというプログラムの形で再現することができます。スクリプトは繰り返し使うことができますので、毎回行いたい定型処理などで便利です。

この章の中で説明するスクリプトについては、米国ベル研究所で開発された S 言語とは異なるものです。S 言語は言語バージョンさえ同じならば、異なるプラットフォームの S-PLUS であっても同じプログラムの利用が可能ですが、スクリプトは S-PLUS for Windows 固有の機能です。

S 言語については、別途参考書籍をご覧ください。もちろん、スクリプト上で S 言語のプログラムを組んだり、関数を定義することも可能です。

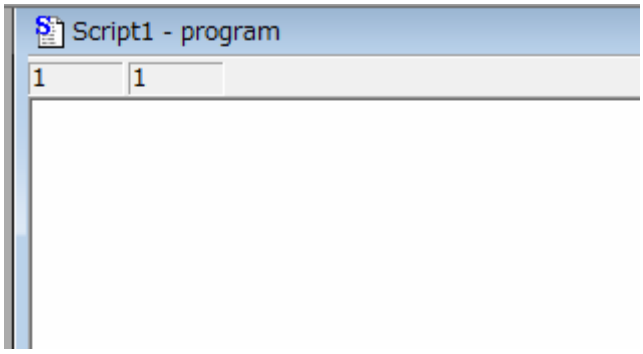
5.1 スクリプトウィンドウ

スクリプトウィンドウはスクリプトを書いたり、その実行を行うウィンドウです。

スクリプトウィンドウを開く

- メニュー **ファイル / 新規ファイル** を選択
- **Script File** を選択

5-1 スクリプトウィンドウを開く

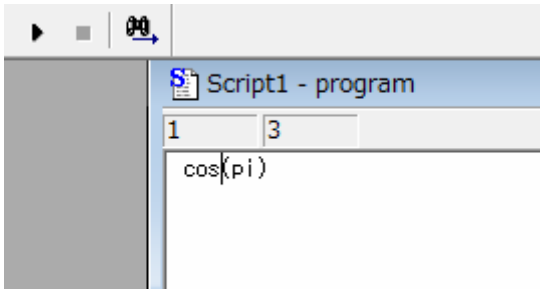


スクリプトウィンドウは 2 つのフレームからなり、上のフレームにはスクリプトを記述し、下のフレームにはプログラムを実行したときのログが残ります。

たとえば、スクリプトウィンドウの上のフレームに次のように入力します。

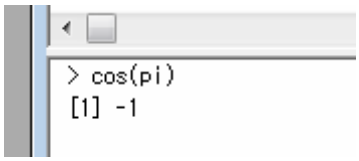
```
cos(pi)
```

5-2 スクリプトを記述し、実行ボタンをクリック



続いて、ツールバーの **実行** ボタンをクリックしますと結果が下のフレームに表示されます。

5-3 スクリプトの実行結果



これは S 言語による簡単な計算の例ですが、実際にはもう少し複雑な操作をスクリプト化しなければなりません。たとえば、32 ページのランダムサンプリングの例をスクリプトにするのには

- 32 ページを参照し、データ `exenvirn` から 10 個のデータを無作為に抽出する
- その際、**Save As** フィールドに `ex2` と入力
- ツールバーから **履歴ログ** ボタンをクリック

スクリプトウィンドウとよく似たウィンドウ **history** が開きます。このウィンドウには、S-PLUS を起動してから今までの操作がスクリプトの形で保存されています。この中から、次のような記述を探します（実行した順に表示されますので、一番下にあるでしょう）。

```
menuRandomSample(data = exenvirn, size = 10, replace = F, save.name =
   "ex2", show.p = T)
```

このスクリプトをコピーペーストで先ほどのスクリプトウィンドウの上のフレーム日貼り付けます。実行ボタンをクリックすると、再び exenvirn データから 10 個、無作為に抽出されます。save.name= 指定を変えることにより、異なる名前で保存することも可能です。

スクリプトは保存をすれば、繰り返し実行することができます。保存はスクリプトウィンドウを選択し、メニュー **ファイル / 名前を付けて保存** とします。

5.2 スクリプトウィンドウの便利な利用

ヘルプを見る

スクリプトウィンドウの上フレームに、オンラインヘルプを参照したい関数名を入力します。関数名を選択したら、右ボタンをクリック、メニューから **ヘルプ** を選択します。

