

S-PLUS for Windows ユーザーズガイド

2012 年2月

所有権について

本ソフトウェア・プログラムと本書の所有権はいずれも株式会社 数理システムおよび Insightful Corporation に帰属します。本プログラムと本書の全著作権はいずれも株式会社 数理システムおよび Insightful Corporation が保有します。

本書の正確な名称は、次の通りです。

『S-PLUS for Windows ユーザーズガイド』株式会社 数理システム

日本語版

著作権

Copyright © 1987-2008, Insightful Corporation. All rights reserved.

Insightful Corporation
1700 Westlake Avenue N, Suite 500
Seattle, WA 98109-3044
USA

株式会社 数理システム
東京都新宿区新宿 2-4-3
フォーシーズンビル 10F

商標

Insightful、Insightful Corporation、Insightful のロゴ、S-PLUS、Insightful Miner、S+FinMetrics、S+SeqTrial、S+SpatialStats、S+ArrayAnalyzer、S+EnvironmentalStats、S+Wavelets、S-PLUS Graphlets、および Graphlet は、米国および他の国々における Insightful Corporation の商標または登録商標です。Intel および Pentium は、米国および他の国々における Intel Corporation またはその系列会社の商標または登録商標です。Microsoft、Windows、MS-DOS、および Windows NT は、米国および他の国々における Microsoft Corporation の登録商標または商標です。本書で使用されるすべての製品名は、それらの会社の商標または登録商標である場合があります。

謝辞

S-PLUS は、AT&T のベル研究所 S チーム (現在の Lucent Technologies)、とりわけ John M. Chambers、Richard A. Becker (現在 AT&T Laboratories)、Allan R. Wilks (現在 AT&T Laboratories)、Duncan Temple Lang、および Lucent の統計学研究部門のスタッフ、William S. Cleveland、Trevor Hastie (現在スタンフォード大学)、Linda Clark、Anne Freeny、Eric Grosse、David James、José Pinheiro、Daryl Pregibon、および Ming Shyu の先駆的研究なしには存在しえなかったでしょう。

Insightful Corporation は、本バージョンとこれまでのバージョンの S-PLUS にご協力頂いた、Douglas M. Bates、Leo Breiman、Dan Carr、Steve Dubnoff、Don Edwards、Jerome Friedman、Kevin Goodman、Perry Haaland、David Hardesty、Frank Harrell、Richard Heiberger、Mia Hubert、Richard Jones、Jennifer Lasecki、W.Q. Meeker、Adrian Raftery、Brian Ripley、Peter Rousseeuw、J.D. Spurrier、Anja Struyf、Terry Therneau、Rob Tibshirani、Katrien Van Driessen、William Venables、および Judy Zeh の各氏に感謝致します。

目次

謝辞	iii
第 1 章 はじめに	1
S-PLUS へようこそ	2
インストール	4
ヘルプ、サポートおよび教材	7
表記規則	15
第 2 章 データによる作業	17
はじめに	18
データの入力、編集および保存	20
データの表示とフォーマット	28
データの操作	41
S-PLUS に含まれるライブラリ	56
第 3 章 データの検索	59
はじめに	60
1次元データの視覚化	61
2次元データの視覚化	66
多次元データの視覚化	92

第 4 章 グラフの作成	121
はじめに	123
1 次元データのプロット	126
2 次元データのプロット	135
多次元データのプロット	147
Trellis グラフ	162
第 5 章 データのインポートとエクスポート	163
はじめに	164
インポートとエクスポートがサポートされたファイルタイプ	165
データファイルからのインポートとデータファイルへのエクスポート	169
ODBC テーブルからのインポートと ODBC テーブルへのエクスポート	180
フィルタ式	188
特定タイプのファイルのインポートとエクスポートに関する注意	191
第 6 章 グラフの編集	197
グラフ	199
グラフのフォーマット	213
グラフオブジェクトで作業する	232
図表示	234
グラフスタイルを使用し色をカスタマイズする	237
グラフシートへのデータの埋込みと抽出	239
オブジェクトのリンクと埋込み	240
グラフを印刷する	243
グラフをファイルにエクスポートする	244
第 7 章 S-PLUS GRAPHLETS™	247
はじめに	248
Graphlet データファイルの作成	252
ウェブページへの Graphlet の埋め込み	263
Graphlet の使い方	269

第 8 章 統計	273
はじめに	276
要約統計	282
標本比較	291
検出力と標本数	335
実験計画	340
回帰	346
分散分析	376
混合効果モデル	382
一般化最小二乗モデル	386
生存時間解析モデル	389
樹形モデル	397
モデルの比較	401
クラスター分析	404
多変量解析	415
QC チャート (品質管理チャート)	421
標本からのサンプリング	426
平滑化	430
時系列解析	434
乱数と分布	441
参考資料	451
第 9 章 オブジェクトおよびデータベースによる作業	453
はじめに	454
オブジェクトタイプとデータベース	455
オブジェクト・エクスプローラについて	461
オブジェクトによる作業	472
作業の構成	482
第 10 章 コマンド・ウィンドウの使い方	489
はじめに	491
コマンド・ウィンドウの基礎	492
S-PLUS 言語の基礎	500

データのインポートと編集	514
データのサブセット抽出	518
S-PLUS のグラフィックス	522
統計	526
定義関数	532
バッチモードでの S-PLUS の使用	533
第 11 章 スクリプト・ウィンドウと レポート・ウィンドウの使い方	535
はじめに	536
スクリプト・ウィンドウ	538
スクリプト・ウィンドウの機能	546
スクリプト使用のための時間節約のヒント	549
レポート・ウィンドウ	554
スクリプトまたはレポートの印刷	556
第 12 章 他のアプリケーションでの S-PLUS の使用	557
Microsoft Excel での S-PLUS の使用	558
SPSS での S-PLUS の使用	577
MathSoft Mathcad での S-PLUS の使用	583
Microsoft PowerPoint での S-PLUS の使用	590
第 13 章 S-PLUS セッションのカスタマイズ	595
はじめに	596
デフォルトと設定の変更	597
起動時と終了時のセッションのカスタマイズ	624

第1章

はじめに

S-PLUS へようこそ	2
インストール	4
システム要件	4
インストール手順	5
S-PLUS を実行する	6
ヘルプ、サポートおよび教材	7
オンライン ヘルプ	7
オンライン マニュアル	10
ワンポイント	11
S-PLUS オン・ザ・ウェブ	11
トレーニングコース	11
技術サポート	12
S-PLUS を使用している書籍	13
表記規則	15

S-PLUS へようこそ

S-PLUS は、Lucent Technologies 社で開発された高性能でオブジェクト指向の S 言語の最新版を利用した S-PLUS の全く新しいバージョンです。S 言語は、対話的データ探索用に設計された豊かな環境を備え、特にデータのビジュアル化と探索、統計モデリングおよびデータプログラミングのために作成された唯一の言語です。

S-PLUS は、お客様のデータ解析および専門的グラフ作成の必要性を満たす最も優れたソリューションです。Microsoft Office 互換のユーザインタフェースによって、ボタンを押すだけでデータ操作、グラフ作成、および統計処理を行うことができます。S-PLUS では、S-PLUS プログラム言語を使って、対話的にプログラムを作成することができます。

注意

S-PLUS には、Enterprise Developer、および Student Edition のデスクトップバージョンがあります。

Enterprise Developer 版は、すべての機能が使用できる無期限の S-PLUS for Windows です。Big Data および S-PLUS Workbench に対応しており、ライセンスは無期限です。

Student Edition 版は、データのインポートを 20,000 セル以下、1,000 列以下に制限しています。Student Edition 版では、Excel および SPSS アドインの使用および Big Data ライブラリの使用はできません。テクニカルサポートは対象外です。

通常の S-PLUS セッションでは、以下のことが可能です。

- ほとんどすべてのデータソースからのデータのインポート
- 使いやすいデータ・ウィンドウでのデータの表示と編集
- ボタンクリックによるプロットの作成
- グラフの細部の制御と、報告書類にエクスポートするための見ばえのよい出力の生成

- メニュー形式の使いやすいダイアログによる統計解析の実行
- コマンドラインで、または**スクリプト・ウィンドウ**を使用したバッチで1つずつ解析機能を実行
- ユーザ独自の関数の作成
- ユーザインタフェースの詳細なカスタマイズ

インストール

ソフトウェアのインストールは、次のように行ってください。詳しくは「インストールガイド」をご覧ください。

1. S-PLUS CD を CD-ROM ドライブに入れます。
2. オペレーティングシステムが AutoPlay をサポートしている場合は、インストールは自動的に進みます。そうでない場合は、CD-ROM のルートディレクトリにある **setup.exe** を実行してください。
3. セットアップ画面の指示に従ってください。このとき、デフォルト設定を使用されることを推奨します。

インストールソフトウェア InstallShield に問題が発生するのを避けるために、S-PLUS のインストール中は、特にウイルスチェッカーなどのアプリケーションを終了しておくことを推奨します。

システム要件

- 推奨最低システム構成：通常のインストールでは、RAM が 512MB 以上の Pentium III 以上のコンピュータ。完全なインストールには、空きディスク容量が少なくとも 450MB 必要です（また、ドライブ C:¥にインストールしない場合は、ソフトウェアを解凍するための空きディスクスペースとしてドライブ C:¥にさらに 50MB が必要です）。
- オペレーティングシステム：Intel プラットフォーム上で動作する Windows 2000、Windows XP Home Edition、Windows XP Professional Edition、Windows Vista、および Windows 2003 Server

注意

S-PLUS は、Win32s（すなわち、Windows 3.1x）をサポートしていません。また、Windows NT 3.51 をサポートしていません。

- SVGA、または他のほとんどの Windows 互換グラフィックスカード および解像度 800×600 以上のモニタ
- ローカルまたはネットワーク接続された 1 台の CD-ROM ドライブ
- Microsoft マウスまたは他の Windows 互換ポインティングデバイス
- Windows 互換のプリンタ（オプション）

インストール 手順

ソフトウェアをインストールし、新しい FLEXnet ライセンスマネージャを起動し、S-PLUS を実行するためには、以下の 3 つの基本ステップがあります。

1. CD の `setup.exe` を実行し、セットアップ画面に従いインストールします。
2. インストール後、ライセンスキーファイルをインストールします。PC がインターネット接続可能ならば、初回の S-PLUS 実行時に自動取得が可能です。
3. S-PLUS を実行します。

既存のバージョンの S-PLUS の上にこのリリースをインストールしないでください。その代わりに S-PLUS 用の新しいインストールディレクトリを指定してください（特に指定をしなければ、新しいインストールディレクトリが作られます）。

詳しくは「インストールガイド」をご覧ください。

S-PLUS を 実行する

Windows では S-PLUS を以下の方法で起動することができます。

- スタートメニュー ▶ プログラム ▶ S-PLUS 8.0 ▶ S-PLUS Interface (GUI) を起動します。
- スタートメニュー ▶ プログラム ▶ S-PLUS 8.0 ▶ S-PLUS Workbench を起動します。
- スタートメニュー ▶ プログラム ▶ S-PLUS 8.0 ▶ S-PLUS Console を起動します。
- 対話形式で使用するには DOS コマンドラインから Windows Console を起動します。
- 「Sqpe.exe infile outfile」を使って Windows バッチファイルからコンソールバージョンを実行します。
- 「S-PLUS BATCH」を使って Windows バッチファイルから S-PLUS GUI バージョンを実行します。
- Excel アドインや SPSS アドインによる自動化、または SpotFire や PharSight などのカスタム自動化アプリケーションから S-PLUS GUI バージョンを実行します。
- DDE によって S-PLUS GUI バージョンを実行します。
- Connect/C++または Connect/Java によって Console バージョンを実行します。

ヘルプ、サポートおよび教材

S-PLUS の修得を支援するための様々な手段を用意しています。この節では、S-PLUS ユーザが利用できる教材とサポート体制について説明します。

オンラインヘルプ

S-PLUS には、S-PLUS の学習と使用を容易にするためのオンライン HTML ヘルプシステムがあります。S-PLUS グラフィカル・ユーザインタフェースの使い方は、ヘルプメニューにあります。さらに、S 言語仕様書で、S-PLUS 言語の各関数の詳細を知ることができます。また、コマンド・ウィンドウで、プロンプトに `help()` と入力することによって、S 言語仕様書を表示させることができます。

様々なダイアログにある **Help** ボタンを押すか、ツールバーにあるヘルプボタンをクリックするか、S-PLUS がアクティブのときに **F1** キーを押すことによって、その状況に対応したヘルプを表示させることができます。

HTML ヘルプ

S-PLUS の HTML ヘルプは、Microsoft Internet Explorer を利用して HTML ウィンドウにヘルプファイルを表示します。HTML ヘルプを利用するときは、次のいずれかの操作を行ってください。

- グラフィカル・ユーザインタフェースのヘルプは、メインメニューから **Help (ヘルプ)** ▶ **Available Help (有効なヘルプ)** ▶ **S-PLUS Help (S-PLUS ヘルプ)** を選択します。
- S-PLUS プログラム言語のヘルプは、メインメニューから **Help (ヘルプ)** ▶ **Available Help (有効なヘルプ)** ▶ **Language Reference (S 言語仕様書)** を選択します。

第1章 はじめに

図 1.1 に示すように、HTML ヘルプ・ウィンドウは、ツールバー、左ペインおよび右ペインの3つの領域に大きく分かれています。

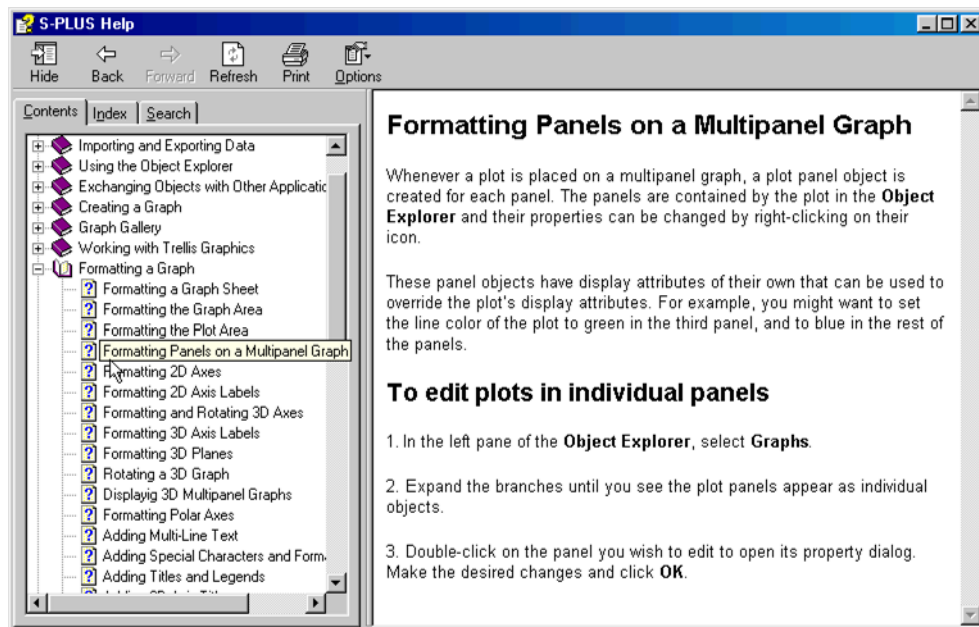


図 1.1 : S-PLUS ヘルプ・ウィンドウ

ツールバーの使い方

表 1.1 に、ヘルプ・ウィンドウのツールバーにある4つのメインボタンをリストします（数がさらに多い場合もあります）。

表 1.1 : ヘルプ・ウィンドウのツールバーボタン

ボタン名	説明
非表示 (または表示)	非表示と示されたボタンを押すと、左ペインが隠され、右ペインがヘルプ・ウィンドウの幅一杯に広がります。表示と示されたボタンを押すと、左ペインが表示され、その結果ヘルプ・ウィンドウが分割されます。
戻る	前に表示されたヘルプトピックに戻ります。
進む	次のヘルプトピックに進みます。
印刷	表示されているヘルプトピックを印刷します。

左ペインの使い方

ヘルプ・ウィンドウ自体と同じように、左ペインは、**目次**、**キーワード**および**検索**のタブのある 3 つの部分に分かれています。

- 関連したヘルプファイルを見つけやすいように、**目次**タブは、ヘルプトピックがそのカテゴリで構成されています。これらのカテゴリは、カテゴリの名前がついた小さい本のアイコンとして表示されています。カテゴリを開くときは、そのアイコンまたはラベルをダブルクリックしてください。カテゴリ内のトピックを選択するときは、その疑問符アイコンまたはトピックタイトルをダブルクリックしてください。
- **キーワード**タブには、ヘルプトピックがキーワードでリストされています。キーワードは、一般に、**S-PLUS** 言語の関数名と、グラフィカル・ユーザインタフェースのトピック名です。キーワードを入力するだけで、**HTML** ヘルプが、それに最も近いキーワードを探します。ヘルプトピックを表示するときは、**表示**をクリック（または、選択したタイトルをダブルクリック）してください。
- **検索**タブは、ヘルプシステム全体を全文検索します。キーワードを入力すると、そのキーワードを含むすべてのヘルプファイルが、リストボックスにリストされます。ヘルプトピックを表示させるときは、必要なトピックを選択し、**表示**をクリック（または、選択したタイトルをダブルクリック）してください。

右ペインの使い方

右ペインには、実際にヘルプ情報が表示されます。通常は、縦と横のスクロールバーが表示されていますが、**HTML** ヘルプ・ウィンドウを広げて右ペインの幅を大きくすることができます。多くのヘルプファイルは長すぎて全体を 1 つの画面に表示できないため、都合のよい **HTML** ヘルプ・ウィンドウの高さを選択して、縦スクロールバーを使ってテキストをスクロールさせます。

第1章 はじめに

右ペインには、トピック内を検索する機能があります。次のように行います。

1. **CTRL-F** を入力して**検索**ダイアログを開きます（このダイアログは、Internet Explorer から受け継いだ HTML ヘルプの機能です）。



2. **検索する文字列**と示されたテキストフィールドに、検索する文字列を入力します。
3. **次を検索**をクリックします。

コマンド・ウィンドウとスクリプト・ウィンドウのヘルプ

コマンド・ウィンドウで作業しているとき、?またはhelp 関数を使ってコマンドについて調べることができます。たとえば、anova のヘルプファイルを開くときは、

```
> help(anova)
```

または、

```
> ?anova
```

と入力します。スクリプト・ウィンドウで作業しているときにコマンドについて調べたいときは、そのコマンドをハイライトし **F1** を押してください。

オンラインマニュアル

このユーザーズガイドの他に、小冊子『Getting Started Guide』『プログラマーズ ガイド』、および『Guide to Statistics 1』『Guide to Statistics 2』の2巻をオンラインで使用することができます。『Getting Started Guide』は、製品を紹介するチュートリアルで、特に S-PLUS を使用し始めたお客さまに役立ちます。

マニュアルをオンラインで見るときは、メインメニューから**ヘルプ ▶ オンラインマニュアル**を選択し、見たいタイトルを選択してください。

注意：文書のオンライン版

オンラインマニュアルは、Acrobat Reader で表示されます。Acrobat Reader は、S-PLUS をインストールする際にオプションとしてインストールすることができます。一般に、マニュアルの最初の目次よりも、Acrobat Reader の“しおり (bookmarks)” (メニューバーの「表示」の項目にある) を利用した方が、参照が容易です。しおりは、常に表示させておくことができ、拡張したり縮小したりして、章のタイトルだけを表示させたり節の見出しを含むように表示させたりすることができます。

ワンポイント

S-PLUS の修得を支援するために、デフォルトではプログラムを起動するたびに簡単な「ワンポイント」が表示されます。(図 1.2 を参照。)



図 1.2 : ワンポイント

また、メインメニューのヘルプ ▶ ワンポイントを選択することによって S-PLUS ワンポイントをいつでも表示させることができます。この機能を無効にする場合は、ダイアログ内の **Show tips on startup** チェックボックスのチェックを外してください。

S-PLUS オン・ザ・ ウェブ

数理システムのウェブサイト <http://www.msi.co.jp/splus/> で、S-PLUS について知ることができます。このサイトには、次のような様々な情報が含まれています。

- FAQ ページ
- 最新のサービスパック
- トレーニングコースの情報
- 製品情報
- 教材の情報

トレーニング コース

数理システムの教育サービスには、S-PLUS でデータを解析する効率と効果を短時間で高めるためにいくつかのトレーニングコースがあります。これらのコースは、統計専門家によって指導されます。このコースの特徴は、体験的な学習アプローチにあり、授業は、講義とオンライン実習とに分かれています。参加者全員に、コースで使用されるテキストが提供されます。

技術サポート

技術サポート窓口

- ファックス : 03-3358-1727
- 電子メール : splus-support@msi.co.jp

またはウェブページ <http://www.msi.co.jp/splus/>

S-PLUS を 使用している 書籍

全般

Becker, R.A., Chambers, J.M., and Wilks, A.R. (1988). *The New S Language*. Wadsworth & Brooks/Cole, Pacific Grove, CA.

Burns, Patrick (1998). *S Poetry*. Download for free from <http://www.seanet.com/~pburns/Spoetry>.

Chambers, John (1998). *Programming with Data*. Springer-Verlag.

Krause, A. and Olson, M. (1997). *The Basics of S and S-PLUS*. Springer-Verlag, New York.

Lam, Longhow (1999). *An Introduction to S-PLUS for Windows*. CANDienst, Amsterdam.

Spector, P. (1994). *An Introduction to S and S-PLUS*. Duxbury Press, Belmont, CA.

データ解析

Bowman, Adrian and Azzalini, Adelchi (1997). *Smoothing Methods*. Oxford University Press.

Bruce, A. and Gao, H.-Y. (1996). *Applied Wavelet Analysis with S-PLUS*. Springer-Verlag, New York.

Chambers, J.M. and Hastie, T.J. (1992). *Statistical Models in S*. Wadsworth & Brooks/Cole, Pacific Grove, CA.

Efron, Bradley and Tibshirani, Robert J. (1994). *An Introduction to the Bootstrap*. Chapman & Hall.

Everitt, B. (1994). *A Handbook of Statistical Analyses Using S-PLUS*. Chapman & Hall, London.

Härdle, W. (1991). *Smoothing Techniques with Implementation in S*. Springer-Verlag, New York.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman & Hall.

Huet, Sylvie, et al. (1997). *Statistical Tools for Nonlinear Regression: with PLUS*. Springer-Verlag.

Kaluzny, S.P., Vega, S.C., Cardoso, T.P., and Shelly, A.A. (1997). *S+SpatialStats User's Manual*. Springer-Verlag, New York.

Marazzi, A. (1992). *Algorithms, Routines and S Functions for Robust Statistics*. Wadsworth & Brooks/Cole, Pacific Grove, CA.

Millard, Steven (1998). *User's Manual for Environmental Statistics*. Companion book to the S+Environmental Stats module. (The S+Environmental Stats module is available through Dr. Millard.)

Selvin, S. (1998). *Modern Applied Biostatistical Methods: Using S-PLUS*. Oxford University Press.

Venables, W.N. and Ripley, B.D. (1999). *Modern Applied Statistics with S-PLUS*, Third Edition. Springer-Verlag, New York.

グラフィック技術

Chambers, J.M., Cleveland, W.S., Kleiner, B., and Tukey, P.A. (1983). *Graphical Techniques for Data Analysis*. Duxbury Press, Belmont, CA.

Cleveland, W.S. (1993). *Visualizing Data*. Hobart Press, Summit, NJ.

Cleveland, W.S. (1994). *The Elements of Graphing Data*, revised edition. Hobart Press, Summit, NJ.

最新情報は S-PLUS のホームページをご覧ください。

表記規則

この『ユーザーズガイド』全体にわたって、以下の表記規則を使用します。

- このフォント (font) は、S-PLUS の式とコードサンプルに使用されます。
- このフォント (**font**) は、S-PLUS ユーザインタフェースの要素、オペレーティングシステムのファイルとコマンド、およびダイアログフィールドへのユーザ入力に使用されます。
- このフォント (*font*) は、本のタイトル (英文) を表示する時に使用されます。
- 大文字／小型大文字は、キー名に使用されます。たとえば、シフトキーは、**SHIFT** と表します。
- 複数のキーを同時に押すときは、2つのキーの名前の間にハイフン (-) を入れて表します。たとえば、**SHIFT** と **F1** のキーの組合せは、**SHIFT-F1** と表します。
- メニューの選択は、メニュー内の選択を示すときは、**ファイル ▶ 新規ファイル**のように矢印記号 (▶) を使って省略した形で表します。

第1章 はじめに

第 2 章

データによる作業

はじめに	18
データの入力、編集および保存	20
データセットを作成する	20
データの入力と編集	22
データを保存する	24
データの表示とフォーマット	28
データセットを表示する	28
データを選択する	32
列をフォーマットする	34
行フォーマットする	39
データの操作	41
データの移動とコピー	41
データを挿入する	45
データを削除する	47
データをソートする	51
その他のデータ操作オプション	53
S-PLUS に含まれるライブラリ	56

はじめに

S-PLUS において、データを表示、編集、フォーマット、操作するための最も重要なツールは、**データ・ウィンドウ**です。このウィンドウは、スプレッドシートと似ていますが、セル指向ではなく列指向です。

下の図 2.1 は、**データ・ウィンドウ**に表示されたサンプルデータセット air を示しています。

	1	2	3	4	5
	ozone	radiation	temperature	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	
6	2.67	99.00	59.00	13.80	
7	2.00	19.00	61.00	20.10	
8	2.52	256.00	69.00	9.70	
9	2.22	290.00	66.00	9.20	
10	2.41	274.00	68.00	10.00	

図 2.1 : データ・ウィンドウに表示されたサンプルデータ

注意

S-PLUS は、内部データベースにいくつかのサンプルデータセットが入った状態で出荷されます。このデータセットは、お客さまが S-PLUS を使い始めたばかりの段階で役に立ちます。このサンプルデータを見るときは、次のように行います。

1. 標準ツールバーの**オブジェクト・エクスプローラ**ボタン  をクリックして、**オブジェクト・エクスプローラ**を開きます。
2. **オブジェクト・エクスプローラ**の左ペインで、SearchPath オブジェクトの左側にある符号「+」をクリックして、検索パスにデータベースの名前を表示させます。
3. そのデータベースに含まれるすべてのオブジェクトを右ペインに表示させるときは、データベース名（たとえば、**data**）の左側のアイコンをクリックします。

オブジェクト・エクスプローラの詳しい説明は、第9章「オブジェクトおよびデータベースによる作業」を参照してください。

データ・ウィンドウを一度にいくつも開き、異なるデータセットを表示させたり、1つのデータセットを同時に表示させたりすることができます。

データ・ウィンドウを開くと、データ・ウィンドウ・ツールバーが自動的に表示されます。図 2.2 に示すツールバーには、よく使われる編集コマンドを簡単に実行するためのボタンがあります。

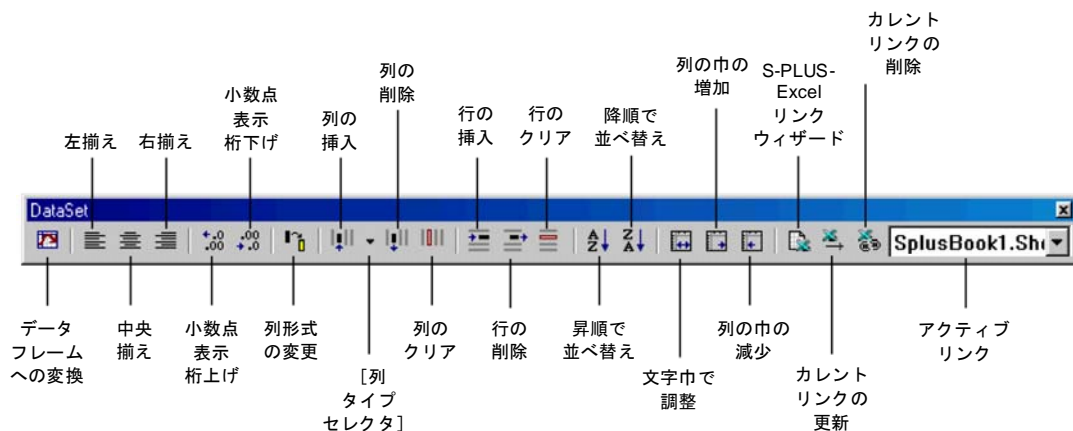


図 2.2 : データ・ウィンドウ・ツールバー

注意

データ・ウィンドウ・ツールバーの Excel 部分の詳しい説明は、560 ページの「S-PLUS to Excel Link Wizard の使い方」を参照してください。

以下の節では、データ・ウィンドウの主な機能を紹介し、一般的な編集作業を行う手順を段階的に説明します。

データの入力、編集および保存

S-PLUS にデータを入力する方法はいくつかあります。最も簡単な方法は、Excel、Lotus、SAS などの別のソースからデータをインポートすることです。また、**データメニュー**には、データを生成するためのいくつかのオプションがあります。たとえば、**変換**オプションを使用して、データセット内の 1 つの列に一連の操作を行い、その結果を別の列に挿入することができます。データを生成するためのもう 1 つの強力なツールは、**コマンド・ウィンドウ**です。S 言語で式を書き込むことによって、たとえば、2 つの列を加えてその結果を第 3 の列に挿入することができます。

S-PLUS にデータを入れる最も基本的な方法は、もちろん単純にキーボードからデータを入力することです。この節では、その操作を説明します。

データセットを作成する

新規データセットを作成するときは、まず以下のいずれかの操作で新しいデータ・ウィンドウを開いてください。



- 標準ツールバーの**新規データセット**ボタン  をクリックします。
- 標準ツールバーの**新規ファイル**ボタン  をクリックするか、メインメニューの**ファイル** ▶ **新規ファイル**を選択します。**New** ダイアログで、**Data Set**を選択し **OK** をクリックします。

図 2.3 に示すように、デフォルト名 SDFx がついた空白の**データ・ウィンドウ**が開きます (x は、連続番号)。

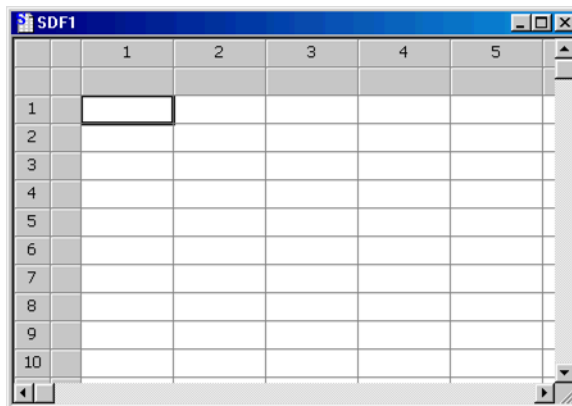


図 2.3 : 空白のデータ・ウィンドウ

新規データセットに名前を付けるときは、以下の操作を行ってください。

1. **データ**・ウィンドウの左上角の灰色のセルをダブルクリックします。
図 2.4 に示すように、**Data Frame** ダイアログが開きます。

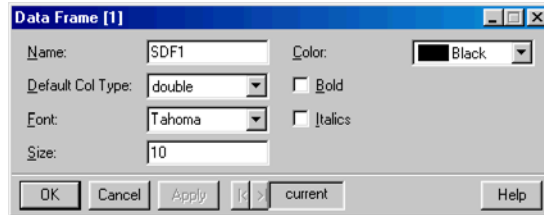


図 2.4 : Data Frame ダイアログ

2. **Name** テキストボックスに新しい名前を入力し、**OK** をクリックします。

注意

データセット名は、文字、数字およびピリオドを含むことができますが、数字で始めることはできません。

また、次のように**データ**メニューを使用して新規データセットを作成し、そのデータセットの名前を同時に変更することもできます。

1. メインメニューから、**データ** ▶ **データの選択**を選択します。図 2.5 に示すように、**Select Data** ダイアログが開きます。

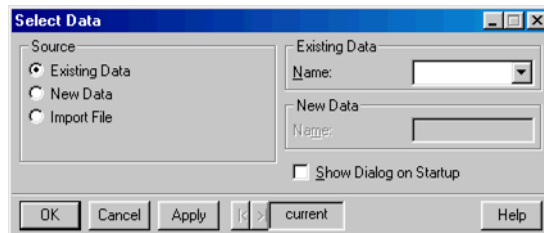


図 2.5 : Select Data ダイアログ

2. **Source** グループの **New Data** ラジオボタンをクリックします。
3. **New Data** グループの **Name** テキストボックスに、データセットの新しい名前を入力し、**OK** をクリックします。

データの 入力と編集

データ・ウィンドウへのデータの入力は、以下のように簡単に行うことができます。

1. データ値を入力したいセルをクリックします。
2. 値を入力します。
3. **ENTER** または矢印キーを押してセル内のデータを確定します。

ENTER を押すとセルに値が入り、カーソルが次のセルに移動します。**S-PLUS** の「スマートカーソル」機能によって、カーソルは前に移動した方向に移動します。データ値を入力した後で矢印キーを押すと、カーソルは矢印の方向に移動します。

注意

S-PLUS では、データセットの列は同じ長さでなければならず、短い列があると **NA** を挿入して他の列と同じ長さにします。

新しい空白の列にデータを入力すると、**S-PLUS** は、その列に、入力したデータの形式に最も近い形式を指定します。新しい列のデフォルトの列形式は、**double** (浮動小数点倍精度実数) です。空白の列に文字データを入力すると、**factor** 列が作成されます (カテゴリーカルデータと解釈する)。

文字データの場合のデフォルト列形式を **factor** から **character** に変更するときは、以下の操作を行ってください。

1. メインメニューから**オプション ▶ 設定**を選択し、**General Settings** ダイアログを開きます。
2. **Data** タブをクリックして、ダイアログの **Data** ページを表示させます。

3. **Data Options** グループの **Default Text Col.** ドロップダウンリストから **character** を選択し、**OK** をクリックします。

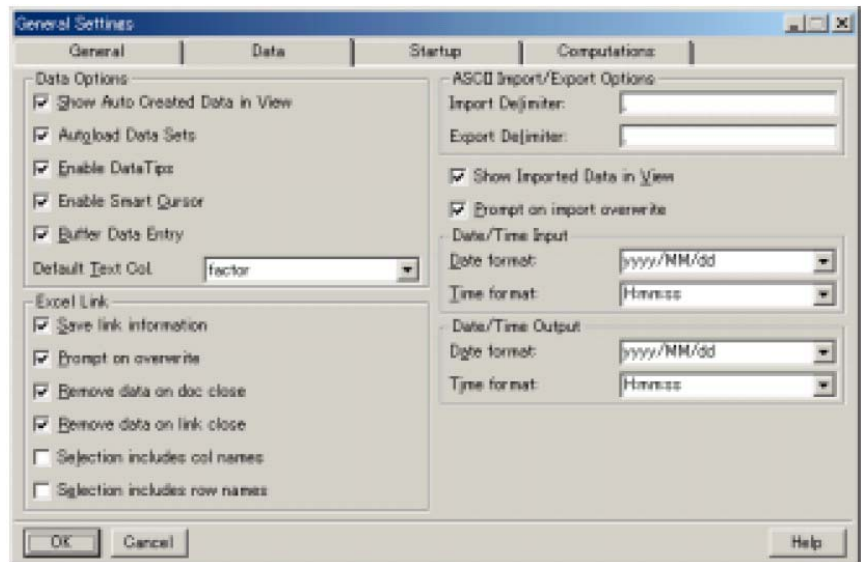


図 2.6 : 文字データのデフォルト列形式の変更

データ・ウィンドウに値を入力した後、その値を編集したい場合があります。セル内の値を編集するには、次の操作を行います。


1. 編集したい値を含むセルをクリックします。
2. **ENTER** を押して編集モードにするか、またはすぐに入力を始めて現在のデータを上書きします。

入力中の変更を無効にするには、**ESC** を押します。

操作を元に戻す

データ・ウィンドウで行った編集を“元に戻す”には、2 つのレベルがあります。最も新しい操作を元に戻したり、データセットを **S-PLUS** セッションの最初の状態に戻したりすることができます。

最も新しい操作を元に戻すときは、次のどちらかの操作を行います。

- **CTRL-Z** を押すか、標準ツールバー上の **元に戻す** ボタン  をクリックします。
- メインメニューから、**編集 ▶ 元に戻す** を選択します。

データセットを初期状態に戻すときは、以下の操作を行ってください。


1. 標準ツールバーの**データオブジェクトを元に戻す**ボタン  をクリックするか、メインメニューから**編集 ▶ データオブジェクトを元に戻す**を選択します。図 2.7 に示すように、**Restore Data Objects** ダイアログが開きます。



図 2.7 : Restore Data Objects ダイアログ

2. ダイアログに表示されたオブジェクトのリストからデータセットを選択します。
3. **Restore to Initial State** ラジオボタンをクリックして、**OK** をクリックします。

注意

また、**Restore Data Objects** ダイアログを使用して、1つ前の状態に戻す操作を行うことができます。データセットを選択し、**Restore to Previous State** ラジオボタンをクリックし、**OK** をクリックします。

元に戻す操作をやり直すときは、以上の手順のいずれかを実行してください。

データを保存する

S-PLUS は、専用の内部データベースにデータを保存することにより、データを自動的に保存します。“作業データ”と呼ばれるこのデータベースは、作成し修正したすべてのデータオブジェクトを、S-PLUS 言語で記述されたすべての関数同様、ユーザが意識することなく自動的に保存されるデータベースです。

オブジェクト・エクスプローラを使用すると、作業データに記憶されたすべてのオブジェクトを簡単に表示させることができます。作業データの詳しい説明とオブジェクト・エクスプローラの使い方は、第 9 章「オブジェクトおよびデータベースによる作業」を参照してください。

新しいデータオブジェクトと修正したデータオブジェクトの S-PLUS への保存をさらに厳密に制御したい場合は、変更を保存するか破棄するかを確認できるダイアログを表示させることができます。このダイアログは、S-PLUS セッションを終了するときに表示されます。

この基本設定を行うときは、以下のような操作を行ってください。

1. メインメニューから、**オプション ▶ 設定** を選択します。**General Settings** ダイアログが開きます。
2. このダイアログの **General** ページの **Prompts Closing Documents** グループで、**Show Commit Dialog on Exit** チェックボックスを選択し **OK** をクリックします。

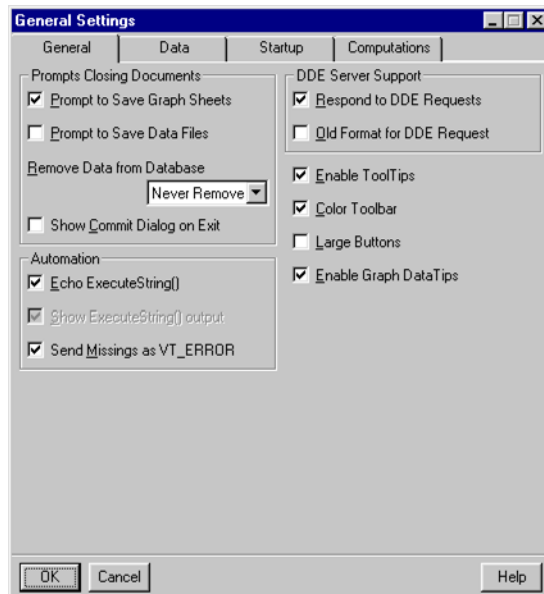


図 2.8 : General Settings ダイアログの General ページ

第2章 データによる作業

この基本設定を行うと、データオブジェクトを作成または修正したセッションを終了するときに、図 2.9 に示す**オブジェクト変更の保存**の保存ダイアログが自動的に開きます。

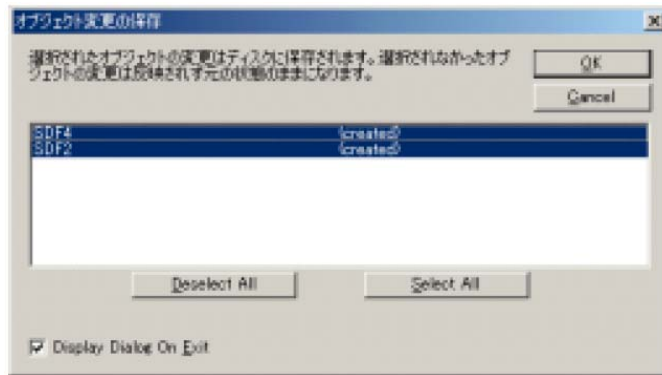


図 2.9 : オブジェクト変更の保存ダイアログ

デフォルトでは、そのセッションで作成または修正されたデータオブジェクトがすべて、**オブジェクト変更の保存**ダイアログで選択されています。リスト内のそれぞれのデータセットに以下の操作のいずれかを行い、**OK**をクリックします。

- 新規データセットまたは既存のデータセットの変更を保存するときは、その名前をハイライトさせたままにします。
- 新規データセットまたは既存のデータセットに行った変更を破棄するときは、その名前を CTRL-クリックして選択を解除します。

注意

General Settings ダイアログのこのオプションを設定した後で、**オブジェクト変更の保存**ダイアログボックス内の **Display Dialog on Exit** チェックボックスのチェックを外すことによって、後からその設定を無効にすることができます。

もちろん、**オブジェクト・エクスプローラ**を使用して、セッション中にいつでも作業データからデータオブジェクトを削除することができます。**オブジェクト・エクスプローラ**の詳しい使い方は、第9章を参照してください。

外部ファイルへ データを保存 する

前に説明したように、データセットを保存する最も簡単で効率的な方法は、そのデータセットを S-PLUS に保存させることです。作業データにデータオブジェクトを記憶することによって、オブジェクト・エクスプローラのすべての機能を使用できるようになります。

しかし、S-PLUS でも、他の標準の Windows 製品と同じように、ファイルメニューを使ってデータセットを外部ファイル (*.sdd) に保存することができます。そのようにデータを管理したい場合は、以下のような操作でいくつかのデフォルトのオプションを設定し直す必要がありますが、これはお勧めしません。

1. 前に説明したように、**General Settings** ダイアログの **General** ページを開きます。
2. **Prompts Closing Documents** グループで、以下の操作を行います。
 - **Prompt to Save Data Files** チェックボックスを選択します。
 - **Remove Data from Database** ドロップダウンリストで、**Always Remove Data** を選択します。
3. **OK** をクリックします。

これらの基本設定を行うと、新規データセットまたは修正されたデータセットが表示されたデータ・ウィンドウを閉じるときに、次の確認メッセージが表示されます。



ダイアログの**はい**をクリックすると、**Save Data Set As** ダイアログが開きます。ファイルにデータを保存するときは、データセットの名前をつけ、希望のフォルダに移動し、**保存**をクリックしてください。

データの表示とフォーマット

上記で説明したように、S-PLUS には、多数のサンプルデータセットが出荷時に含まれています。**Select Data** ダイアログを使用すると、そのようなサンプルデータセットならびに作業データに記憶された自分のデータセットのどれでも表示させることができます。

データセットを表示する

S-PLUS データベースに記憶されたデータセットを表示させるときは、次の操作を行ってください。

1. メインメニューから、**データ ▶ データの選択**を選択します。図 2.10 に示すように、**Select Data** ダイアログが開きます。



図 2.10 : Select Data ダイアログ

デフォルトでは、**Source** グループの **Existing Data** ラジオボタンが選択されています。

2. **Existing Data** グループの **Name** フィールドに、開きたいデータセットの名前を入力するか、ドロップダウンリストからその名前を選択して **OK** をクリックします。

ヒント

オブジェクト・エクスプローラ内のデータセットの名前をダブルクリックすることによって、そのデータセットを表示させることもできます。オブジェクト・エクスプローラの詳しい説明は、第9章「オブジェクトおよびデータベースによる作業」を参照してください。

データ・ウィンドウで最後に開いた（または、オブジェクト・エクスプローラで最後に選択した）データセットを、“カレント”データセットと呼びます。カレントデータセットを変更するときは、カレントデータセットにしたいデータのデータ・ウィンドウ内でクリックするか、ウィンドウメニューの下にあるリストから、そのデータセットを選択してください。操作対象のデータセット