関数の定義を見る

S 言語による関数定義の内容（プログラム）を見るには、スクリプトウィンドウの上側フレームに関数名を入力、右ボタンをクリック、メニューから **インプレース展開** を選択します。

6　S-PLUS のカスタマイズ

6.1 オブジェクトエクスプローラによるデータ管理

これまで、オブジェクトエクスプローラを用いて、データ検索、グラフ作成、統計解析結果の保存などを説明しましたが、オブジェクトエクスプローラには他にもいろいろな機能があります。

データベースのリスト

17 ページの 2.6 S-PLUS のデータベースでは、オブジェクトエクスプローラの **SearchPath** アイコンをクリックすることにより、データベースを一覧表示できることを説明しました。**SearchPath** で表示されるデータベースの 1 番最初のものがユーザーの作業用のデータベースとなり、ユーザーがインポート、入力して作成したデータ、あるいは統計解析結果などはこのユーザーデータベースに自動的に保存され､S-PLUS を起動すれば､常に利用可能になります。

SearchPath の機能について、もう少し説明します。**SearchPath** の左側にある 展開ボタン をクリックすると、すぐ下にデータベースが複数表示されます。リストの 1 番目はユーザーのデータベースですが、2 番目以降は S-PLUS がインストールされた時にインストールされるデータベースで、S 言語の関数などが用途に応じて保存されています。そのうちの 1 つを選ぶと、右側に関数などが表示されます。これらはリードオンリーのデータベースに保存されているので、参照は可能ですが、オブジェクトを選択して削除などの操作はできません。

同じ名前のオブジェクトが存在する場合

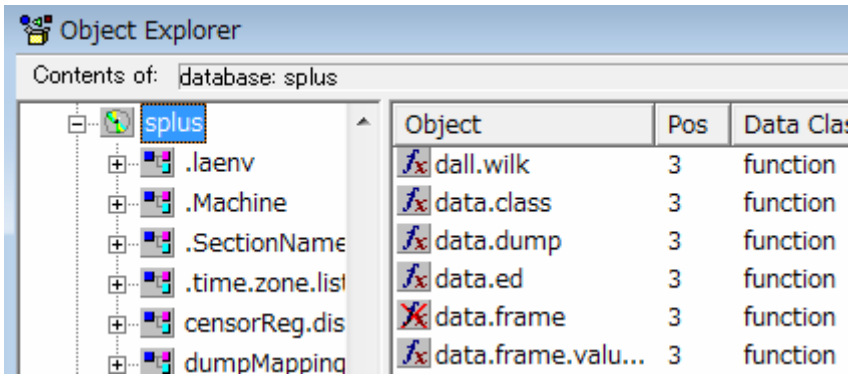
オブジェクトエクスプローラで、自分のデータベースにある不要なデータを 1 つ選択し、右ボタンクリック、メニューから **Properties** を選択し、**Name** フィールドに `data.frame` と入力します。

`data.frame` という関数が、S-PLUS には存在しますので、調べてみます。

SearchPath の 展開ボタン をクリック、左側から、**splus** を選択すると、右側には

splus に存在するオブジェクトが一覧表示されます。右スクロールバーでスクロールして、`data.frame` を探します。

6-1 splus データベースの検索



data.frame のアイコンに赤で×印が付いていることがわかります。これは、このオブジェクトと全く同じ名前のオブジェクトが（データ、関数など種類にかかわらず)、リストの上位データベースに存在することを示しています。ここでは `data.frame` という名前を付けたデータがユーザーのデータベースに存在することを意味します。ユーザーデータベースは **splus** データベースより上位のデータベースです。関数そのものが上書きされているわけではありません。ユーザーのデータベースから、`data.frame` を削除するか、異なる名前にすれば、赤い×印は消えます。このように、同じ名前のオブジェクトが複数のデータベースに存在する場合、上位のデータベースに存在するオブジェクトが、下位のデータベースのオブジェクトを一時的に利用できなくすることを、「マスクする」といいます。場合によっては、問題を起こすことがありますので、このような名前を付けることは避けた方が無難です。避けた方がよい名前としては

- 1 文字のアルファベット
- `matrix, vector, mean,sum` のように、数学でよく用いられる英単語

が挙げられます。

6.2 異なるデータベースを利用する

データベースをプロジェクトによって変えたり、用途に応じて変更することができます。また、利用可能なデータベースを増やすことも可能です。

データベースの変更

オブジェクトを保存するユーザーデータベースを変更するには

- メニュー **ファイル / チャプター / 新規ワーキングチャプター** を選択
- **Browse** ボタンをクリック、書き込み権を持つフォルダを 1 つ選択
- **Label** には、そのデータベースの通称を指定。**SearchPath** で表示されるデータベース名と同じものは利用できません

OK をクリックすると、以降作成されたオブジェクトはすべて、新しいデータベースに保存され、今までユーザーのデータベースとして使ってきたデータベースはオブジェクトの保存のみならず、参照もできなくなります。S-PLUS をシャットダウンすると、データベースは初期状態に戻ります。

現在のユーザーデータベースを変えずに、利用可能なデータベースを増やす

✓ 複数のユーザーで S-PLUS を利用しており、データベースを共有したい
✓ データベースはプロジェクトごとに変更しても、自作の関数などは常に使いたい

このような場合には、利用可能なデータベースを増やすことが有効です。

- メニュー **ファイル / チャプター / チャプターのアタッチ** を選択
- **Browse** ボタンをクリック、書き込み権を持つフォルダを 1 つ選択
- 見ることはできても、書き込み禁止にしたい場合は、**Attach read-only** チェックボックスにチェック
- **Label** には、そのデータベースの通称を指定、**SearchPath** で表示されるデータベース名と同じものは利用できません
- **Position** には、そのデータベースを参照リスト(**SearchPath**)の何番目にするかを指定します。**1** とすれば、ユーザーデータベースとなります。**2** 以降なら、参照するだけのデータベースになります。**1** とする際には、**Attach read-only** チェックボックスにチェックをしないように注意してください。

この設定も、S-PLUS をシャットダウンすると、初期状態に戻ります。

データベース間でのデータのコピーや移動は、オブジェクトエクスプローラで元のデータを選択し、ツールバーのコピーボタンや右ボタンクリックでコピーを選択し、保存先デ

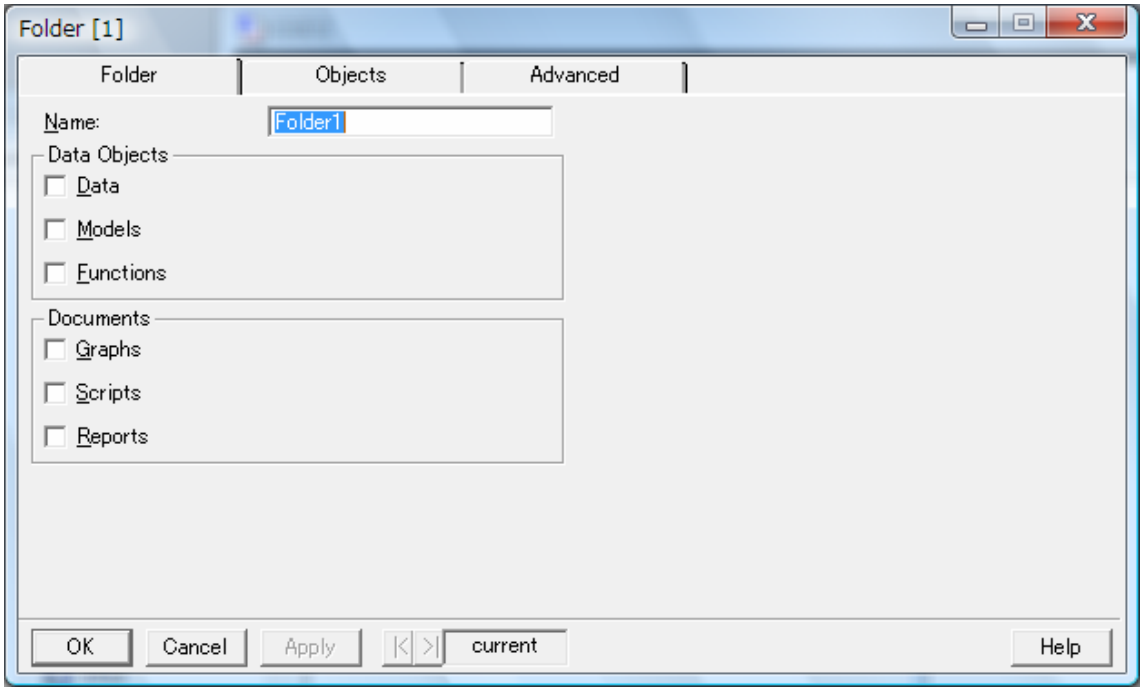
ータベースを **SearchPath** で選んで、ペーストボタンや、右ボタンクリックのペーストを選択します。**Read-only** になっている、S-PLUS のデータベースにはペーストできませんので、注意してください。