が明示的に指定されないときは、カレントデータセットがデフォルトの操作対象になります。

データセットを 同時表示する

大きなデータセットの場合は、いくつかの異なるデータを別々のデータ・ウィンドウに表示させた方が都合のよいことがあります。

データセットの同時表示を行うときは、次の操作を行ってください。

1. **Select Data** ダイアログを使用してデータ・ウィンドウにデータを表示させます。
2. メインメニューから、**ウィンドウ ▶ 新規ウィンドウ**を選択します。

注意

元または複製のどのデータ・ウィンドウ内のデータも編集することができます。行った変更は、直ちにすべてのデータ・ウィンドウに反映されます。

元のデータ・ウィンドウのタイトルバーに表示されるデータセット名には、一時的に**:1** が付きます。2 番目のデータ・ウィンドウでは、名前に**:2** が付きます。この一時的命名規則は、追加のウィンドウが開かれるときに常に適用されます。ただし、複製したウィンドウを閉じると、データセットの元の名前が復元します。

データ・ ウィンドウを 操作する


S-PLUS には、データ・ウィンドウを簡単に表示させるのに便利なキーボードおよびマウス・ショートカットがいくつかあります。そのようなショートカットを下の表 2.1 に示します。

表 2.1 : データ・ウィンドウを表示させるキーボードとマウスのショートカット

機能	キーボード	マウス
画面を左に 1 列移動させる。	CTRL-左矢印	スクロールバーの左矢印をクリックする。
画面を右に 1 列移動させる。	CTRL-右矢印	スクロールバーの右矢印をクリックする。
第 1 列の第 1 行に移動する。	CTRL-HOME	スライダを上矢印と左矢印までドラッグし、セルをクリックする。
最終列の最終行に移動する。	CTRL-END	スライダを下矢印と右矢印までドラッグし、セルをクリックする。

第2章 データによる作業

表 2.1 : データ・ウィンドウを表示させるキーボードとマウスのショートカット (続き)

機能	キーボード	マウス
同じ行の第1列に移動する。	HOME または CTRL←	横スライダを左矢印までドラッグして、セルをクリックする。
同じ行の最終列に移動する。	END または CTRL→	横スライダを右矢印までドラッグして、セルをクリックする。
現在の列の第1行に移動する。	CTRL-PAGE UP	縦スライダを上矢印までドラッグし、セルをクリックする。
現在の列の最終行に移動する。	CTRL-PAGE DOWN	縦スライダを下矢印までドラッグして、セルをクリックする。
列を選択する。	CTRL-スペースバー	列見出しをクリックする。
行を選択する。	SHIFT-スペースバー	行見出しをクリックする。
データ・ウィンドウ全体を選択する。	CTRL-SHIFT-スペースバー または CTRL-A	データ・ウィンドウ左上角の一番上のセルをクリックする。
カーソルを選択モードにし、カーソルを移動させてブロック選択を行う。	SHIFT-矢印キー	マウスでセル上をクリックし、ドラッグする。
オンラインヘルプを表示する。	F1	標準ツールバーのヘルプボタン  をクリックし、次にデータ・ウィンドウ内でクリックする。
Go To Cell ダイアログを表示する。	F5	メインメニューから、表示 ▶ セルジャンプを選択する。
カーソルを編集モードにし、列名を編集する。	F9	列見出しの名前ボックスをダブルクリックする。

Go To Cell ダイアログを使用して、**データ**・ウィンドウ内の特定のセル位置に簡単にジャンプすることができます。

1. **F5** を押すか、メインメニューから**表示 ▶ セルジャンプ**を選択します。図 2.11 に示すように、**Go To Cell** ダイアログが開きます。

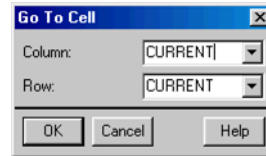


図 2.11 : **Go To Cell** ダイアログ

2. 列名を選択し、ジャンプしたいセルの行番号を入力します。最終列／最終行の位置に移動するときは、**Column** ドロップダウンリストと**Row** ドロップダウンリストから**END**を選択または入力します。
3. **OK** をクリックします。

Go To Cell ダイアログを使用して、セルの選択を拡張することもできます。アクティブなセルからダイアログで指定した位置まで選択を拡張するときは、**SHIFT** キーを押しながら **OK** をクリックしてください。たとえば、第 1 列第 5 行がアクティブなセルで、**Go To Cell** ダイアログで第 5 列第 5 行を指定し、**SHIFT-OK** を押すと、第 1 列第 5 行から第 5 列第 5 行まで選択範囲が拡張されます。

データ・ ウィンドウを カスタマイズ する

図 2.12 に示すように、**Data Frame** ダイアログを使用して、フォーマットの基本設定に合わせて**データ**・ウィンドウをカスタマイズすることができます。このダイアログを開くときは、以下の操作のいずれかを行います。

- **データ**・ウィンドウの左上角の灰色のセルをダブルクリックします。

第2章 データによる作業

- データ・ウィンドウをアクティブにし、メインメニューから書式 ▶ シートを選択します。

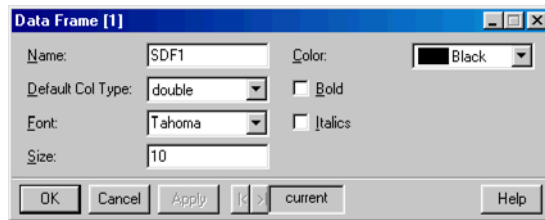


図 2.12 : Data Frame ダイアログ

このダイアログを使用すると、データセットの名前を変更したり、新しい列のデフォルト形式を変更したり、データ・ウィンドウのフォント、フォントサイズや他のフォーマット特性を指定したりすることができます。

デフォルト値を設定する

新しい空白のデータ・ウィンドウを開くとき、そのフォーマットには、一組のデフォルト設定が使用されます。たとえば、新しい列のデフォルト形式は、double の数値データです。Data Frame ダイアログを使用して、特定のフォーマット設定が反映されるようにデフォルト設定を変更することができます。

新しいデフォルト設定を行うときは、まず、データ・ウィンドウの Data Frame ダイアログで必要な変更を行い、次に **OK** をクリックし、変更を確定します。次に、以下の操作のいずれかを行ってください。

- メインメニューから、オプション ▶ ウィンドウサイズ・プロパティの今の状態をデフォルト値にするを選択します。
- データ・ウィンドウの左上角の灰色のセルを右クリックし、Save Data Frame as Default を選択します。

データを選択する

データをフォーマットまたは処理するときは、最初に処理するデータを選択します。1 つのセル、セルブロックまたは 1 つ（あるいは複数）の列（あるいは行）を選択することができます。最初にデータ・ウィンドウ内のデータを選択することにより、いくつかのメニューオプションの適用範囲を制限することもできます。

セルとブロック を選択する

1つのセルを選択するときは、選択したいセルをクリックします。

セルブロックを選択するときは、以下の操作のいずれかを行います。

- ブロック選択を始めたいセル内でマウスボタンを押したまま、カーソルをドラッグして、ハイライトされたブロックの大きさを変化させます。必要な領域がハイライトされたら、マウスボタンをはなします。
- ブロック選択を始めたいセル内をクリックし、次に、選択したいブロックを表す列と行の位置のセル内で **SHIFT** クリックします。

ヒント

SHIFT キーを押したまま矢印キーを押すことにより、セルの選択範囲を拡張することができます。

列と行を選択 する

データ・ウィンドウ内のすべてのセルを選択するときは、データ・ウィンドウの左上角の空白の灰色部分をクリックしてください。

1つの列／行を選択するときは、列見出し／行見出しをクリックします。

連続する列／行のブロックを選択するときは、以下の操作のいずれかを行います。

- 選択を始める最初の列／行の、列見出し／行見出しをクリックし、次に、選択したいブロックを示す最後の列／行の、列見出し／行見出しを **SHIFT** クリックします。
- ブロック選択を始めたい最初の列／行の、列見出し／行見出しでマウスボタンを押したまま、選択したい列／行の端までカーソルをドラッグし、マウスボタンをはなします。

連続していない列／行のグループを選択するとき、または一群の列／行を特定の順序で選択するときは、次の操作を行います。

- 選択したいそれぞれの列／行の見出しを、選択したい順序で **CTRL** クリックします。

注意事項

CTRL-クリック選択の大きな特徴は、選択プロセスに順序がつけられることです。これと対照的に、カーソルをドラッグするか SHIFT-クリックを使用すると、選択操作自体が実際にどのように行われたとしても、選択の順序はデフォルトとして、列の場合は左から右に、行の場合は上から下になるように解釈されます。したがって、これらの方法でデータを選択するときは、次の点に注意してください。

- 連続していない列または行を、選択順序を考慮して選択しなければならないときは、CTRL-クリックを使用してください。
- 列/行が連続している場合でも、列/行のグループを特定の順序で選択したいときは、CTRL-クリックを使用してください。
- 左から右または上から下の順序で選択したい場合には、カーソルをドラッグするか SHIFT-クリックして連続した列または行のブロックを選択することができます。

列を フォーマット する

データセットの列は、通常、所定の変量のデータからなる縦グループです。S-PLUS は、列指向であるため、フォーマットおよびデータ操作ツールは、列を単位として機能します。

S-PLUS は、データセットの各列に自動的に番号を付けます。その列番号は、列見出しに表示され、データ・ウィンドウ内の列の位置を示します。

列名を変更する

空白の列にデータ値を入力すると、S-PLUS は、その列に自動的にデフォルト名 (Vx、x は連続番号) を付け、その名前が、列番号の下の見出しに表示されます。デフォルト名を使って列を参照することもできますが、わかりやすい名前に付け直すことをお勧めします。

列に名前をつけるためのヒント

- 列名は、データセット内に同じものがあってはいけません。
- 列名は、文字で始まらなければなりません。文字、数字およびピリオドの任意の組合せを含むことができます。ただし、“É” などの拡張 ASCII 文字を含めることはできません。
- S-PLUS 関数名や他の予約語は、列名に使用することはできません。

列は名前でも番号でも参照することができますが、いくつかの操作では列の番号がつけ直されるため、列は名前で参照した方がよい場合があります。たとえば、第5列と第6列の間に列を挿入すると、第5列の右側の列はすべて番号がつけ直されます。参照に列の番号を使用している場合は、その後の操作で必ず新しい番号を使用してください。

列名を変更するときは、以下の操作を行ってください。

1. 列見出しの名前ボックスをダブルクリックするか、列内の任意のセルをアクティブにした状態で、**F9** を押します。
2. 新しい列名を入力するか、既存の名前を編集します。
3. **ENTER** を押すか、**データ・ウィンドウ**内の他の場所をクリックし、変更を確定します。

列名のプロパティダイアログを使って列名を変更するときは、以下の操作を行います。

1. 列見出しの番号ボックスをダブルクリックするか、列をクリックし、メインメニューから**書式 ▶ 選択したオブジェクト**を選択します。図 2.13 に示すように、列プロパティダイアログが開きます。

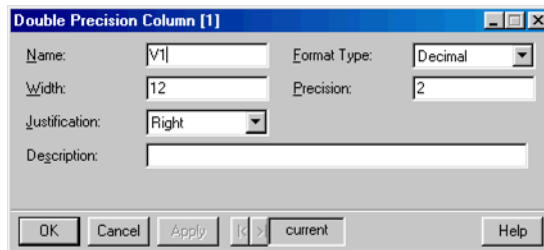


図 2.13 : Double Precision Column ダイアログ

2. **Name** テキストボックスに、新しい列名を入力するか、または既存の名前を編集し **OK** をクリックします。

注意

ダイアログのタイトルバーに表示されているプロパティダイアログの名前は、ダイアログを開くときに選択したオブジェクトの型によって決まります。たとえば、倍精度の列には、**Double Precision Column** ダイアログが開き、文字の列には、**Character Column** ダイアログが開きます。

列説明の追加と編集

列には、番号と名前の他に説明を加えることもできます。列説明を指定すると、その説明が、グラフのデフォルトの軸タイトルと凡例テキストとして使用されます。説明を指定しない場合は、代わりに列名が使用されます。

列説明を指定するためのヒント

- 列説明は、75文字までです。(全角かなの場合37文字まで)
- 列説明は、文字、数字、記号およびスペースの任意の組合せが可能です。

列説明を追加または編集するときは、以下のように行います。

- 35 ページで説明したように列プロパティダイアログを開きます。
Description テキストボックスに、新しい列説明を入力するか、既存の説明を編集し、**OK** をクリックします。

列見出しの名前ボックス上でマウスカーソルを止めると、図 2.14 に示すように、データタイプに列説明が表示されます。

	1	2	3	4	5	6
	Weight	Disp.	Mileage	Fuel	Type	
1	Eagle Summit 4	2560.00	97.00	33.0	3.03	Small
2	Ford Escort 4	2345.00	114.00	33.00	2.70	Small
3	Ford Festiva 4	1845.00	81.00	37.00	2.12	Small
4	Honda Civic 4	2250.00	91.00	32.00	2.12	Small

図 2.14 : データタイプに列説明が表示される

列リストを作成する

列リストは、操作する列のグループまたは順序を指定するためのダイアログフィールド内で使われる列名または列番号のリストです。たとえば、列名 **Weight** と **Type** を選択すると、列リスト **Weight** と **Type** が作成されます。

列リストを作成するときは、ダイアログフィールドのドロップダウンリストから列名 (必要に応じて **CTRL**-クリックを使用) を選択します。

注意

ダイアログフィールドには、列番号ではなく列名だけが表示されます。

また、カンマで区切った列番号を入力することにより、ダイアログフィールド内に列リストを作成することができます。たとえば、**1,3,4** は、第 1 列、第 3 列および第 4 列を示します。連続する列を指定するときは、最初と最後の列名または番号をコロンで区切って入力してください。たとえば、**3:7** は、第 3 列から第 7 列までを示します。データセット内のすべての列を指定するときは、**<ALL>** を選択してください。

列幅を変更する

目で確かめながら列の幅を広くしたり狭くしたりするときは、カーソルをドラッグするか、ツールバーボタンを使用します。



ドラッグして列幅を変更するときは、次の操作を行います。

1. 列見出し右側の縦線の上にカーソルを置きます。マウスポインタが、サイズ変更ツールになります。


1	2
Weight	Disp.
2560.00	97.00
2345.00	114.00

2. 列の幅を狭くするときはサイズ変更ツールを左にドラッグし、幅を広くするときは右にドラッグします。

ツールバーのボタンを使って列幅を変更するときは、次の操作を行います。

1. 列をクリックします。
2. データ・ウィンドウ・ツールバーの列の巾の増加ボタン  または列の巾の減少ボタン  をクリックします。クリックするたびに 1 文字ずつ列の幅が増減します。

列内の最も幅の広いセルに合わせて列幅を調整するときは、次の操作を行います。

1. 列をクリックします。
2. データ・ウィンドウ・ツールバーの文字巾で調整ボタン  をクリックします。


正確な列幅を設定する必要がある場合は、列プロパティダイアログを開き、デフォルトのフォントおよびポイントサイズの文字の数で希望の幅を指定してください。

データ形式を変更する

入力することができるデータの形式は、列全体のデータ形式によって決まります。たとえば、文字型の列には文字データだけを入力することができ、整数型の列には整数データだけを入力することができます。

S-PLUS のデータ形式は、character、complex、double、factor、integer、logical、single および timeDate です。最も一般的に使用される2つのデータ形式は、double（浮動小数点倍精度実数）と factor（カテゴリカルデータ）です。S-PLUS のデータ形式の詳細は、『Programmer's guide』を参照してください。

列のデータ形式を変更するときは、次の操作を行ってください。

1. 列をクリックし、次に**データ・ウィンドウ・ツールバーの列形式の変更ボタン**  をクリックするか、メイン・メニューから**データ ▶ 列形式の変更**を選択します。図 2.15 に示すように、**Change Data Type** ダイアログが開きます。

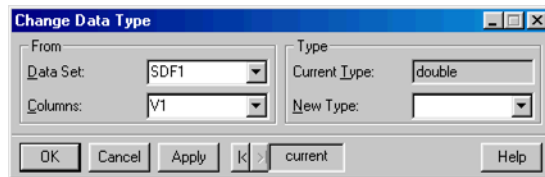


図 2.15 : Change Data Type ダイアログ

2. **Type** グループの **New Type** ドロップダウンリストから新しいデータタイプを選択し、**OK** をクリックします。

列見出しの番号ボックスの上でマウスカーソルを止めると、図 2.16 に示すように、データタイプに列形式が表示されます。

		1	2	3	4	5	6
		Weight	Disp.	Mileage	Fuel	Type	factor
1	Eagle Summit 4	2560.00	97.00	33.00	3.03	Small	
2	Ford Escort 4	2345.00	114.00	33.00	3.03	Small	
3	Ford Festiva 4	1845.00	81.00	37.00	2.70	Small	

図 2.16 : データタイプに列形式が表示される

フォーマット タイプを変更 する

S-PLUS は、数値列 (double 列) に、**Mixed**、**Number**、**Decimal**、**Scientific**、**Currency**、**Financial**、**Date**、**Date&Time**、**Time**、**Elapsed_H:M:S** などの標準的な Windows フォーマットタイプをあてはめることができます。



列のフォーマットタイプを変更するときは、以下の操作を行ってください。

- 35 ページで説明したように、列プロパティダイアログを開きます。**Format Type** ドロップダウンリストから別のフォーマットタイプを選択し、**OK** をクリックします。

表示精度を変更 する

列の表示精度は、数字の表示方法だけを変化させます。内部の計算には影響を及ぼさず、常に倍精度を維持します。

列の表示精度を変更するときは、以下の操作のいずれかを行います。

- 表示精度を高くするか低くするときは、列をクリックし、次に**データ・ウィンドウ・ツールバー**にある**小数点表示桁上げボタン**  または**小数点表示桁下げボタン**  をそれぞれクリックします。
- 35 ページで説明したように、列プロパティダイアログを開きます。**Precision** テキストボックスに、小数点の後に表示させたい桁数を入力し (最大 17 まで)、**OK** をクリックします。

デフォルト値を 設定する

フォーマット基本設定を反映させることで、行揃え、精度、幅などの列のデフォルト設定を変更することができます。たとえば、文字の列を数値の列と異なるデフォルト幅にした方がよい場合があります。

列デフォルト値を設定するときは、以下の操作を行ってください。

1. 35 ページで説明したように、列プロパティダイアログを開きます。
2. 新しいデフォルト設定として保存したい変更を行い、**OK** をクリックします。
3. 列を右クリックし、ショートカットメニューから **Save [Column Type] Column as default** を選択します。

行を フォーマット する

S-PLUS は、データセットの各行に自動的に番号を付けます。行番号は、行見出しに表示され、**データ・ウィンドウ**における行の位置を示します。S-PLUS は、列指向であるため、ほとんどのフォーマットオプションは、列だけに適用されます。しかし、行に名前を付けることもできます。

行名の追加と 変更

行名は、行番号の右側の見出しに表示されます。

行名を追加または変更するときは、以下の操作を行ってください。

1. 行見出しの名前ボックスをダブルクリックします。
2. 行名を入力するか、既存の名前を編集します。
3. ENTER を押すか、**データ**・ウィンドウ内の他の場所をクリックし、変更を確定します。

行リストを作成 する

行リストは、操作する行のグループまたは順序を指定するためのダイアログフィールド内で使われる行番号のリストです。行リストを作成するときは、カンマで区切って行番号を入力してください。たとえば、**1,3,4** は、第1行、第3行および第4行を表します。連続する行を指定するときは、最初と最後の行番号をコロンで区切って入力してください。たとえば、**3:7** は、第3行から第7行までを表します。データセット内のすべての行を指定するときは、**<ALL>**を入力してください。

データの操作

S-PLUS は、様々なデータ操作ツールを提供します。データ・ウィンドウ・ツールバーのボタンは、通常の作業を行うのに便利ですが、データメニューでさらに多くのオプションを利用することができます。

データの移動とコピー

後で説明するような様々な方法を使って、1 つのデータ・ウィンドウ内または異なるデータ・ウィンドウ間でデータを移動したりコピーしたりすることができます。

セルおよびブロックの移動とコピー

セル／セルブロックをドラッグで移動あるいはコピーするときは、以下の操作を行ってください。

1. 移動あるいはコピーしたいセル／セルブロックを選択します。
2. 選択したセル／ブロック内にカーソルを置きます。図 2.17 に示すように、カーソルが矢印になります。

	1	2	3	4	5
	NOx	C	E		
1	3.74	12.00	0.91		
2	2.29	12.00	0.76		
3	1.50	12.00	1.11		
4	2.88	12.00	1.02		
5	0.76	12.00	1.19		
6	3.12	9.00	1.00		
7	0.64	9.00	1.23		

図 2.17 : データ・ウィンドウ内のセルブロックの選択

3. 選択したセル／ブロックを新しい位置にドラッグします。セル／ブロックを移動するときは、マウスボタンをはなしてください。セル／ブロックをコピーするときは、CTRL キーを押したままマウスボタンをはなしてください。図 2.18 を参照してください。

注意

既にデータを含む目標位置にデータを移動またはコピーすると、既存のデータが上書きされます。また、セル／ブロックを移動すると、移動元の空白のセルに、欠損値を示す **NA** が入ることに注意してください。

第2章 データによる作業

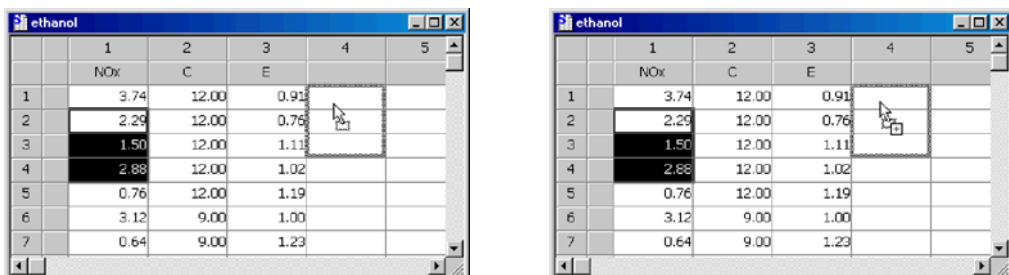





図 2.18 : データ・ウィンドウにおけるセルブロックの移動（左上）とコピー（右上）

ヒント

データ・ウィンドウ間でドラッグ アンド ドロップを使ってデータを移動またはコピーするときは、移動元のセルと移動先のセルの両方の位置が見えるようにウィンドウを調整してください。