オブジェクトエクスプローラに新規フォルダを作る

オブジェクトエクスプローラに多くのデータが表示され、目的とするデータを表示させるのが困難になることがあります。オブジェクトエクスプローラにフォルダを作り、そこにある条件に従ったデータだけを集めて表示させることができます。この場合のフォルダはあくまでショートカットであり、実在のフォルダではないので注意してください。

オブジェクトエクスプローラでメニュー **挿入 / フォルダ** を選択
挿入されたフォルダを選択して、右ボタンクリック、**Folder** を選択

6-2 Folder ダイアログ



タブ **Folder** では、表示させるオブジェクトの種類を選択します。

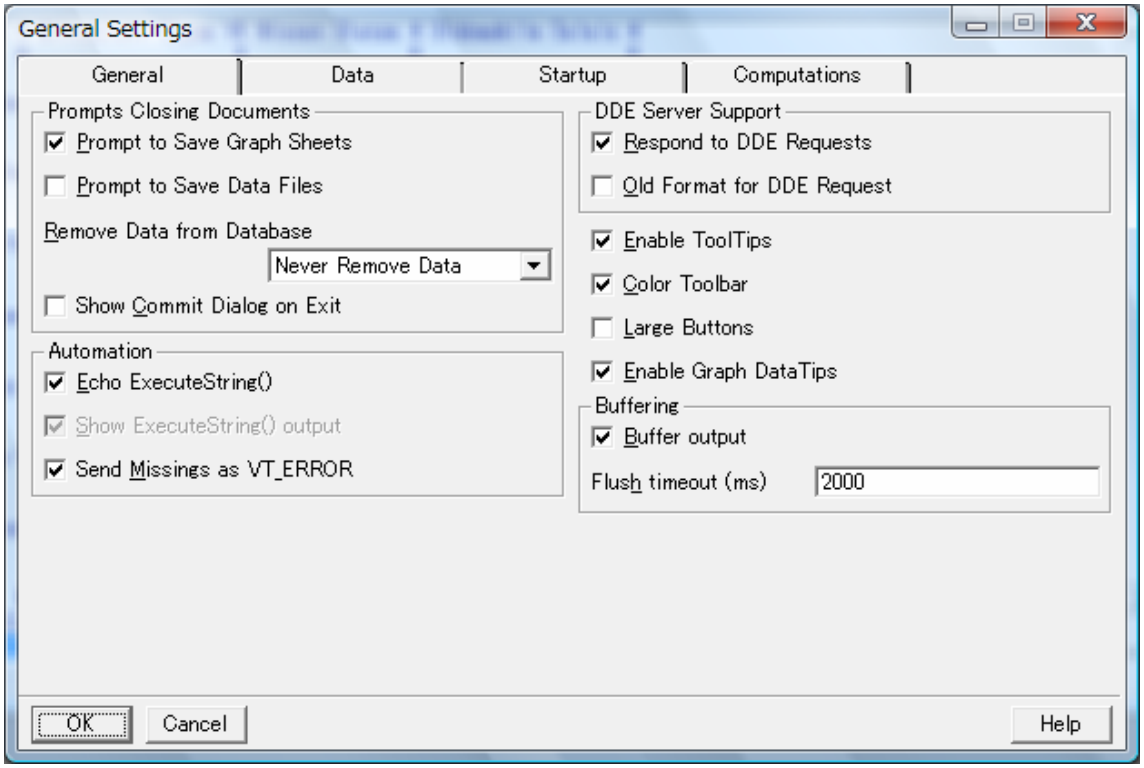
Data	データフレーム
Models	統計解析の結果、各統計ダイアログの **Save As** で指定した場合

Functions	関数、スクリプトではないので注意

6.3 オプションの指定

S-PLUS の利用環境をカスタマイズするためのオプションがメニュー **オプション / 設定** にあります。

6-3 General タブ

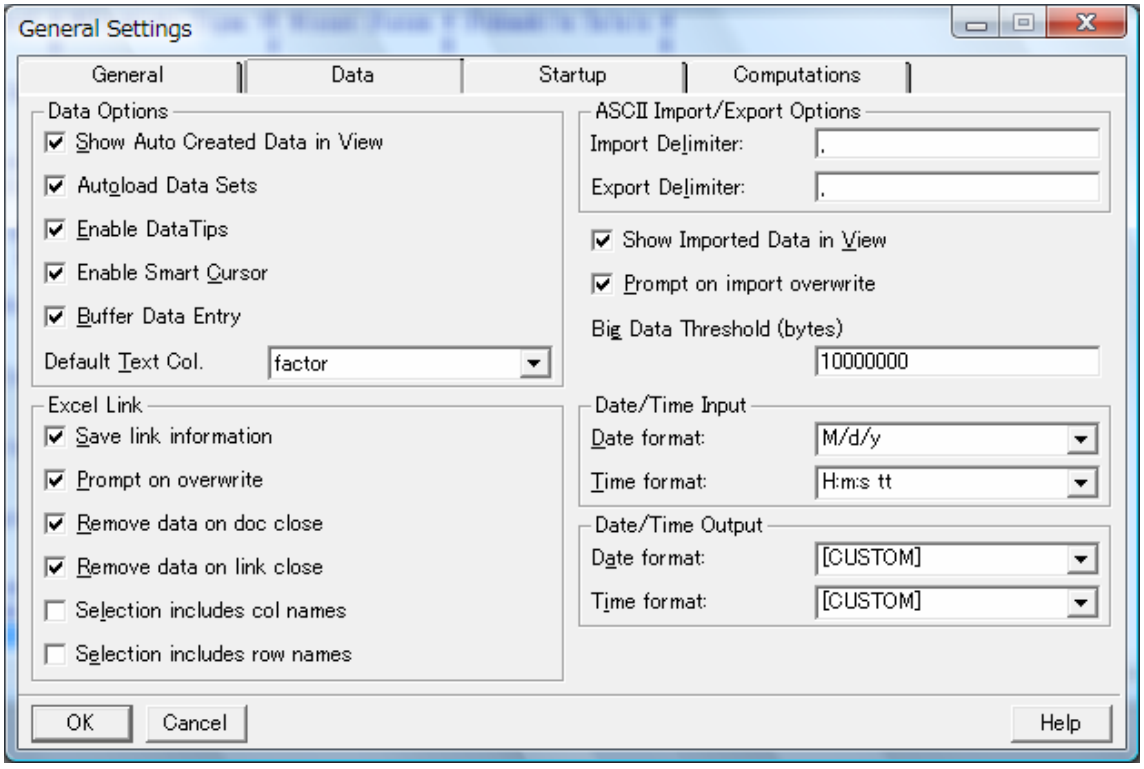


GUI 操作に関連のあるオプション

Prompt to Save Graph Sheets	グラフウィンドウを閉じる際に、保存をするよう促すかどうか
Prompt to Save Data Files	データウィンドウを閉じる際に、保存をするよう促すかどうか（データベースには自動保存されますが、*.sdd ファイルにも保存するかどうか)
Remove Data from Database	データベースからデータを削除するかどうか
Show Commit Dialog on Exit	終了時に、データをデータベースからのぞくか

	どうかの確認ダイアログを出すかどうか
Enable ToolTips	ツールティップ（マウスをボタンに近づけると表示される文字）を表示するかどうか
Color Toolbar	カラーツールバーにするかどうか
Large Button	ボタンを大型にするか
Enable Graph DataTips	データティップ（グラフにマウスポインタを近づけると表示される文字）を表示するかどうか
Buffer Output	出力をバッファするか（まとめて出力するか、逐次出力するか）