切り取り、コピーあるいは貼り付けを使ってセル／セルブロックをコピーするときは、以下の操作を行ってください。

1. 移動あるいはコピーしたいセル／セルブロックを選択します。
2. 以下の操作のいずれかを行ってください。
 - セルまたはブロックを移動させるときは、CTRL-X を押すか、標準ツールバーの切り取りボタン  をクリックするか、編集またはショートカットメニューから切り取り (Cut) を選択します。
 - セルまたはブロックをコピーするときは、CTRL-C を押すか、標準ツールバーのコピーボタン  をクリックするか、編集またはショートカット・メニューからコピー (Copy) を選択します。
3. データ・ウィンドウ内の新しい位置でマウスをクリックします。
4. CTRL-V を押すか、標準ツールバーの貼り付けボタン  をクリックするか、編集またはショートカット・メニューから貼り付け (Paste) を選択します。

データメニューを使ってセル／セルブロックを移動あるいはコピーするときは、以下の操作を行ってください。

1. メインメニューから、セル／ブロックを移動するときは**データ ▶ 移動 ▶ ブロック**を選択し、セル／ブロックをコピーするときは**データ ▶ コピー ▶ ブロック**を選択します。選択に応じて、図 2.19 に示すように、**Move Block** ダイアログまたは **Copy Block** ダイアログが開きます。

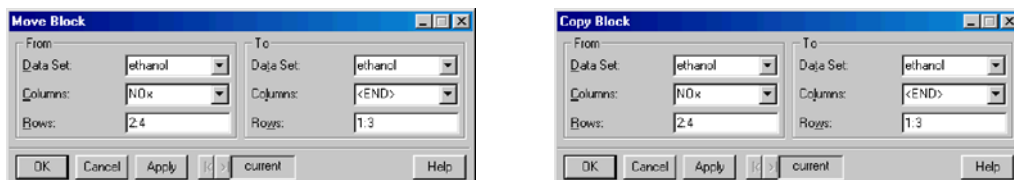


図 2.19 : Move Block ダイアログと Copy Block ダイアログ

2. **From** グループの **Columns** および **Rows** フィールドで、移動あるいはコピーしたいセル／セルブロックを、列と行の位置で指定します。
3. **To** グループの **Columns** および **Rows** フィールドで、移動先を列と行の位置で指定し、**OK** をクリックします。

ヒント

セル／ブロックを別のデータセットに移動あるいはコピーするときは、**To** グループの **Data Set** ドロップダウンリストからその名前を選択します。ターゲット・データセットを新たに作成するときは、このフィールドに新しい名前を入力します。

列および行の移動とコピー

列／行の移動あるいはコピーの手順は、前に述べたようなセル／ブロックの移動あるいはコピーの手順と同じです。ただし、以下の点に注意してください。

列／行をドラッグで移動あるいはコピーするときは、以下の操作を行ってください。

- 列／行をドラッグするときは、選択したセル内の列見出し／行見出し以外の場所にカーソルを置きます。
- S-PLUS は、名前を含む列／行全体を 1 つの単位として移動あるいはコピーします。コピーした列と行の名前には、.1 が付けられます。
- 既にデータを含む位置にデータを移動あるいはコピーすると、既存のデータが上書きされます。

第2章 データによる作業

切り取り、コピーあるいは貼り付けを使って列／行を移動あるいはコピーするときは、以下の点に注意してください。

- S-PLUS は、列／行内のデータ値だけを新しい場所に移動あるいはコピーします。
- 既にデータを含む位置にデータを移動あるいはコピーすると、既存のデータが上書きされます。

図 2.20 に示すように、列／行を移動あるいはコピーするためのデータメニューのダイアログは、セル／ブロック用のものとよく似ています。

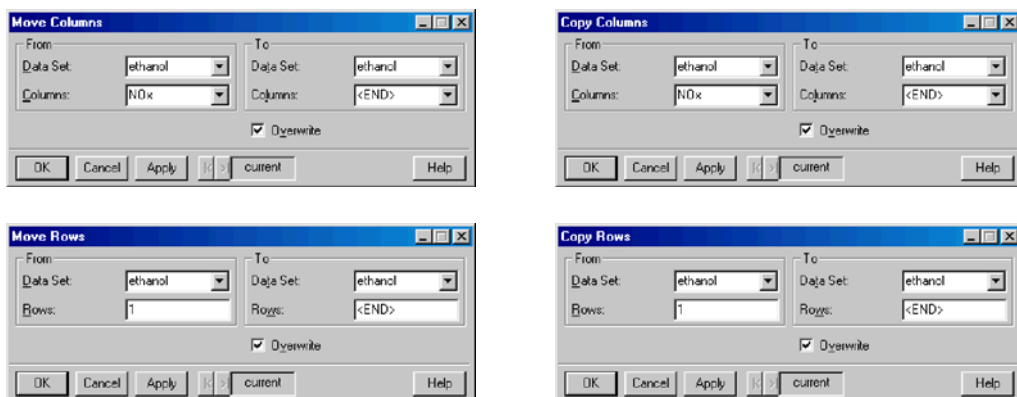


図 2.20 : Move Columns、Copy Columns、Move Rows、Copy Rows ダイアログ

データメニューを使って列／行を移動あるいはコピーするときは、以下の点に注意してください。

- S-PLUS は、名前を含む列／行全体を 1 つの単位として移動あるいはコピーします。コピーした列と行の名前には、.1 が付けられます。
- 既にデータを含む位置にデータを移動あるいはコピーすると、デフォルトでは、既存のデータが上書きされます。ただし、ダイアログの下にある **Overwright** チェックボックスのチェックを外すことにより、既存のデータの上書きを防ぐことができます。このチェックボックスが選択されていない場合、S-PLUS は、移動あるいはコピーするデータの場所を作るために、既存の列を右にずらすか、既存の行を下にずらします。

ヒント

Copy Columns ダイアログを使用して、**From** または **To** グループのどちらかの **Columns** ドロップダウンリストから専用キーワード<**ROWNAMES**>を選択するだけで、**データ・ウィンドウ**内にある列を行名（灰色）にコピーしたり、行名を列にコピーしたりすることができます。

データを挿入する

データ・ウィンドウ内でセル、ブロック、列あるいは行を挿入すると、新しいセルの場所を作るために既存のセルが下または右にずれます。

セルとブロックを挿入する

セルやセルブロックを挿入するときは、以下の操作を行います。

- メインメニューから、**挿入 ▶ ブロック** を選択します。図 2.21 に示した **Insert Block** ダイアログが開きます。

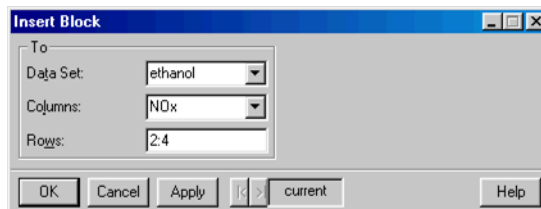


図 2.21 : Insert Block ダイアログ

Columns および **Rows** フィールドで、挿入したいセルやセルブロックを、列と行の位置で指定し、**OK** をクリックします。

列を挿入する

列を挿入するときは、以下の操作のいずれかを行います。


- 新しい列の場所を作るために、右にずらしたい列をクリックします。デフォルト形式または前回最後に挿入した列と同じ形式の新しい列を挿入するときは、**データ・ウィンドウ・ツールバー**の**列の挿入**ボタン をクリックします。特定の形式の新しい列を挿入するときは、**列の挿入**ボタン（図 2.22 を参照）の右側にある列形式を選択する矢印をクリックし、挿入したい列の形式を選択します。



図 2.22 : 特定形式の列の挿入

- メインメニューから、**挿入 ▶ 列**を選択します。図 2.23 に示すように、**Insert Columns** ダイアログが開きます。

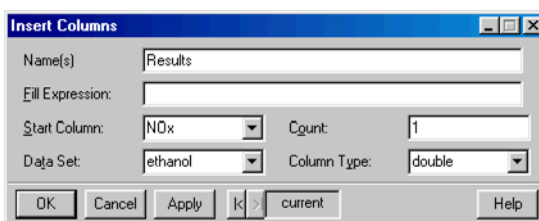


図 2.23 : Insert Columns ダイアログ


新しい列の場所を空けるために右にずらしたい列を、**Start Column** ドロップダウンリストから選択します。**Name(s)**テキストボックスに、新しい列の名前を入力し、**OK** をクリックします。

ヒント

Insert Columns ダイアログを使って複数の列を挿入することができます。挿入したい列の数を **Count** テキストボックスに入力し、カンマで区切った名前を **Name(s)** テキストボックスに入力します。

行を挿入する

行を挿入するときは、以下の操作のいずれかを行ってください。

- 新しい行の場所を作るためにずらしたい行をクリックし、次に、データ・ウィンドウ・ツールバーの**行の挿入**ボタン  をクリックします。
- メインメニューから**挿入 ▶ 行**を選択します。図 2.24 に示すように、**Insert Rows** ダイアログが開きます。

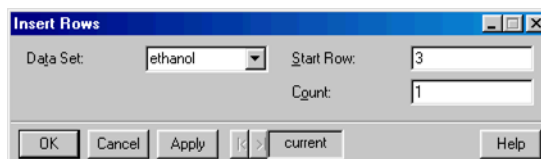


図 2.24 : Insert Rows ダイアログ

Start Row テキストボックスに、新しい行の場所を作るためにずらしたい行の行番号を入力し、**OK** をクリックします。

ヒント

Insert Rows ダイアログを使って複数の行を挿入することもできます。**Count** テキストボックスに挿入したい行の番号を入力するだけです。

データを削除する

データ・ウィンドウ内のデータを削除するときは、セルを残したままデータ値をクリアすることも、セルとその内容の両方を削除してデータセットのサイズを小さくすることもできます。データをクリアすると、その値が、欠損値を示す NA と置き換えられることに注意してください。

注意

DELETE キーを押すか、**編集**またはショートカット・メニューから**クリア (Clear)** を選択してセル、ブロック、列または行をクリアすると、データは、クリップボードに入りません。データをクリアしてそのデータをクリップボードに入れたいときは、**切り取り (Cut)** を選択してください。

セルおよび ブロックの クリアと削除

セルやセルブロックを消去するときは、以下の操作のいずれかを行います。

- セルやセルブロックを選択し、**編集**またはショートカット・メニューから**クリア (Clear)** を選択します。
- メインメニューから、**データ ▶ クリア ▶ ブロック**を選択します。図 2.25 に示すように、**Clear Block** ダイアログが開きます。

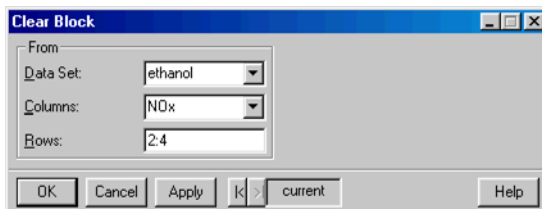


図 2.25 : Clear Block ダイアログ

Columns と **Rows** フィールドで、クリアしたいセルやセルブロックの列と行の位置で指定し、**OK** をクリックします。

ヒント

データ・ウィンドウ内のすべてのデータをクリアするときは、データ・ウィンドウの左上角の空白の灰色部分をクリックしてデータセット内のすべてのデータを選択し、**編集**またはショートカット・メニューから**クリア (Clear)** を選択します。

セルやセルブロックを削除するときは、以下の操作のいずれかを行います。

- セルやセルブロックを選択してから、**DELETE** キーを押すか、**編集**またはショートカット・メニューから**切り取り (Cut)** を選択します。
- メインメニューから、**データ ▶ 削除 ▶ ブロック**を選択します。図 2.26 に示すように、**Remove Block** ダイアログが開きます。

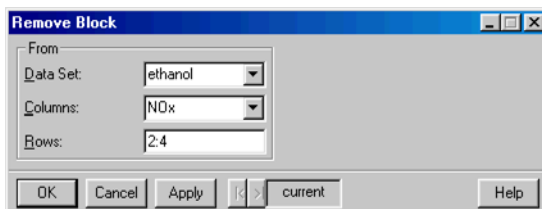



図 2.26 : Remove Block ダイアログ

Columns および **Rows** フィールドで、削除したいセルやセルブロックの列と行の位置を指定し、**OK** をクリックします。

列のクリアと削除

列をクリアすると、その列内のデータは削除されますが、その列の位置、名前、フォーマット情報はそのまま残ります。

列をクリアするときは、以下の操作のいずれかを行います。

- 列をクリックし、次にデータ・ウィンドウ・ツールバーの**列のクリア** ボタン  をクリックします。
- 列を選択し、**編集**またはショートカット・メニューから**クリア (Clear)** を選択します。
- メインメニューから、**データ ▶ クリア ▶ 列**を選択します。図 2.27 に示すように、**Clear Columns** ダイアログが開きます。

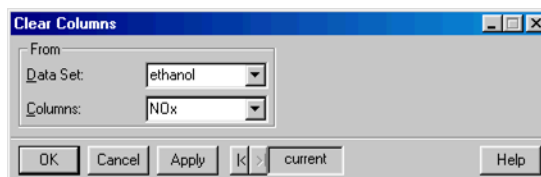


図 2.27 : Clear Columns ダイアログ

Columns ドロップダウン・リストからクリアしたい列を選択し、**OK** をクリックします。

列を削除すると、列全体が削除され、データセットのサイズが小さくなります。

列を削除するときは、以下の操作のいずれかを行ってください。


- 列をクリックし、次にデータ・ウィンドウ・ツールバーの**列の削除** ボタン  をクリックします。
- 列を選択し、次に **DELETE** キーを押すか、**編集**またはショートカット・メニューから**切り取り (Cut)** を選択します。
- メインメニューから、**データ ▶ 削除 ▶ 列**を選択します。図 2.28 に示すように、**Remove Columns** ダイアログが開きます。



図 2.28 : Remove Columns ダイアログ

Columns ドロップダウン・リストから削除したい列を選択し、**OK** をクリックします。

行のクリアと削除

行をクリアすると行内のデータが削除されますが、行の位置と名前（名前がある場合）はそのまま残ります。

行をクリアするときは、以下の操作のいずれかを行ってください。



- 行をクリックし、次にデータ・ウィンドウ・ツールバーの行のクリアボタン  をクリックします。
- 行を選択し、編集またはショートカット・メニューからクリア (**Clear**) を選択します。
- メインメニューから、データ ▶ クリア ▶ 行を選択します。図 2.29 に示すように、**Clear Rows** ダイアログが開きます。



図 2.29 : Clear Rows ダイアログ

クリアしたい行の行番号を **Rows** テキストボックスに入力し、**OK** をクリックします。

行を削除すると行全体が削除され、データセットのサイズが小さくなります。
行を削除するときは、以下の操作のいずれかを行ってください。

- 行をクリックし、次にデータ・ウィンドウ・ツールバーの行の削除ボタン  をクリックします。
- 行を選択し、次に DELETE キーを押すか、編集またはショートカット・メニューから切り取り (Cut) を選択します。
- メインメニューから、データ ▶ 削除 ▶ 行を選択します。図 2.30 に示すように、Remove Rows ダイアログが開きます。

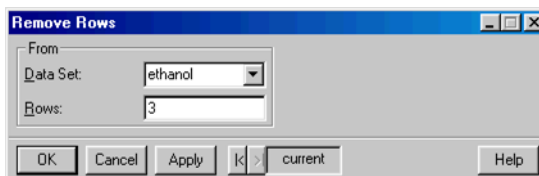


図 2.30 : Remove Rows ダイアログ



削除したい行の行番号を **Rows** テキストボックスに入力し、**OK** をクリックします。

データを ソートする

S-PLUS は、データセット全体を簡単にソートできるツールバー・ボタンと、ソートパラメータをカスタマイズすることができるダイアログを提供します。

クイックソート

所定のデータセット内のすべての列を、アクティブセルを含む列で素早くソートしたいときは、以下の操作を行ってください。

- ソートしたい列をクリックし、次に、データ・ウィンドウ・ツールバーの昇順で並べ替えボタン  または降順で並べ替えボタン  をクリックします。

ソートの カスタマイズ

ソートパラメータを指定しやすくするために、**データ**メニューから **Sort Columns** ダイアログを使用します。このダイアログで、次のことを行うことができます。

- データセット全体をソートするか、列の一部分だけをソートするかを指定します。
- ソートする複数の列を選択します。ソートする複数の列を指定するときは、まず、選択した最初の列によってデータに順序が付けられます。最初の列で値が等しいデータの場合は、次に選択した列で順序が決定され、それ以後同じように順序がつけられます。
- 元のデータに上書きしたくない場合は、ソートの結果を保存するための別のデータセットまたは列を指定します。

ソートをカスタマイズするときは、以下の操作を行います。

1. メインメニューから、**データ ▶ 構築 ▶ ソート**を選択します。図 2.31 に示すように、**Sort Columns** ダイアログが開きます。

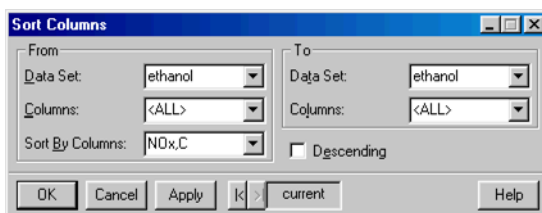


図 2.31 : Sort Columns ダイアログ

2. **From** グループの **Columns** ドロップダウンリストからソートしたい列を選択します。データセット内のすべての列をソートするときは、専用キーワード **<ALL>** を選択します。
3. **Sort By Columns** ドロップダウンリストからソートキーとする 1 つまたは複数の列を選択します。複数の列をソートキーとするときは、**CTRL**-クリックで列を希望の順序で選択してください。

4. 次のように、**To** グループでソート結果のターゲットデータセットを指定します。
 - 適切にソートするために、**Data Set** ドロップダウンリストと **Columns** ドロップダウンリストからそれぞれ、対応する **From** グループで選択したものと同じデータセットと列を選択してください。

注意

Columns フィールドで<ALL>よりも少ない列が選択され、結果を上書きにすると、ソートするとき列が一致しなくなることがあります。

- ソート結果を別のデータセットに送るときは、**Data Set** ドロップダウンリストからそのデータセットを選択し（または、このフィールドに新しい名前を入力してデータセットを作成し）、**Columns** ドロップダウンリストから希望の列を選択してください。

注意

To グループで選択した列の数は、**From** グループで選択した列の数と一致しなければなりません。また、ターゲット列の既存のデータが上書きされることに注意してください。

5. デフォルトでは、列は、昇順にソートされます。降順でソートするときは、**Descending** チェックボックスを選択してください。
6. **OK** をクリックします。

その他の データ操作 オプション

以上で説明した基本ツールの他に、**データメニュー**には、多数の便利なデータ操作オプションがあります。次に、まだ説明していないオプションについて簡単に説明します。**乱数**、**分布関数**、**テーブル**、および**ランダムサンプリングツール**については、第 8 章を参照してください。データ操作ダイアログの使い方の詳細は、「オンラインヘルプ」を参照してください。

転置

Transpose Columns ダイアログと **Transpose Rows** ダイアログを使用して、列を行に変換したり行を列に変換したりすることができます。**Transpose Block** ダイアログを使用して、テキストのブロックを転置させる（すなわち、ブロックを横に回転させる）ことができます。

交換

Exchange Columns ダイアログと **Exchange Rows** ダイアログを使用して、異なるデータセット間で、列または行の位置を交換することができます。

構築

追加

Append Columns ダイアログを使用して、データの列を他の列の下に追加することができます。

パック

Pack Columns ダイアログを使用して、列内の欠損値を削除し、残りの値をずらしてスペースを詰めることができます。

スタック

Stack Columns ダイアログを使用して、複数列のデータを別の列に積み重ね、各値を適切に複製しながら1つの列とすることができます。

スタックの回復

Unstack Columns ダイアログを使用して、1つの列を、指定した長さのいくつかの列に分割することができます。

埋める

Fill Numeric Columns ダイアログを使用して、データセットの列を、NAまたは生成した一連の数字で埋めることができます。

置換

Recode ダイアログを使用して、指定した列内に生じた特定のすべての値を、指定した新しい値に置き換えることができます。

新しい列を 変換式で作成

Transform ダイアログを使用して、他の変数の変換式に基づいて新しい変数を作成することができます。

数値を因子に 変換

Create Categories ダイアログを使用して、数値変量から新しいカテゴリカルな変数を作成したり、因子水準の名前を変えたり結合することによって、既存の因子水準の変量を再定義することができます。

乱数

Random Numbers ダイアログを使用して、指定した分布から乱数を生成することができます。

分布関数

Distribution Functions ダイアログを使用して、指定した分布から確率密度、累積確率、および確率点を計算することができます。

行の分割

Split Data by Group ダイアログを使用して、分割変数の値に基づいて、データセットを複数の新しいデータセットに分割することができます。

- 行の部分集合** **Subset** ダイアログを使用して、条件抽出式に基づいてデータセットの部分集合を作成することができます。このダイアログには、式の作成を支援をするツールがありますが、S 言語についてある程度の知識が必要です。
- マージ** **Merge Two Data Sets** ダイアログを使用して、2 つのデータセットを 1 つのデータセットに結合することができます。
- テーブル** **Tabulate** ダイアログを使用して、データセットからデータのテーブルサマリーを作成することができます。選択したデータセットの列が変量として識別され、変量値の各組合せの件数が返されます。数値変量を分類した後でカウントを行うことができます。
- 件数のテーブルは、印刷することができ、複数パネルの条件付けプロットに適したデータセットで返すこともできます。統計やその他のサマリー情報については、**統計 ▶ データサマリー ▶ クロス表** を選択してください。
- 組み合わせ
データフレーム
を作成** **Expand Grid** ダイアログを使用して、既存のデータセットの値のすべての組合せを含む新規データセットを作成することができます。各組の値は、列内がすべて同一の値でも、列内の一連の値をカバーする指定された数値でもかまいません。このダイアログは、関数を評価したりモデルから予測したりするために、値の格子を表す列を生成するのに役立ちます。
- ランダムサンプリング** **Random Sample of Rows** ダイアログを使用して、ランダムサンプルを生成したり、データセット内の観測値を入れ替えたりすることができます。