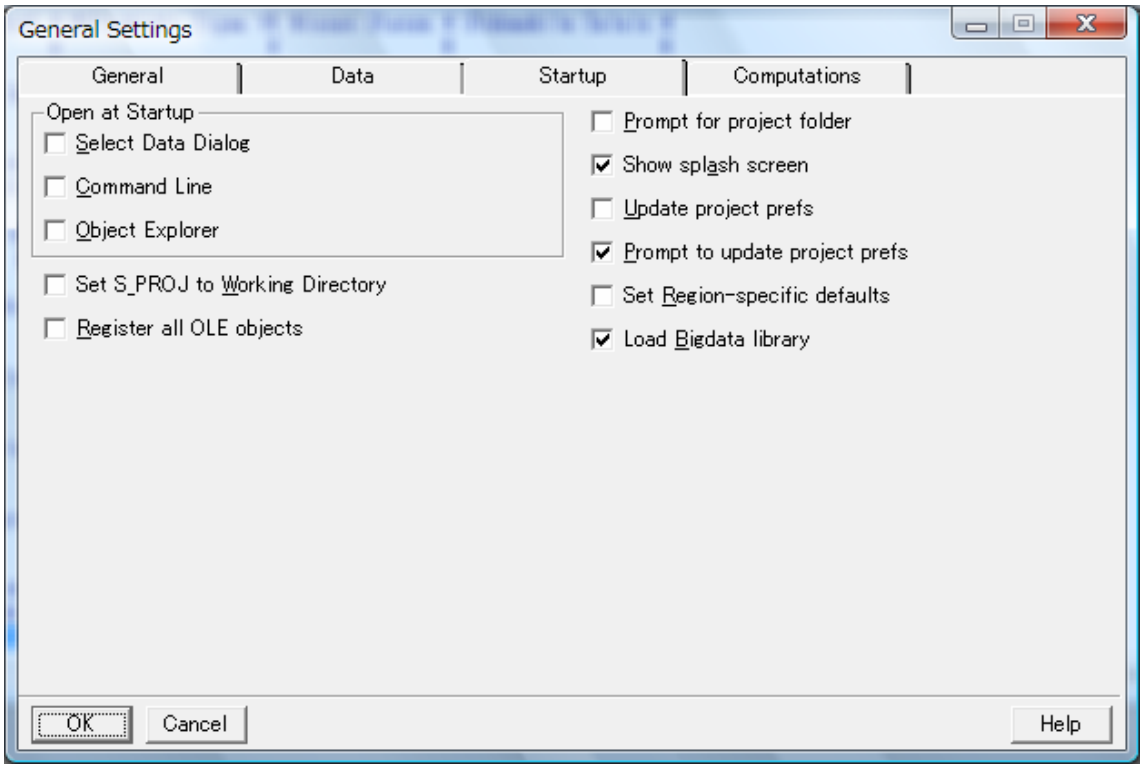
6-4 Data タブ



Show Auto Created Data in View	統計解析やデータメニューで自動的にできたデータをウィンドウに表示するかどうか
Autoload Data Sets	データを変更したときに、グラフをその変更に合わせるかどうか
Enable DataTips	データティップを表示するかどうか
Enable Smart Cursor	入力時にエンターした際の、カーソルを縦、横いずれに進ませるかを、その前の移動方向に倣う

Buffer Data Entry	データを逐次入力するかどうか
Default Text Col.	文字データを character にするか、factor にするか
Import Delimiter	インポート時の標準区切り文字
Export Delimiter	エクスポート時の標準区切り文字
Show Imported Data in View	データインポート時にデータをウィンドウに開くかどうか
Prompt on import overwrite	インポート時に既に同じ名前のデータがあったら、上書きするかどうかを確認するダイアログを表示するかどうか

6-5 Startup タブ



Open at Startup グループ	起動時にどのウィンドウを開くか **Select Data Dialog:**データ選択のダイアログ **Command Line:**コマンドウィンドウ **Object Explorer:**オブジェクトエクスプローラ
Set S_PROJ to Working Directory	ショートカット起動しているときのみ有効、チェックがされていれば、ショートカットのオプショ

	ンで S_PROJ の指定が必要
Register all OLE objects	OLE オブジェクトの登録
Prompt for project folder	起動時にワーキングチャプターフォルダの選択をするかどうか
Show splash screen	S-PLUS の Tips を表示するかどうか
Update project prefs	プロジェクトの.Prefs(ウィンドウやメニューなどの設定を保存するフォルダ)をアップデートするかどうか
Set Region-specific Defaults	地域固有のデフォルト値を用いるかどうか、日付フォーマットなどに適応されます

索引

AIC 95
NA 7, 9
オブジェクト・エクスプローラ 15, 75
character 22
QQ プロット 60
グラフ・ウィンドウ 39
グリッドデータ 58
作業データベース 18
SAS データ 10
条件モードのオンオフ 62
スクリプト 101
スクリプトウィンドウ 101
ステップワイズ法 95
3D プロットパレット 68
データウィンドウ 18
DataTips 19
データの名前 13
データフレーム 7, 12
2D プロットパレット 40
p-value 83
非線形回帰 94
標準正規分布 30
factor 22, 25, 36
変数名 7
マスクする 105
4 方向に回転角で分割 70
乱数 29
履歴ログ 102
列名 7
レポートウィンドウ 81, 89