S-PLUS に含まれるライブラリ

S-PLUS 内のデータセットはすべて、ライブラリに記憶されています。ただし、本書で「S-PLUS」について説明するときは、通常、起動時に自動的にロードされるライブラリ内のプログラムおよびオブジェクトのことを指します。しかし、S-PLUS には、そのようなコアとなるライブラリよりも多くのライブラリが含まれています。表 2.2 は、S-PLUS に標準で付属する追加ライブラリのリストです。

表 2.2 : S-PLUS に含まれている追加ライブラリ

名前	説明
chron	日付と時間を処理する関数
class	W.N. Venables and B.D. Ripley による『S-PLUS による統計解析』からの例
Defunct	S-PLUS でサポートされなくなった関数
design	Frank Harrell による実験計画法の例
examples	『The New S Language』からの例
example5	S-PLUS 5.x 以降のための例
hmisc	Frank Harrell からの有効な例
maps	投影によるマップの表示
Mass	W.N. Venables and B.D. Ripley による『S-PLUS による統計解析』からの例
missing	欠損データに対するモデルベース法と多重代入法
nlme2	古い混合効果モデル関数

表 2.2 : S-PLUS に含まれている追加ライブラリ (続き)

名前	説明
Nnet	W.N. Venables and B.D. Ripley による『S-PLUS による統計解析』からのニューラルネットの例
robust	最先端のロバストモデル近似と外れ値検出
spatial	W.N. Venables and B.D. Ripley による『S-PLUS による統計解析』からの空間解析

これらのライブラリはすべて、メインメニューから**ファイル ▶ ライブラリ**の**ロード**を選択するか、コマンド・ウィンドウからの `library` 関数を使用することによってロードすることができます (第 10 章を参照)。robust ライブラリや Frank Harrell と Brian Ripley により提供されたライブラリを含むライブラリの多くは、グラフィカル・ユーザインタフェースを含んでいます。examples や example5 などは、簡単なコマンドライン関数を含んでいます。

そのようなライブラリで何ができるかの例として maps ライブラリをロードし、コマンド・ウィンドウ内でその 2、3 のコマンドを試してみてください。

```
> library(maps)
> map("county", "Washington") # ワシントンの地図
> map() # 境界線付きの USA 地図
> graphsheet()
> usa() # 異なる USA 地図
```

map によって作成された USA マップは、usa によって作成されたものよりはるかに優れています。

第 3 章

データの探索

はじめに	60
1 次元データの視覚化	61
探索的プロット	62
2 次元データの視覚化	66
散布図	68
線形近似と曲線近似を含む散布図	70
ノンパラメトリック曲線近似を含む散布図	75
線グラフと時系列プロット	81
多次元データの視覚化	92
散布図と対散布図	92
Trellis グラフ	102
3 次元グラフ	110
ダイナミックグラフ	118

はじめに

この章では探索的データ解析の概念について説明し、データ構造を探索するための様々な図表示の種類を紹介します。ここでは、データを探索する手段としてのグラフ作成技術を中心に取り上げています。さらに、**S-PLUS** はグラフを十分にカスタマイズして、プレゼンテーションに使用できるだけの品質のグラフに変換する様々なオプションを備えています。その手順は、第6章「グラフの編集」で詳しく説明します。

1次元データの視覚化

1次元データオブジェクトは、(1つの) データサンプル、一組の1変量観測値、あるいは単にデータのかたまりと呼ばれることがあります。この節では、1次元データオブジェクトの分布の形を探索するのに役立つ、いくつかの基本的なグラフについて説明します。これらのグラフは、データの分布の性質を素早く把握するのに役立つ、簡単で高性能な探索的データ解析ツールです。これを理解すれば分布があきらかに非正規なのに、正規(ガウス)分布にしかならない方法を使用するような統計的推論方法の誤用を回避することができます。

マイケルソン・データ

グラフを作成する最初のステップは、グラフ作成の対象とするデータを作成したり探したりすることです。データが大きい場合は、データベースまたはMicrosoft Excelなどのスプレッドシートにデータを保存したいかもしれません。データが小さい場合なら、**データ・ウィンドウ**に直接データを入力した方が便利です。この節では、最初にデータセット例、マイケルソン・データ(exmichel)を作成します。

1876年に、フランスの物理学者コルニュが、光の速さ c について299,990km/秒の値を報告しました。1879年には、アメリカの物理学者A.A.マイケルソンが、いくつかの実験を行ってコルニュの値を検証し精度を高めました。

マイケルソンは、光の速さに関して次のような20の測定値を得ました。

850 740 900 1070 930 850 950 980 980 880
1000 980 930 650 760 810 1000 1000 960 960

km/秒で表したマイケルソンの実測値を得るためには、上記の各値に299,000km/秒を加算します。

この20の観測値は、共通で未知の平均値 μ を持つ20の確率変数の観測値と見なすことができます。光の速さを測定する実験のセットアップにバイアスがないとすると、 μ が真の光の速さであると仮定するのが適切です。

第3章 データの探索

以下の節では、これらの観測値の分布を調べます。第8章の「統計」では、データの平均化に関するいくつかの疑問を取り上げ、その疑問に答える様々な統計的検定を行います。

データは順序が付いた一組の観測値からなるので、1つの変量からなるデータとして表現するのが適切です。データ・ウィンドウを使って、前述の20個の観測値を含む新規データセットを作成します。

1. メインメニューから、**データ ▶ データの選択**を選択して **Select Data** ダイアログを表示させます。
2. **Source** グループで、**New Data** ラジオボタンをクリックして選択します。
3. **New Data** グループの **Name** フィールドに **exmichel** と入力し、**OK** をクリックします。
4. 次に、第1列に20個のデータ点を入力します。
5. **v1** をダブルクリックし、**speed** と入力して、列（または変量）名をデフォルト **v1** から変更します。**ENTER** を押すか、データ・ウィンドウ内の他の場所をクリックして、変更を確定します。

探索的 プロット

マイケルソン・データの有用で探索的な図を得るために、箱型図、ヒストグラムと確率分布および正規QQプロットを作成します。

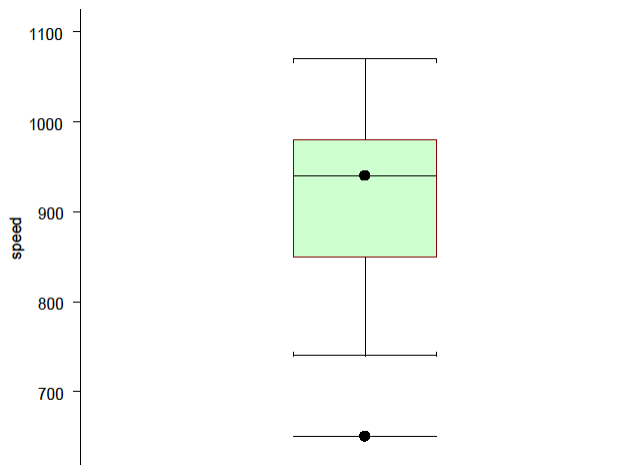


図 3.1 : マイケルソン・データの箱型図

箱型図は、中央値が約 950 で、分布がおそらく少し小さい値の方に偏っていることを示しています。また、650 は外れ値である可能性を示しています。

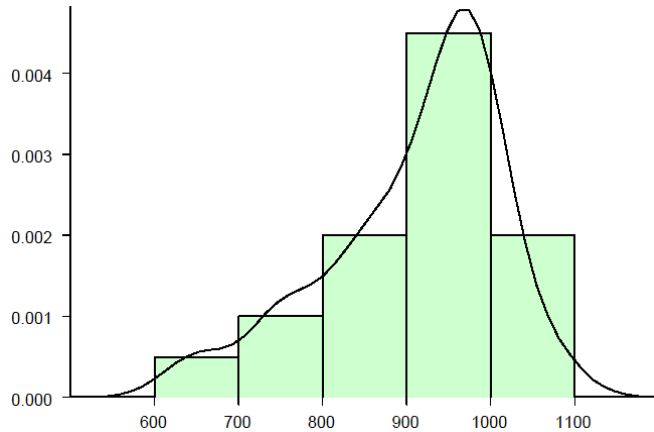


図 3.2 : マイケルソン・データのヒストグラムによる確率分布の推定

図 3.3 に示した正規 QQ プロットのデータ点は、グラフに描かれた直線にそれほど近くなりず、データが正規分布していない可能性を示しています。

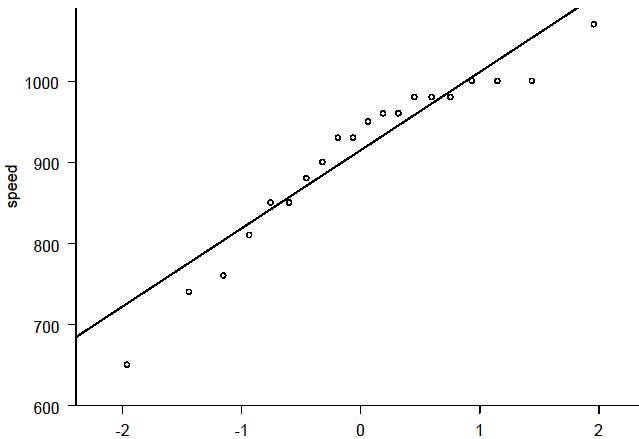
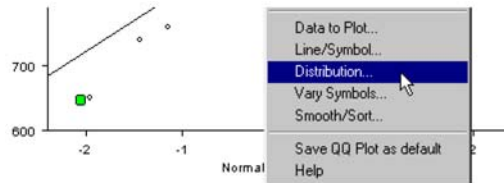


図 3.3 : マイケルソン・データの参照線を含む正規 QQ プロット

他の分布の QQ プロットを探索する

1. 他の分布の QQ プロットを作成してみましょう。任意のデータ点を右クリックして、ショートカットメニューを表示させます。ショートカットメニューで **Distribution** を選択すると、**QQ Plot** ダイアログの **Distribution** ページが開きます。



2. **Function** コンボボックスの **t** を選択し、**df 1** (自由度) ボックスに **5** を入力し、**OK** をクリックします。

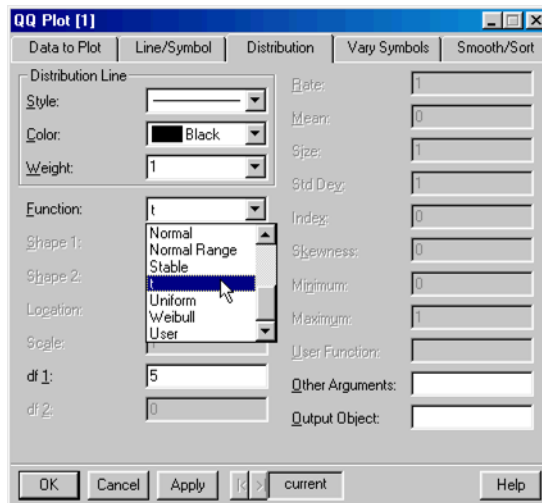


図 3.4 : QQ Plot ダイアログの Distribution ページ

t 分布 QQ プロットは線形に見えますか? **Uniform** などの他の分布を QQ プロットに試してみてください。

サンプルのサイズがとても小さいことを考慮すると、正規 QQ プロットがサンプルごとに固有の変動性を持つのではないかとと思われるかもしれません。データと同じ長さ (exmichel の場合は 20) を有する正規乱数のサンプルを作成し、作成したそれぞれの正規乱数ベクトルごとに QQ プロットを作成し、

QQプロットの変動を観察することはよい実習になります。乱数の作成については、第8章で説明します。

2次元データの視覚化

前の節では、1次元データの分布の形を素早く視覚的に考察するいくつかのグラフの作成方法を学習しました。この章では、散布図、線グラフ、および他のいくつかの2次元データのグラフ（2Dプロット）の作成方法を学習することによって、視覚的な探索的データ解析ツールについて詳しく説明します。

2次元データはしばしば2変量データと呼ばれ、データの個々の1次元成分を変量と呼ぶことがあります。たとえば、2次元グラフは、2つの変量の関係を素早く理解するのに役立ちます。たとえば、関係は線形か非線形か、変量間に高い相関があるか、外れ値があるか、明らかなクラスターがあるかなどです。2変量データの2次元グラフ表示を、（たとえば、箱型図やヒストグラムを使って）2つの変量のそれぞれの分布の1次元的表示とあわせて見ると、データを完全に理解することができます。

Main Gain データ

表 3.1 の“Main Gain”データは、住宅着工件数と新規の主要電話機増設件数との関係を表しています。第1列の“新規住宅着工”は、ニューヨーク周辺の地域のある年から次の年までの新規住宅着工件数の変化を、（機密保持のために）桁数を減らして示したものです。第2列の“主要住宅用電話機増設件数の増加”は、同じ地域における主要住宅用電話機増設件数の増加を、やはり桁数を減らして示したものです。この節では、これら2つの変量の関係を探索します。


表 3.1 : Main Gain データ

新規住宅着工	主要住宅用電話機増設件数の増加
0.06	1.135
0.13	1.075
0.14	1.496
-0.07	1.611
-0.05	1.654
-0.31	1.573
0.12	1.689
0.23	1.850

表 3.1 : Main Gain データ (続き)

新規住宅着工	主要住宅用電話機増設件数の増加
-0.05	1.587
-0.03	1.493
0.62	2.049
0.29	1.942
-0.32	1.482
-0.71	1.382

データは、2つの変数の関係を見るのに使われます。

1. 標準ツールバーの**新規データセット**ボタン  をクリックします。
2. 上のリストの 14 個の観測値を入力します。列 (変数) 名をデフォルトの V1 と V2 から `diff.hstart` と `tel.gain` にそれぞれ変更します (変数名を変更するときは、V1 と V2 をダブルクリックします)。
3. データ・ウィンドウの左上角の灰色のセルをダブルクリックし、**Name** フィールドに **exmain** と入力し、**OK** をクリックして、データセットの名前を変更します。

散布図

もしあなたが、次の年の新規住宅用電話機増設を計画する担当者で、次の年の新規住宅着工件数が推定可能な場合は、`diff.hstart`（毎年の新規住宅着工件数の増加）と `tel.gain`（毎年の住宅用電話機増設件数の増加）との間に強い関係があるかどうか、すなわち `diff.hstart` を利用して `tel.gain` を予測できるかどうかに関心をもつことでしょう。これらの2つの変量の間に関係があるかどうかを評価する最初のステップとして、図 3.5 に示すような散布図を作成します。

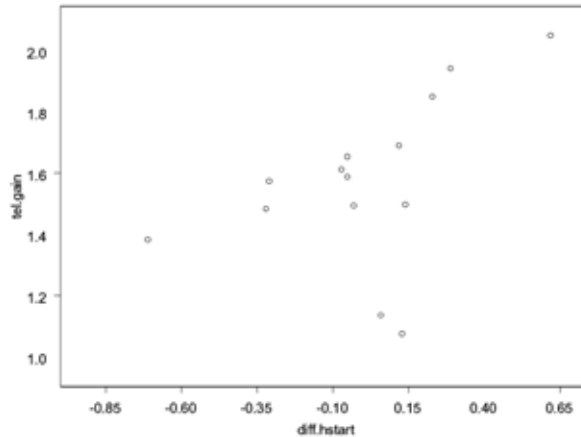


図 3.5 : `tel.gain` と `diff.hstart` の関係を示す散布図

このグラフから、データに2つの重要な特徴があることがすぐに分かります。それは、データ点のうち2つを除き、新規住宅着工件数と住宅用電話機増設件数の増加との間には、正でほぼ線形の関係があることです。例外的な2つのデータ点は、データの残りの部分から大きく離れており、そのようなデータ点を外れ値と呼びます。

外れ値を識別する

マウスポインタを離れた点のうちの1つの真上に移動させてください。図 3.6 に示すように、その点における 2 つの変量の値を示すデータティップが現われます。

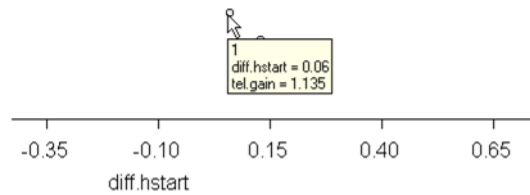




図 3.6 : ポインタ位置における変量値を示すデータティップ

データティップの最初の行に現われる数に注意してください。この数は、その点に対応するデータセットの行番号を示しています。図 3.6 では、データティップは、この点が第 1 行であることを示しています。次に、マウスポインタを別の外れ値の上の移動させます。この点のデータティップは、その点が第 2 行であることを示します。したがって、データセット内の最初の 2 つの観測値は、外れ値です。

点の選択とハイライト

散布図内のデータ点を色でハイライトして、データの残りの部分と区別することができます。exmain データの散布図で、2 つの外れ値をハイライトしましょう。


1. グラフツールバーのグラフツールボタン  をクリックして、グラフツールパレットを開きます。
2. グラフツールパレットの矩形内のデータ選択ボタン  をクリックします。マウスカーソルが、小さい四角形の注釈がついた十字線になります。

七

3. 四角形を 2 つの外れ値のまわりにドラッグして選択します。これで、外れ値がデフォルトでは赤でハイライトされます。(CTRL キーを押したままマウスボタンをはなすと別の点をハイライトすることができます。)

注意


散布図内の点を選択すると、その点は、データが表示されているどの**データ・ウィンドウ**内でも選択されます。

4. **グラフツールパレットのオブジェクトの選択ボタン**  をクリックして、十字線のマウスポインタを通常のマウスポインタに戻します。
5. 処理が終わったら**グラフツールパレット**を閉じてください。すべての点の選択を解除するときは、**データ・ウィンドウ**内のセルをクリックしてください。

線形近似と 曲線近似を 含む散布図

散布図のデータに直線をあてはめて、そのあてはめをデータと重ねることができます。そのようなあてはめを利用すると、データが 2 つの変量間の線形関係とどれだけ一致するかを視覚的に評価することができます。線形のあてはめが十分と思われるときは、直線グラフは、2 変量間の勾配と直線近似に関するデータの変動を両方とも視覚的によく表しています。

最小二乗法による直線近似

最小二乗法によって `exmain` の 2 変量データに直線をあてはめ、その結果をデータの散布図に重ねて表示させることができます。散布図を作成するときは前と同じように行い、今度は**散布図**ボタンではなく、**2D プロットパレット**の**回帰直線**ボタン  をクリックします。この結果を、図 3.7 に示します。

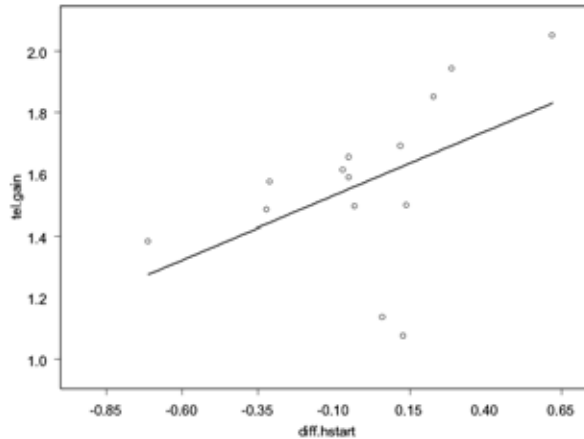



図 3.7 : `tel.gain` と `diff.hstart` の関係を示す最小二乗回帰線を含む散布図

データの 2 つのはずれ値が最小二乗法による回帰線の傾きを下げていることに注意してください。

ロバスト近似

直線による最小二乗法による近似は、外れ値が回帰線に対して大きく影響する可能性があるという点で頑健ではありません。ロバスト法は、外れ値がどれほど大きくてもそれほど大きい影響を受けません。刈り込み最小二乗法 (LTS) と呼ばれる方法でロバスト線をあてはめ、その結果を表示させるときは、データを選択し、**2D プロットパレット**の**ロバスト**ボタン  をクリックしてください。その結果を、下の図 3.8 に示します。**グラフシート**を `exmain.sgr` として保存します。

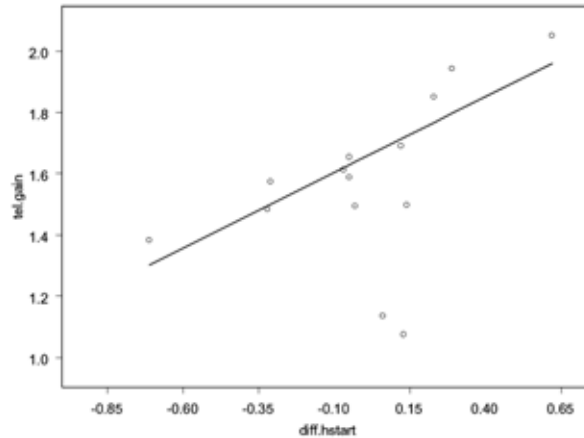



図 3.8 : ロバスト LTS 線を含む tel.gain と diff.hstart の関係を示す散布図

図 3.7 と図 3.8 を比較して、2 つの外れ値が最小二乗法による回帰線に及ぼす影響の大きさに注意してください。

選択された点を削除した近似

exmain データの回帰直線が 2 つの外れ値の影響を受けることが明らかなため、これらの 2 つの点を削除して最小二乗法による近似を行うときの影響を確認するのもよいでしょう。これは次のようにすると、とても簡単に行うことができます。

1. tel.gain と diff.hstart の最小二乗法による回帰線を含む散布図を作成し、前に行ったように 2 つの外れ値を選択します (図 3.9 を参照)。外れ値のデータ点を選択した後、**グラフツール**パレットのオブジェクトの**選択ボタン**  をクリックして、カーソルを通常の形に戻すのを忘れないでください。

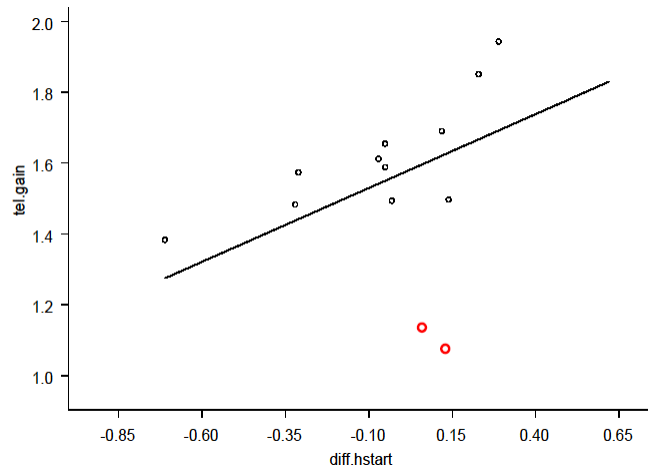


図 3.9 : 最小二乗線を含み外れ値の点を選択した散布図

2. メインメニューから、書式 ▶ 指定データを除いて使用を選択します。この結果、図 3.10 に示すように、外れ値が全くないデータをあてはめる新しい回帰線が表示されます。縦軸の目盛りが変更され、削除した2つの外れ値がグラフに現われていないことに注意してください。

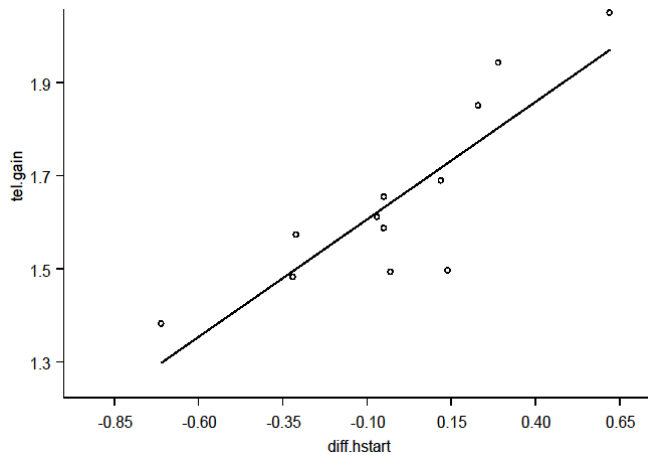



図 3.10 : 新たな回帰線を含み外れ値の点を除去した散布図

適切なロバスト線あてはめ法の大きな特徴は、データが外れ値を含まないときに、最小二乗法で得られるものに近い直線近似が得られることです。作成したばかりのグラフに刈り込み最小二乗 (LTS) ロバスト線を加えることによって、この最小二乗法による回帰に対するロバスト LTS 近似を確認することができます。

3. プロットする列が選択された状態で、グラフ領域を選択し、SHIFT キーを押し、**2D プロットパレット**の**ロバスト**ボタン  をクリックします。図 3.11 に示したように、外れ値を削除した線型回帰線と外れ値を含む LTS 近似は、実際互いにかかなり似たものであることが分かります。散布図が元の軸範囲を表示しており、線型回帰線からは削除した 2 つの外れ値が、ロバスト LTS 線と共に表示されていることに注意してください。

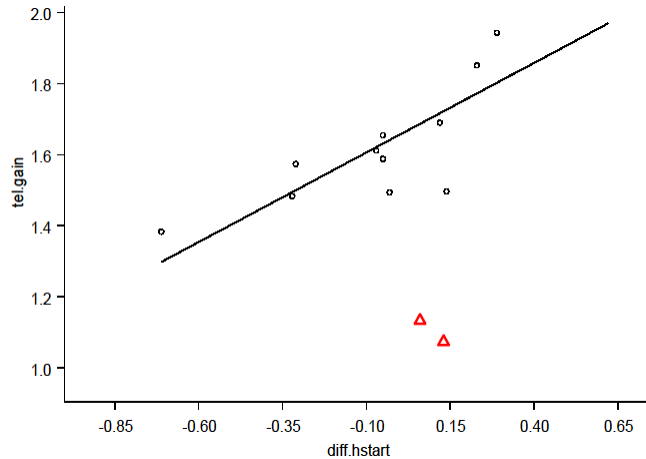


図 3.11 : 最小二乗法による回帰線 (外れ値なし) とロバスト LTS 線を含む散布図

4. 2 つの外れ値を除いたデータの線型回帰線を表示し、さらに外れ値を追加して線形回帰線がどのように変化するかを簡単に確認することができます。メインメニューから**書式 ▶ すべてのデータを使用**を選択すると、図 3.12 に示すグラフが作成されます。すべて点が用いられ、グラフ上のすべての既存のプロットが再計算されます。

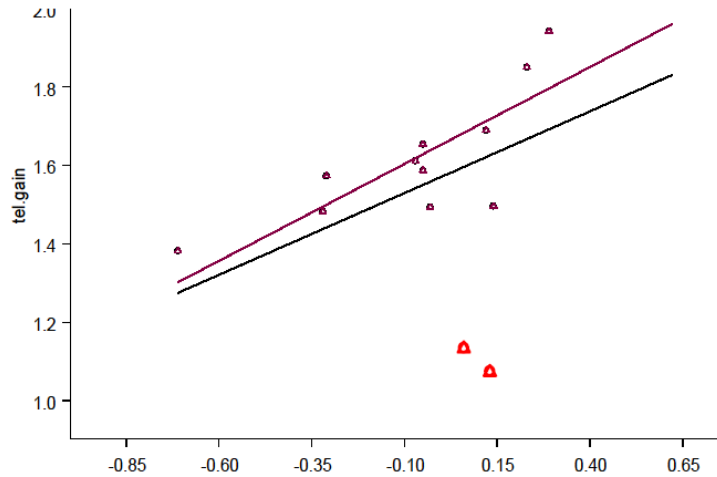


図 3.12 : 外れ値を含む線形回帰線とロバスト LTS 線

ノンパラメ トリック曲線 近似を含む 散布図

前の節では、線形パラメトリック関数を散布図のデータにあてはめました。しばしば、どの関数を使用すべきか決定するのに十分な情報がないことがあります。そのような場合、特定の非線形関係を想定しないノンパラメトリック曲線をあてはめることができます。

ノンパラメトリック曲線近似は、データの一般的な傾向を示す平滑曲線を作成するため、スプーザーとも呼ばれます。最も簡単なスプーザーは移動平均の応用です。特定の x 値におけるあてはめ値は近くの点の y 値の加重平均として計算され、その加重はその x 値と対象となる x 値との間の距離が増加すると減少するよう設定されます。最も簡単なタイプの移動平均スプーザーにおいて、対象となる点からの一定の距離（ウィンドウ）内にあるすべての点が、その点の平均計算に使用されます。

ウィンドウ幅は、スプーザーの帯域幅と呼ばれます。帯域幅を広くするとよりなめらかな近似を得ることができますが、局所的に変化する特徴をとらえられなくなることがあります。帯域幅を狭くすると、スプーザーは局所的に変化する特徴をよりの確に追跡できるようになりますが、曲線は複雑な形になります。

さらに進んだ平滑化もこの移動平均アプローチの応用です。なめらかに減少する加重や、局所回帰が用いられます。スムーザーにはすべて、曲線の滑らかさを制御する滑らかさのパラメータ（帯域幅）があります。

適切な帯域幅を選択する問題は複雑で、多くの統計に関する研究論文で扱われてきました。しかし、実際のデータにいくつかのスムーザーをあてはめることによって、帯域幅を変化させて得られる実際の感覚をつかむことができます。

この節では核関数による平滑化、スプライン平滑化、局所平滑化の異なる3つのスムーザーの使用法と、曲線近似の滑らかさの度合い（すなわち、データの“平滑度合”）を制御する帯域幅の選択方法について説明します。

空気中の様々なレベルの窒素酸化合物（NO_x）に対する、異なる8つの半導体素子センサの反応値からなるサンプルデータセット `sensors` を使用します。これらのセンサを設計した技術者は、1つではなく2つのセンサを使用することによって、NO_xの濃度をより正確に測定できるようになるかどうかを確認するために、これらの8つのセンサの反応値の関係を研究しています。これまでの調査では、2つのセンサの反応値に非線形関係があることは分かっていますが、その詳しい関係についてはよく分かっていません。

核関数による平滑化

核関数による平滑化は、様々な重み関数（核）を使用して、点と点との間を、単純な局所平均より滑らかにつなぐ局所平均の一般化です。デフォルトの核は、この章の始めに説明した局所平均の方法を提供する箱型です。

センサ5とセンサ6との関係を示す散布図を作成し、単純な移動平均スムーザーの帯域幅（しばしば、“ボックスカー”スムーザーと呼ばれる）を試してみます。最初に、2次元グラフの機能を使って散布図と移動平均をデフォルト帯域幅の選択値とした核平滑化曲線を作成します。

ボックスカースムーザー

1. メインメニューから、**データ ▶ データの選択**を選択します。
2. **Source** グループの **Existing Data** が選択されていることを確かめます。

3. **Existing Data** グループにおいて、**Name** フィールドに **sensors** と入力し、**OK** をクリックします。
4. 列 v5 と v6 を選択します。
5. メインメニューから、**グラフ ▶ 2D プロット・グラフ** を選択して **Insert Graph** ダイアログを開きます。
6. 図 3.13 に示すように、**Plot Type** リストボックスで **2D 核関数による平滑** を選択し、**OK** をクリックします。

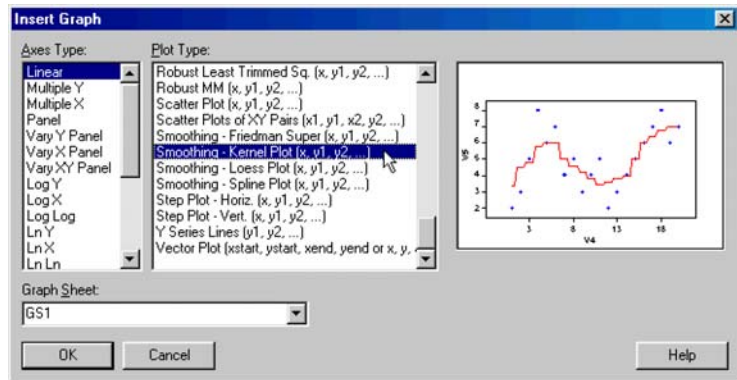


図 3.13 : Insert Graph ダイアログ

その結果、図 3.14 に示すようなグラフが表示され、あまりデータにあてはまらず滑らかでない曲線が生成されます。これは平滑化帯域幅がこのデータに対し小さすぎるためです。

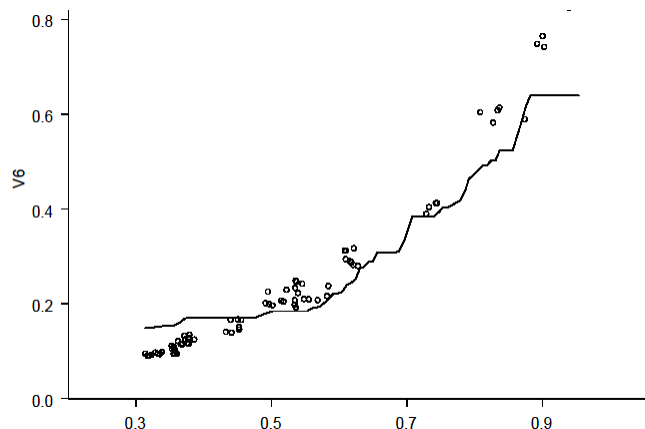


図 3.14 : ボックス核平滑化線を含むセンサ 5 とセンサ 6 の関係

帯域幅を変える

1. 散布図（または、曲線近似線）の点のうちの1つを右クリックし、ショートカットメニューから **Smooth/Sort** を選択します。
2. 現れた **Line/Scatter Plot** ダイアログの **Smooth/Sort** ページで、(**Kernel Specs** グループの **Bandwidth** ボックス内にある) スムーザーの帯域幅のデフォルト値に注意してください。**Bandwidth** ボックスに 0.1~0.6 の間の様々な帯域幅選択値を入力し、**Apply** をクリックして、試してみてください (**Line/Scatter Plot** ダイアログは開いたままです)。“見たところ” 曲線近似が最もよくあてはまる帯域幅はどれでしょうか。図 3.15 に、帯域幅選択値 0.3 のスムーザーを示します。

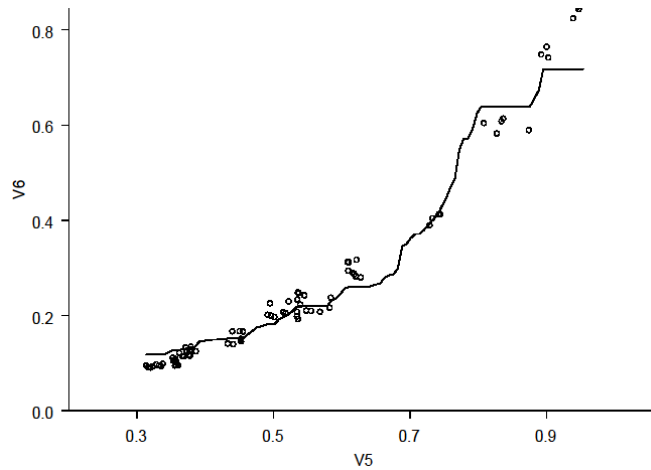


図 3.15 : 帯域幅を 0.3 としたボックス核平滑化線を含む
センサ 5 とセンサ 6 の関係

核平滑化のタイプを Parzen スムーザーに変更する

- **Line/Scatter Plot** ダイアログの **Smooth/Sort** ページが開いた状態で (このページが閉じている場合はもう 1 度開き)、**Kernel Specs** グループの **Kernel** プルダウンリストから **Parzen** を選択し、**Apply** をクリックします。(Parzen カーネルは、三角形にたたみ込まれたボックスです) 再び、帯域幅選択値の選択を試してみてください。滑らかな曲線近似が得られましたか。図 3.16 に、帯域幅 0.15 の場合の Parzen 核平滑化を示します。

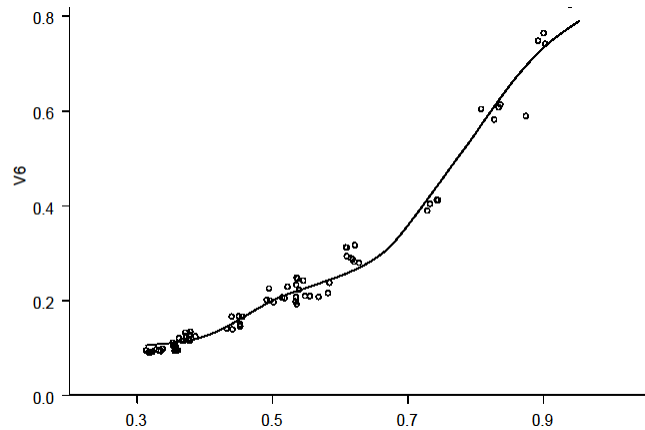



図 3.16 : 帯域幅を 0.15 とした Parzen 核平滑化線を含む
センサ 5 とセンサ 6 の関係

スプライン 平滑化

3 次平滑化スプラインは、一連の局所的な 3 次多項式をつなぎ合わせるこ
によって計算されます。滑らかさは、隣り合った多項式の値、勾配および曲
率を、それらが合うところで一致させることによって保証されます。“平滑化”
パラメータは、曲線近似の滑らかさの度合いとデータ値に対するあてはまり
のよさとのトレードオフにより決定され、多項式内の曲率の大きさを制御し
ます。3 次平滑化スプラインのデータ値へのあてはめが正確になるほど曲線が
複雑になり、この逆のことも言えます。

S-PLUS は、データ値に基づいて理論的に調整する技術を用いて自動的に平
滑化パラメータを選択します。あるいは平滑化パラメータ値を指定して、ス
プライン平滑化の滑らかさを制御することができます。

スプライン平滑化のあてはめ

- 自動帯域幅選択による 3 次平滑化スプライン曲線を含むセンサ 5 とセン
サ 6 の関係を表す散布図を作成します。グラフ内の点をクリックし、
次に 2D プロットパレットの **スプライン** ボタン  をクリックするこ
とにより、核型平滑プロットをスプラインプロットに変換することが
できます。得られたグラフを、図 3.17 に示します。

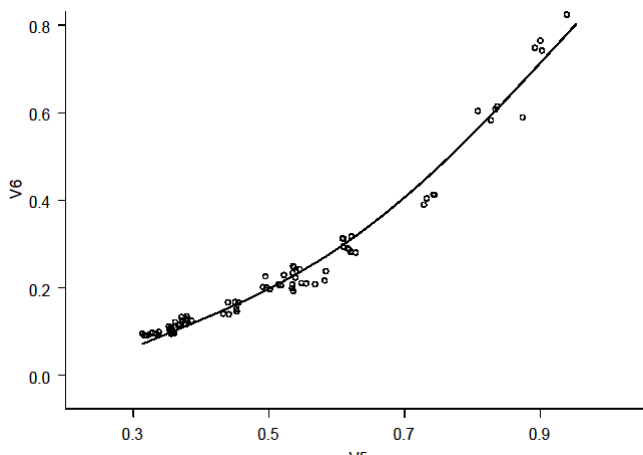


図 3.17 : スプライン平滑化線を含むセンサ 5 とセンサ 6 の関係

帯域幅を選択する

- プロットのショートカットメニューから、前と同じように **Smooth/Sort** を選択し、**Line/Scatter Plot** ダイアログの **Smooth/Sort** ページを開きます。**Smoothing Spline Specs** グループに、スプライン近似の滑らかさを制御するために用いられる **Deg. of Freedom** (自由度) プルダウンリストの使用可能な選択値を試してください。図 3.18 に、自由度 6 のスプライン平滑化の結果を示します。

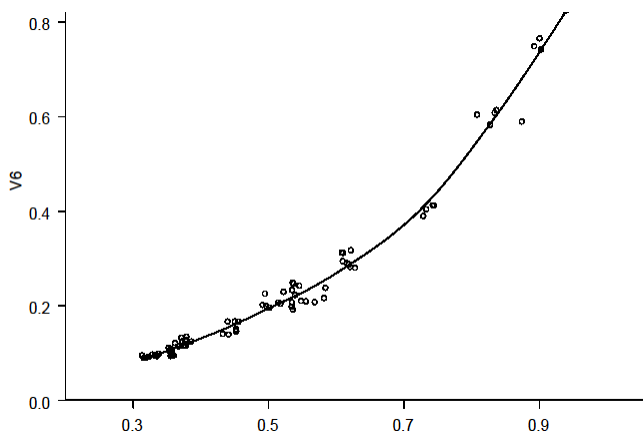



図 3.18 : 自由度を 6 としたスプライン平滑化線を含むセンサ 5 とセンサ 6 の関係

局所平滑化

ベル研究所の W.S.Cleveland と共同研究者によって開発された局所平滑化は、データに対する局所線形回帰または局所二次近似に基づく優れた手法です。

局所平滑化のあてはめ

- グラフ内の点をクリックしてそのグラフを選択し、次に **2D プロット** パレットの **局所重み付き最小二乗法** ボタン  をクリックして、**sensors** データに局所曲線近似をあてはめ、そのデータにより曲線近似線を表示します。その結果を図 3.19 に示します。

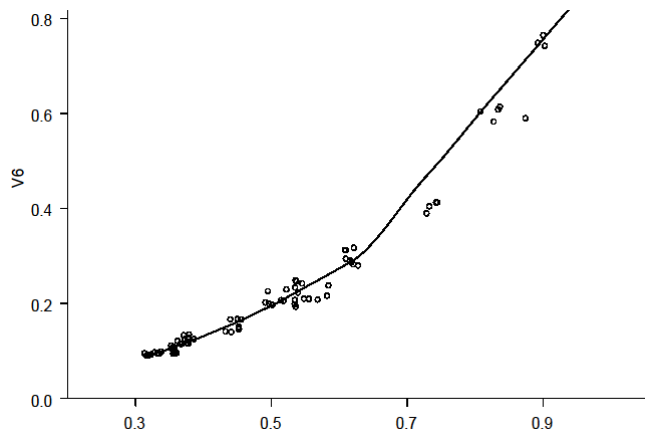


図 3.19 : 帯域幅をデフォルトとした局所平滑化線を含む
センサ 5 とセンサ 6 の関係

局所平滑化の帯域幅を変更する

- **Line/Scatter Plot** ダイアログの **Smooth/Sort** ページを開きます。**Loess/Friedman Specs** グループで、**Span**、**Degree** および **Family** とラベルのついた 3 つのボックスがあります。0.5 (デフォルト値) よりも上と下の **Span** 値を試してみてください。この数は帯域幅を、データの x 軸値の範囲に対する割合として決定します (この場合、センサ 5 の値の範囲)。また、**Degree** ボックスの値として、デフォルト値の “One” の代わりに “Two” も試してみてください。これにより、局所線形回帰と局所二次近似のどちらを使用すべきかが決まります。

線グラフと 時系列 プロット

散布図は、横軸の変量に特定の順序があるかどうかに関係なく、2 つの変量の関係を視覚化するのに便利なツールです。一方で、データによってはしばしば、散布図で視覚化したい 2 つの変量のうち 1 つはデータセットの行にしたがって、最も小さいものから最も大きいものの順で並べられることもあります。

第3章 データの探索

直線近似またはノンパラメトリック曲線近似の線を散布図に重ねたときに分かったように、線グラフは2つの変量間の関係を視覚化するのに便利です。このような場合には、 y 軸の変量に対応する x 軸の値の順序付けは、直線近似または曲線近似を行う際に、水面下で自動的に実行されます。


それ以外でデータの順序付けが重要なのは時系列データです。この場合、観測値は、連続する時間の瞬間瞬間で測定されます。データに順序がある場合は、単にデータの散布図を作成するよりも、連続するデータの値を直線で結んだ「線グラフ」を作成した方が有効です。直線で点を結ぶことにより、順序付けしたデータ値の全体の傾向、または形をはっきりと知ることができます。

この章の最初の方で紹介した `exmain` データセットに含まれる変量 `tel.gain` および `diff.hstart` に戻ります。これらは両方、1971年から14年間、年に1度、1月1日に記録された時系列の値です。これらの変量を使用して線グラフと時系列プロットを作成します。

線グラフ

簡単な線グラフ

`tel.gain` 変量の線グラフを作成するには、

1. 67 ページで説明したように、**データ ▶ データの選択**を使って `exmain` を開くかまたはデータセットを作成します。
2. `tel.gain` 列を選択します。
3. **2D プロットパレットの線グラフボタン**  をクリックします。得られたグラフを、図 3.20 に示します。

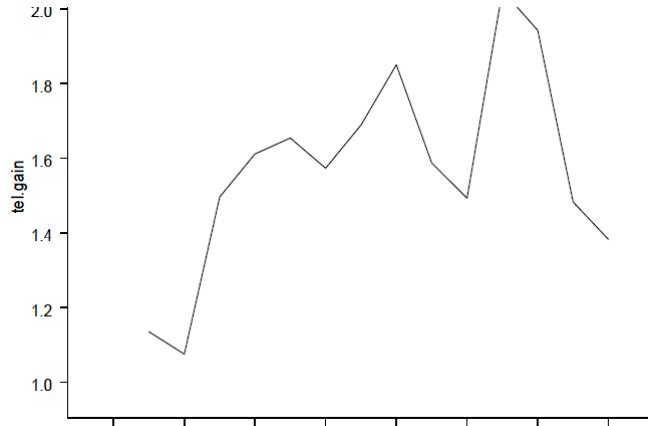



図 3.20 : tel.gain の線グラフ

プロットする変数が 1 つのとき、S-PLUS はこの変数が y 軸の変量であると想定し、 x 軸の変量として整数 $1, 2, \dots, n$ を自動的にあてはめます。 n は、変量の長さ（すなわち変量のデータ数）です。tel.gain と diff.hstart の両方を選択すると、S-PLUS は選択した最初の変量が x 軸の変量であると想定するため、意味のない線グラフが作成されます。

点を含む線

記号を重ねた線グラフを作成するには、

1. 線グラフをクリックして選択します。
2. 2D プロットパレットの線付き散布図ボタン  をクリックします。得られたグラフを、図 3.21 に示します。

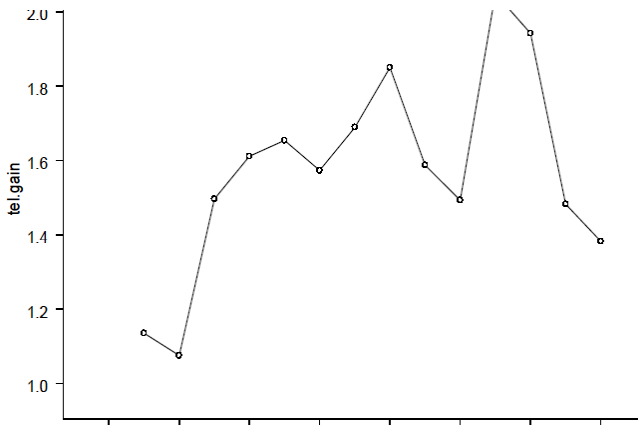



図 3.21 : tel.gain の線付き散布図

新規住宅用電話機増設件数の増加は、14年間のうち最初の2年間は最も小さく、3年目に急に大きくなり、6年目から大きく上下することに注目してください。

飛び線付き散布図

点をさらにはっきり見るために、散布図の点が線から少し離れた線グラフを作成することができます。このグラフを作成するには、

1. グラフ内の任意の点または線をクリックして、そのグラフを選択します。

2. 2D プロットパレットの飛び線付き散布図ボタン  をクリックします。得られたグラフを、図 3.22 に示します。

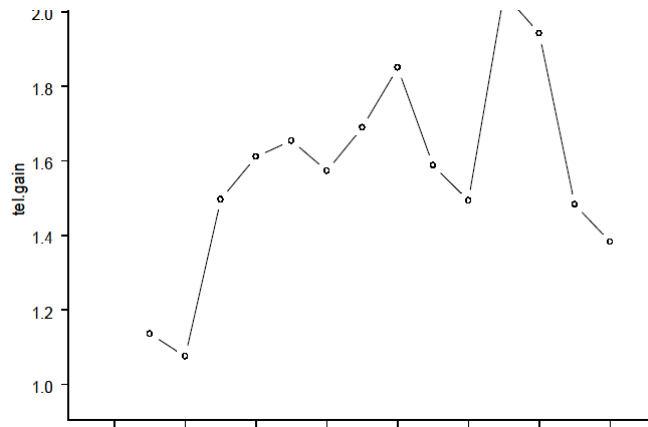


図 3.22 : tel.gain の飛び線グラフ

複数の線グラフ


これまで tel.gain の時系列の挙動を見てきましたが、1つのグラフに2つの時系列の線グラフを作成して、diff.hstart の挙動も同時に見たいケースもあります。

これは、次のような2つの方法で行うことができます。最初の方法は、線グラフの一方がすでに作成されており、別の線グラフを追加したい場合に適しています。2番目の方法は、両方の線グラフを一度に作成したい場合に適しています。

第2番目の線グラフを追加する

前に行ったように、第1番目の線グラフがすでに作成されている場合は、次のように通常の方法で第2番目の線グラフだけを追加することができます。

1. グラフ領域の内側（プロット上ではない）をクリックし、既存のグラフを選択します。グラフの境界に沿って緑の8つの四角形が現れ、グラフが選択されたことを示します。
2. データ・ウィンドウの変量 `diff.hstart` を選択します。

3. SHIFT キーを押しながら、2D プロットパレットの飛び線付き散布図ボタン  をクリックします。これにより、図 3.23 に示すグラフが作成されます。

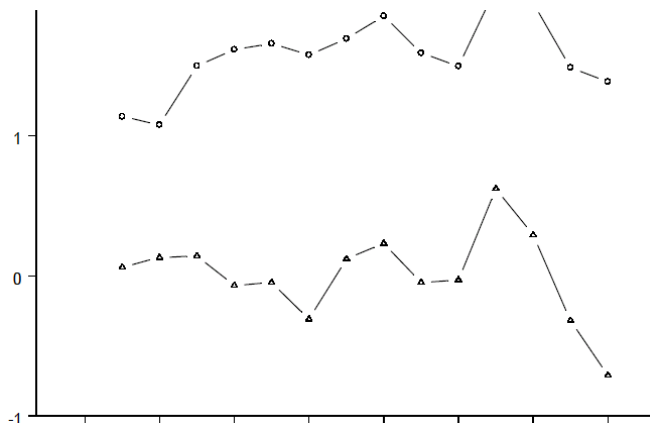




図 3.23 : tel.gain と diff.hstart の飛び線付き散布図

時系列グラフボタンの使い方

前の例は、既存のグラフに別の線グラフを追加する方法を示しました。2D プロットパレットの時系列グラフボタン  を使って、複数の線グラフを一度に作成することができます。

1. データ・ウィンドウの tel.gain と diff.hstart を (任意の順序で) 選択します。

2. 2D プロットパレットの**時系列グラフ**ボタン  をクリックします。得られたグラフを図 3.24 に示します。これは、図 3.23 とよく似ています。

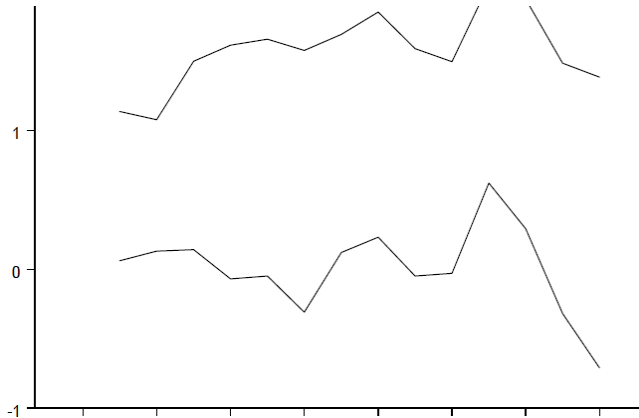



図 3.24 : tel.gain と diff.hstart の Y 系列プロット

タイトルと凡例を追加する

メインタイトル “Main Gain”、 x 軸のタイトル “Year” および凡例を追加するには、

1. メインメニューから、**挿入 ▶ タイトル ▶ メインタイトル**を選択します。**@Auto** と現われた場所に **Main Gain** と入力し、次に編集ボックスの外でクリックします。
2. 目盛りの真上（目盛りラベルではない）をクリックして、 x 軸を選択します。
3. メインメニューから、**挿入 ▶ タイトル ▶ 軸**を選択します。**@Auto** と現われた場所に **Year** と入力し、次に編集ボックスの外でクリックします。

4. グラフツールバーの**凡例の自動作成**ボタン  をクリックします。凡例が、自動的に作成され、グラフに入ります。凡例を選択し（凡例の境界のすぐ内側をクリックする）希望の位置にドラッグして、凡例を位置変更することができます。得られたグラフを、図 3.25 に示します。

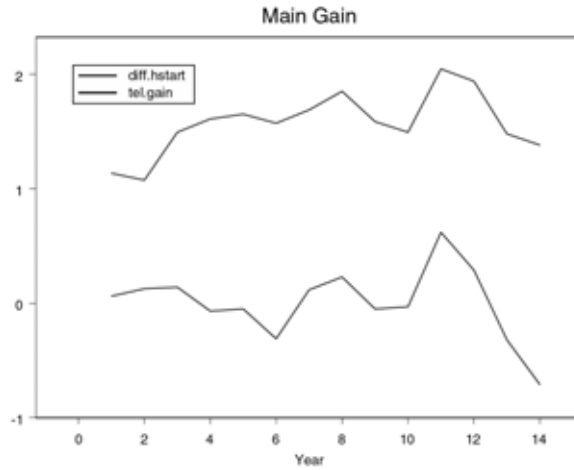




図 3.25 : tel.gain と diff.hstart のフォーマットされたグラフ

別々のパネルにグラフを作成する

2つの時系列はスケールが少し異なるため、別々のパネルで見ることができると便利です。

1. グラフ領域の内側（プロット上ではない）をクリックしてそのグラフを選択します。軸上に8つの緑の四角形が現れ、グラフが選択されたことを示します。
2. グラフツールバーの**グラフツール**ボタン  をクリックします。

3. グラフツールパレットのパネル分割 (Y) ボタン  をクリックします。図 3.26 を参照してください。

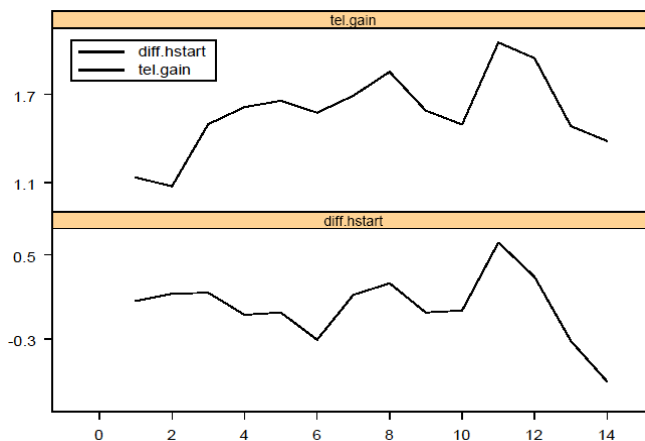


図 3.26 : 異なる y 軸により分割されたパネル

Main Gain プロットを解析する

時間に対する `tel.gain` と `diff.hstart` の線グラフ（時系列プロット）を作成することによって、ただ単に 2 変量のグラフを作成しただけよりも詳細な分析が可能です。両方の図表示を使用すれば、データの理解がより完全なものになります。

この章の始めに、外れ値がデータセットの最初の 2 行にあると判断しました。時系列プロットは、この最初の 2 年間の `tel.gain` の値が、14 年分のデータの中で最も小さかったことを表しています。同時に、最初の 2 年間の `diff.hstart` の値は 14 年間の全体の平均に近いものでした。さらに最初の 4 年を除き、2 つの系列の間には著しい相関パターンがあり、片方が増加すると他方も増加することに注意してください。最初の 2~4 年の 2 つの変量の相対的挙動は、その他の年とは異なることが分かります。そのため、14 年よりも後の予測モデルとして最初の 2 年を削除した線形回帰か、ロバスト LTS 近似を用いたモデルに確信が持てるはずで

時系列プロット

exmain データの線グラフは、整数 1、2、... 14 を使用するあいまいなラベル付けではなく、*x* 軸（時間軸）に実際の日付を出力した方がよいものになります。このためには、第 3 の変数 Year を生成し、観測された年を指定します。

日付変数を作成する

Year と名前を付けた整数変数を作成するには、

1. exmain データ・ウィンドウにフォーカスを当て、列の一番上をクリックして、exmain データセットの第 3（現在空白である）列を選択します。

	1	2	3	4
	diff.hstart	tel.gain		
1	0.0600	1.1350		
2	0.1300	1.0750		
3	0.1400	1.4960		
4	-0.0700	1.6110		
5	-0.0500	1.6540		

2. メインメニューから、挿入 ► 列を選択して **Insert Columns** ダイアログを表示させます。
3. **Name(s)**テキストボックスに **Year**、**Fill Expression** テキストボックスに **1971:1984** と入力します。**Column Type** として **integer** を選択し、**OK** をクリックします。（図 3.27 を参照。）

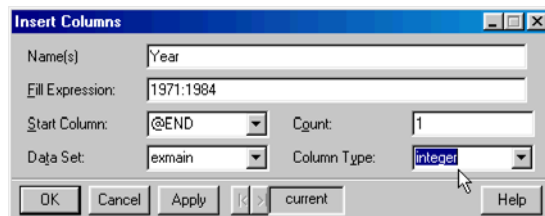
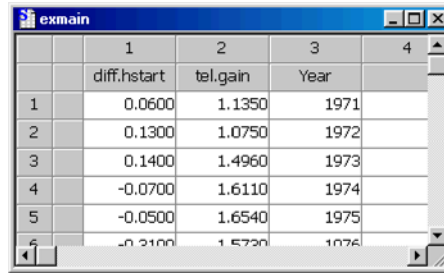


図 3.27 : Insert Columns ダイアログ

これで、第3の列には、下に示すような年を示す整数が入ります。



	1	2	3	4
	diff.hstart	tel.gain	Year	
1	0.0600	1.1350	1971	
2	0.1300	1.0750	1972	
3	0.1400	1.4960	1973	
4	-0.0700	1.6110	1974	
5	-0.0500	1.6540	1975	
6	-0.2100	1.5720	1976	


4. Year 列を選択します。CTRL キーを押し、tel.gain と diff.hstart を選択します（どの順序でもかまいません）。
5. 2D プロットパレットの飛び線付き散布図ボタン  をクリックします。得られたグラフを、図 3.28 に示します。



図 3.28 : tel.gain と diff.hstart の年次時系列プロット

多次元データの視覚化

これまでの節では、単純な1次元と2次元のデータの視覚化について説明しました。1次元と2次元のデータの場合は、データの基本的なすべての情報を一組のグラフで簡単に表すことができます。それぞれのグラフ（散布図、箱型図、ヒストグラム）は異なる種類の情報を提供しますが、どのグラフを使用するかを決定するのは簡単です。

多次元データの場合の視覚化はさらに複雑になります。1変量と2変量の関係の他に、2つの変量間の関係が他の変量値によって変化するような交互作用がある場合があります。この節では標準的な多次元視覚化方法と、Trellis グラフなどの新しい方法について説明します。

まず最初に、多次元データの散布図と対散布図について説明します。次に、データに関してさらに高次元の関係を見つけるために特に役立つTrellis グラフについて説明し、次に3次元データのために設計されたグラフを紹介します。最後に、ブラッシングとスピンを行う対話式のダイナミックグラフツールについて説明します。

散布図と対散布図

2次元散布図を使用して、色/スタイル/サイズなどの記号の属性を他の列の値により変更させて、多次元データを表示させることができます。

対散布図は、多次元データを素早く視覚化することができる強力なグラフツールです。データの解析作業を行うとき、最初のステップはそのデータについてよく知ることです。対散布図を生成することで、そのプロセスがとても容易になります。

燃料データ

fuel.frame データセットは、『Consumer Reports』1990年4月号から引用された60種類の自動車に関する情報を含んでいます。変量は、次の通りです。



- Weight : 自動車重量
- Disp : エンジン排気量 (6リッター、8リッターなど)
- Mileage : 1ガロン当たりの走行マイル数
- Fuel : 100/mileage
- Type : 車の分類 (Large、Medium、Small、Compact、Sporty、Van)

これらの変量の間には、どのような関係があるでしょうか。

カラープロット

最初に、重量、燃費および車の種類の関係を見てみましょう。まず、車の種類によって記号の色が異なる散布図を生成します。

凡例付きのカラープロットを生成する

1. **Select Data** ダイアログを使って、データ・ウィンドウに `fuel.frame` データセットを表示させます。
2. `Weight` をクリックし、次に `Mileage` と `Type` を **CTRL**-クリックします。
3. **2D プロットパレットのカラープロットボタン**  をクリックします。
4. **グラフツールバーの凡例の自動作成ボタン**  をクリックします。

記号のスタイルを変更する

車の様々な種類を識別するために、記号のスタイルを変更することができます。

1. 任意の点を右クリックし、ショートカットメニューから **Vary Symbols** を選択します。
2. **Vary Style By** を **z Column** に変更します。(**Data to Plot** ページを見て、`Type` が `zColumns` と指定されていることを確認してください。)
3. **OK** をクリックします。(記号のスタイルと色のデフォルト設定を変更することができます。詳細は、第 13 章「S-PLUS セッションのカスタマイズ」を参照してください。) 得られた色のプロットを、図 3.29 に示します。

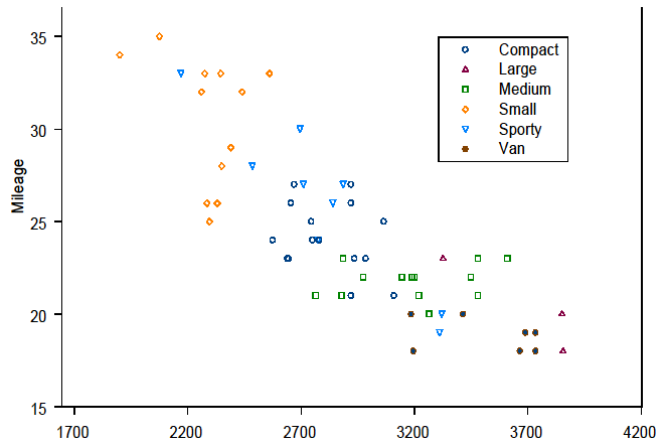


図 3.29 : fuel.frame データの車の種類によって色と記号のスタイルが異なる Weight と Mileage のカラープロット

すべての車に関して、重量が増えるほど燃費が低下する傾向があることに注目してください。ただし、この関係は車の種類によって異なります。たとえば、Sporty 車は重量と燃費の範囲がどちらも広がっています。

対散布図


次に対散布図を作成して、データセット内のすべての変量間の関係を同時に表示させます。

対散布図を生成する

1. すべての変量をプロットするため、**データ・ウィンドウ**の左上角をクリックして、列をすべて選択します（すべての列が黒くなります）。

注意

データ列を選択する順序により、データが対散布図に現われる順序が決まります。上のケースでは順序を考慮していませんでしたが、順序を考慮する場合は、CTRL-クリックで列を希望の順序で選択してください。

2. 2D プロットパレットの対散布図ボタン  をクリックします。得られた対散布図を、図 3.30 に示します。

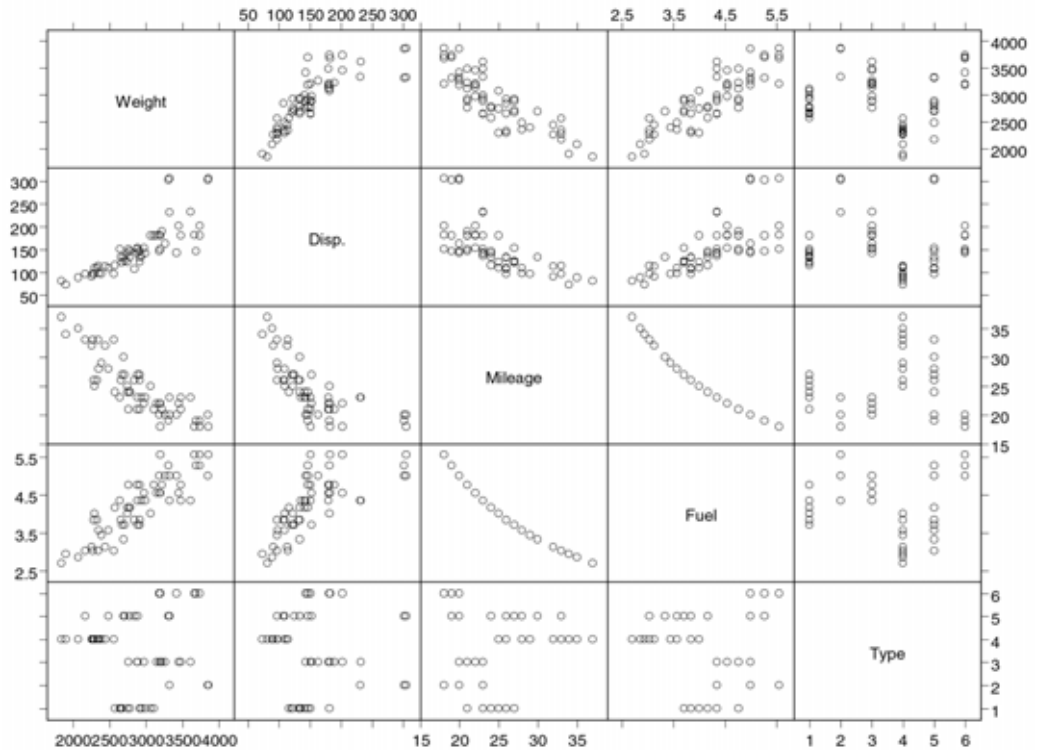


図 3.30 : fuel.frame データの対散布図。いくつかの強い関係が現われている。

グラフを見てみると、いくつかの強い線形関係があることがすぐに分かります。たとえば、予想していたかもしれませんが、重量とその燃料消費量には正の線形関係 (Weight が増えるほど Fuel も増える) があります。対散布図を利用すると、データにおける 2 変量のより明らかな関係を素早く、簡潔に理解することができます。

ヒストグラムを追加する

次に、図 3.30 に示された 5 つの変量のヒストグラムを追加しましょう。

1. 任意のデータ点を右クリックしてグラフのショートカットメニューを表示させ、**Line/Histogram** を選択します。

2. **Scatter Plot Matrix** ダイアログの **Line/Histogram** ページで、**Draw Histograms** チェックボックスをクリックします。

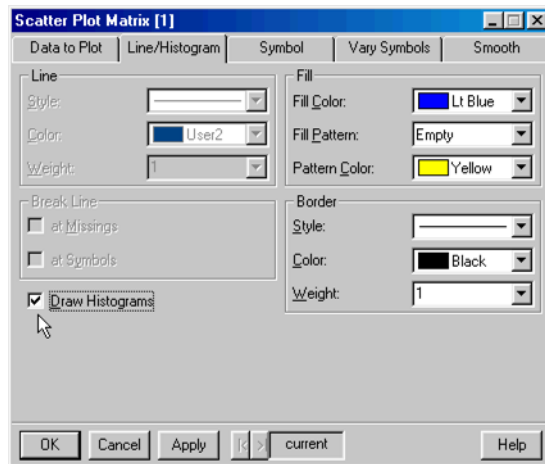


図 3.31 : Scatter Plot Matrix ダイアログの Line/Histogram ページ

3. **OK** をクリックすると、対散布図とヒストグラムが再描画されます。

ヒストグラムは、その他にも有用な情報を提供します。たとえば、変数 *Weight*、*Disp.* および *Fuel* は、正規分布または正規分布に近い分布を持つことが分かります。これと対照的に、変数 *Mileage* は、正規分布していないようです。結論を導き出す前に、これらの推測をさらに調査する必要があります。

最小二乗法による回帰線を追加する

S-PLUS の散布図では、様々な線のあてはめを行うことができます。先の対散布図を使用して、線形回帰/ロバスト/核/局所/平滑化スプラインをあてはめることができます。たとえば、図 3.30 に最小二乗法による回帰線を追加するときは、次のように行います。

1. 任意のデータ点を右クリックしてプロットのショートカットメニューを表示させ、**Smooth** を選択します。

2. **Scatter Plot Matrix** ダイアログの **Smooth** ページで、**Smoothing Type** プルダウンリストから **Least Squares** を選択し、**Apply** をクリックします。

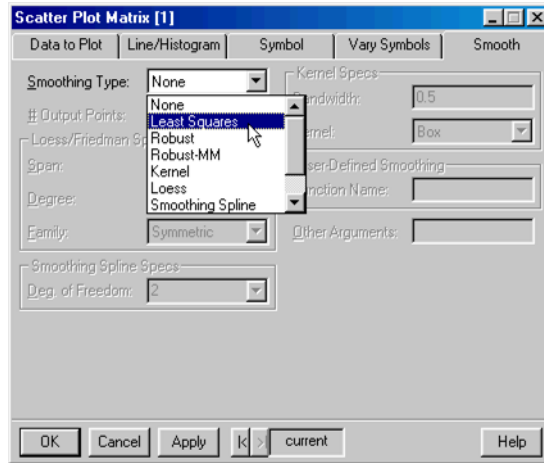


図 3.32 : Scatter Plot Matrix ダイアログの Smooth ページ


3. いろいろなタイプの線と平滑化パラメータを試してみてください。終了したら **OK** をクリックし、ダイアログを閉じてください。ウィンドウをすべて閉じます。


センサのデータ

sensors データは、たとえば汚染監視のために空気中のガスに対して様々な値に反応するように設計された、異なる 8 つの半導体の反応値を含んでいます。データは、80 の観測値を含みます。問題は、複数のセンサの追加によって汚染物質の検出能力が実際に高まるかどうかです。

バブルカラー・プロット

まず、 x 値、 y 値、記号のサイズ、記号の色を使って、各変量を表示するバブルカラー・プロットを作成します。これは、1 つの散布図で 4 つの変量を表すことができます。

1. **Select Data** ダイアログを使ってデータ・ウィンドウに sensors データを表示させます。
2. 最初の 4 列のデータを選択します。
3. **2D プロットパレット** の **バブルカラー** ボタン  をクリックします。通常の散布図と同じように、 v_1 が x 軸値として使用され、 v_2 が y 軸値として使用されます。記号のサイズは、 v_3 の値で変化します。同様に、記号の色は、 v_4 の値で変化します (v_4 が大きいほど明るい値になります)。

- 次に、グラフツールバーの**色凡例の自動作成**ボタン  をクリックして、 v_4 の値を色と関連づける凡例を追加します。(Line/Scatter Plot ダイアログの **Vary Symbols** ページで記号のサイズと色を変更することができます。) 得られたグラフを、図 3.33 に示します。

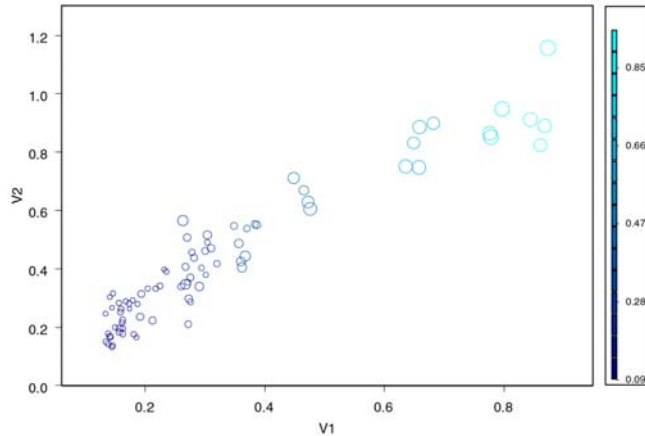




図 3.33 : sensors データのバブルカラー・プロット

このデータセットでは、 $v_1 \sim v_4$ の 4 つの変量と一緒に増減する傾向があります。たとえば、グラフの右上の方に行くほど記号が大きく明るくなります。

グラフ内の 複数プロット

データの関係と比較するもう 1 つの有効な方法は、1 つのグラフに複数の散布図を作成することです。そうすることにより、同じ x 変量を共用する多数の y 変量がプロットされます。

- sensors データの最初の 4 列を選択します。
- 2D プロットパレットの**散布図**ボタン  をクリックします。1 つのグラフ上に 3 種の散布図が作成されます。

3. 凡例の自動作成ボタン  をクリックして凡例を追加します。得られたグラフを、図 3.34 に示します。

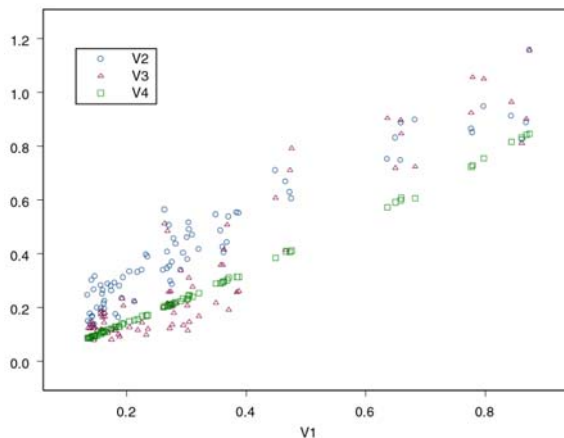




図 3.34 : sensors データのグラフ内の複数プロット

個別パネルでのプロット

x 軸は共通の個別のパネルに 3 つのプロットを表示させることもできます。

1. グラフの内側（点の上ではない）をクリックして、グラフを選択します。
2. グラフツールバーのグラフツールボタン  をクリックして、グラフツールパレットを開きます。
3. グラフツールパレットのパネル分割 (Y) ボタン  をクリックします。個別のパネルに 3 つの散布図が作成されます。

4. 凡例を選択し **DELETE** キーを押して、凡例を削除します。得られたグラフを、図 3.35 に示します。

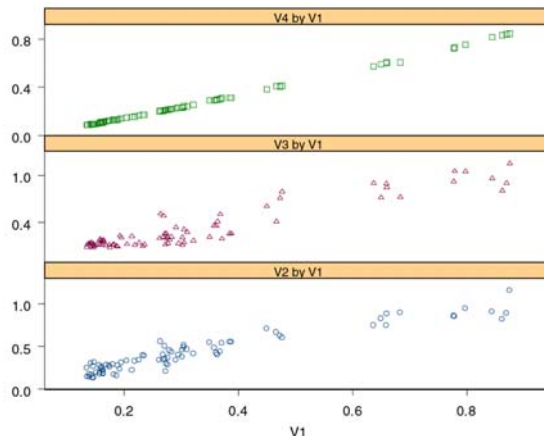



図 3.35 : 個別のパネルに描いた sensors データの複数プロット

V1 と V4 の関係が非常に強いことに注意してください。点はほぼ直線に沿って並んでいます。V1 と他の 2 つの変量 (V2 と V3) との関係には、それよりも大きな変動があります。

対散布図

今度はすべての変量を表示するために、対散布図を作成します。

1. データ・ウィンドウの左上角をクリックして、sensors データのすべての列を選択します。

- 2D プロットパレットの対散布図ボタン  をクリックします。得られたグラフを、図 3.36 に示します。

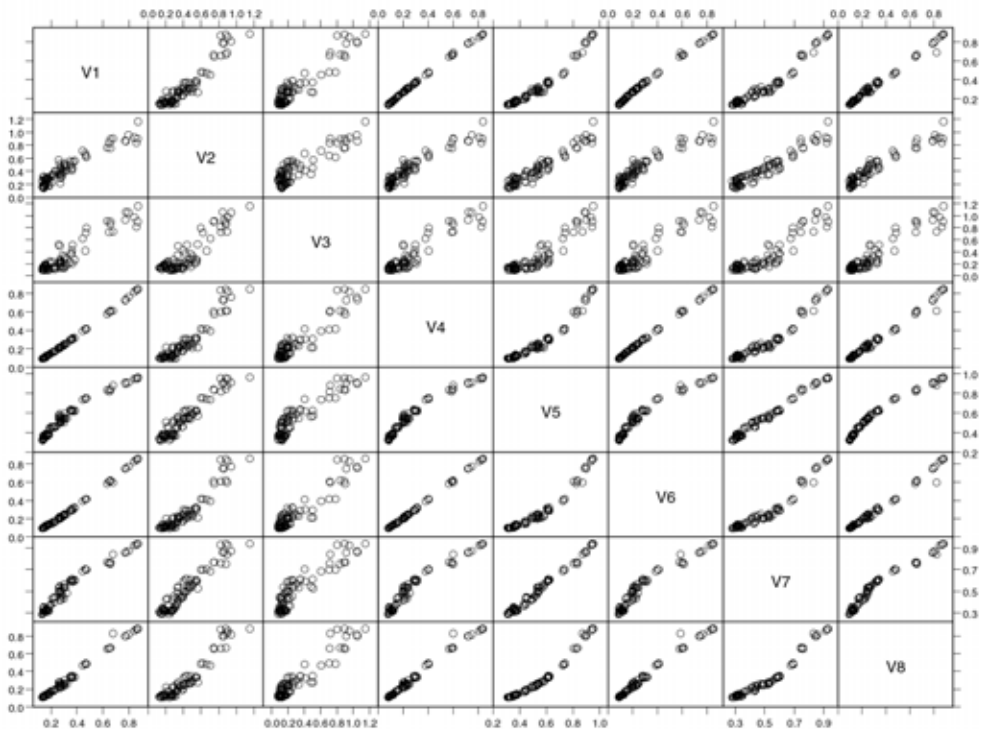


図 3.36 : 8 つの異なる汚染センサの対散布図

グラフをよく見てください。センサの汚染検出の能力が異なる場合は、ある程度非線形の関係が現れることが予想できます。センサ 4 と 5、5 と 6、6 と 7、7 と 8 の関係をよく見てください。非線形関係（線の曲がり）が分かりますか。これらのセンサのうちいくつかはガスの検出能力が異なり、したがって複数のセンサを追加するとガスの検出能力を高めることができると思われます。

Trellis グラフ


Trellis グラフでは、条件付けによってデータの様々な変量の関係を表わすことができます。複数の変量に基づくデータがあるとしたら、2つの変量のプロットが、第3の“条件付け”変量の変動によって、どのように変化するかを見たいことでしょう。Trellis グラフを使用すると、条件付け変量の区間に分けられた、元データのサブセットを表す一連のパネルに、データをグラフとして表示させることができます。

Trellis グラフを使用すると、様々なグラフの条件付けを行うことができます。

エタノールの データ

データセット例 ethanol は、自動車の排気中の汚染物質の測定値を含んでいます。データは、排気中の窒素酸化物を分析するために収集されました。実験は、エタノール燃料の単気筒エンジンで行われました。空気と燃料の混合気の濃度測定値である当量比 (E) と、エンジンに設定する圧縮比 (C) のエンジンの2つの因子を調査しました。実験は88回行いました。次に、窒素酸化物 (NOx) と圧縮比の関係調べます。

Loess プロットを作成する

1. **Select Data** ダイアログを使ってデータ・ウィンドウに ethanol データを表示させます。
2. C 列を選択します。CTRL キーを押し、列 NOx を選択します。
3. **2D プロットパレットの局所重み付き最小二乗ボタン**  をクリックすると、図 3.37 に示したような局所平滑化の散布図が作成されます。

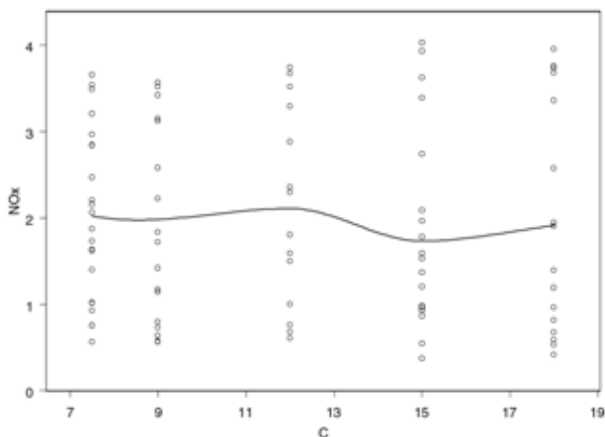


図 3.37 : ethanol データセットの圧縮比と窒素酸化物の関係 (デフォルトの平滑化)

- 現在のグラフシートと ethanol データ・ウィンドウを除くすべてのウィンドウを閉じてください。次に、メインメニューから、ウィンドウ ▶ 縦に並べるを選択してください。

平滑化パラメータを調整する

Loess プロット線の任意の点を右クリックして、グラフのショートカットメニューを表示させ、**Smooth/Sort** を選択します。**Loess/Friedman Specs** グループで、**Span** ボックスに **1** と入力し、**OK** をクリックします。図 3.38 に示すように、**Span** パラメータを大きくすると線が滑らかになります。

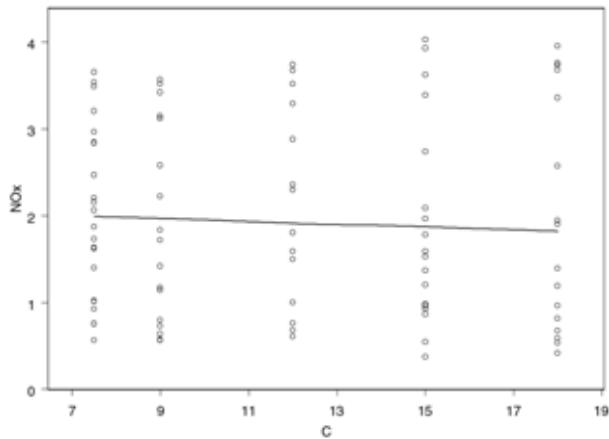


図 3.38 : ethanol データセットの圧縮比と窒素酸化物の関係 (Span が 1 のとき)

Trellis グラフ

圧縮比に対する窒素酸化物の依存性はほとんど見られません。しかしこの散布図では、データの点ごとに E が異なることが無視されています。次のステップで、プロットに E による条件付けを行い、関係をさらに詳しく調べます。

1. **データ・ウィンドウの列 E** を選択します。カーソルを列のデータセル部分に移動させます。(マウスポインタが下向きの黒矢印から左斜め上向き白矢印に変わります) マウスボタンを押したままにします。ボタンを押したまま、カーソルをグラフの最上部までドラッグします。カーソルをグラフにドラッグすると、四角形のドロップターゲットが現われます。



2. カーソルが点線の四角形の中にある状態で、マウスボタンをはなしてください。パネルに、4つの E のレベルごとに C と NOx の関係を表すグラフが再描画されます。

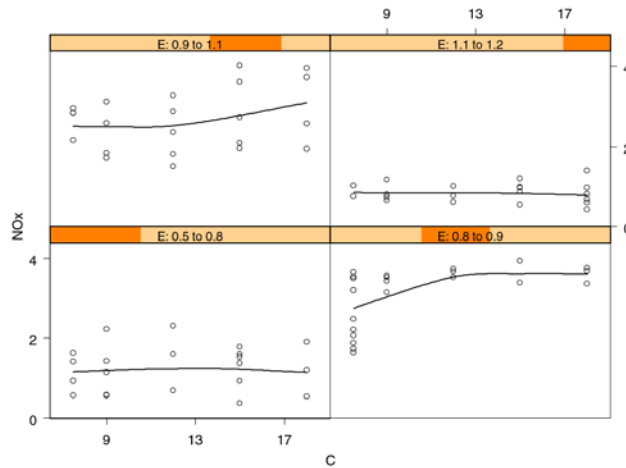


図 3.39 : 当量比で条件付けた圧縮比と窒素酸化物の関係 (デフォルトのとき)

条件付けの規則を修正する

デフォルトでは、条件付け変量によってすべてのデータはグループごとにほぼ同数となるような 4 つのグループに分けられます。さらに多数のグループに分けたり、条件付けのための他の値を変更することもできます。

1. グラフ内の空白部分を右クリックして、グラフのショートカットメニューを表示させ、**Multipanel** を選択します。
2. **Continuous Conditioning** グループで、**# of Panels** を **9** に、**Frac. Shared Pts** を **0.25** に設定します。
3. **Layout** グループで、**# of Rows** を **2** に設定します。
4. ダイアログの **Position/Size** タブをクリックします。**Aspect Ratio** を **2.5** に設定し、**OK** をクリックします。得られたグラフを、図 3.40 に示します。

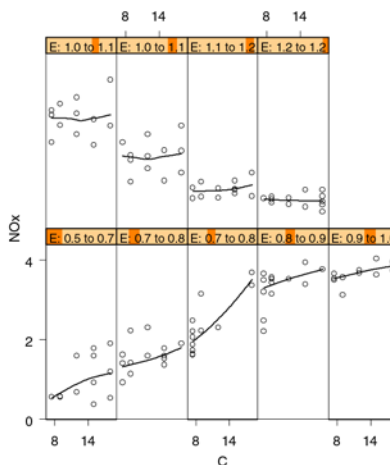



図 3.40 : 当量比で条件付けした圧縮比と窒素酸化物の関係

このとき、Trellis グラフは 2 行 9 つのパネルに表示され、各パネルの縦横比は 2.5 です。各パネルに使用される E の範囲は、重ね合わせを 25%としたデフォルトの等カウント法 (1 つのパネルに同数の観測値を含む) によって決定されます。このアルゴリズムによって、各区間の端点がデータから計算されます。最も低い区間の左側の端点がデータの最小値であり、最も高い区間の右側の端点がデータの最大値です。端点は区間内に含まれる点の数ができるだけ等しくなるように、かつ隣り合う区間によって共有される点の割合が目標とする割合にできるだけ近くなるように選択されます。通常重なり合う部分の範囲では、重なり合わない部分と比べて、不均一性の検出感度が高くなります。

条件付けしたグラフでは、 E の値が小さいとき、窒素酸化物と圧縮比との間に明らかな正の関係があります。 E の値が大きいときは、勾配が 0 に近くなります。各パネルのパターンは直線に見えます。

グラフの種類を変更する

パターンが直線のように見えるため、回帰直線を使って Trellis グラフを再描画しましょう。

1. グラフの任意の点をクリックし、グラフを選択します。
2. **2D プロットパレットの回帰直線ボタン**  をクリックします。グラフが、回帰直線を使って再描画されます。

大麦のデータ


データセット例 `barley` は、大麦の収穫高の研究を行う農業実地試験によるデータを含んでいます。ミネソタ州の 6 つの生産地で、2 年間のそれぞれの年に大麦の 10 の品種を育てました。データは生産地、品種および年のすべての組合せによる収穫高です。したがって、 $6 \times 10 \times 2 = 120$ の観測値があります。

この大麦の実験は、1930 年代に行われました。データは最初、実験者による 1934 年発行の報告書で公表されました。それ以来、データの解析が繰り返行われました。R.A.Fisher は、彼の代表的な著書『*The Design of Experiments*』の中で、5 つの生産地のデータを提示しています。この本の出版によりこのデータは有名になり、新しい統計的方法を実証するために、多くの人々が次々とこのデータを解析しました。


1990 年代の初め、AT&T（現在 Lucent Technologies）の Bill Cleveland は Trellis グラフを使用して、再びデータを解析しました。その結果は、非常に驚くべきものでした。50 年間にわたって何度も解析されたにもかかわらず、データの重要な特徴が見いだされていなかったのです。次に、基本的な解析内容を説明します。

Trellis グラフ

大麦の収穫高が、他の変量の組合せによってどのように異なるかについての探索を行います。Trellis グラフは特に、共変量の影響とその交互作用を表すために役立ちます。条件付けをしたカラープロットを使用して、第 4 と第 5 の変量を表します。

1. **Select Data** ダイアログを使ってデータ・ウィンドウに barley データを表示させます。
2. 列 yield を選択します。次に、CTRL キーを押して variety、year および site を選択します。選択した変量の順序は、グラフでどのように用いられるか決定します。第 1 の変量 (yield) は x として、第 2 の変量 (variety) は y として使用され、第 3 の変量 (year) は記号の色を決定するために用いられます。最後の変量 (site) は条件付け変量として用いられます。
3. 次に、**2D プロット** パレットを開きます。標準ツールバーの**条件モードのオン・オフ** ボタン  をクリックして、条件付けモードに切り換えます。(2D プロットパレット上のそれぞれのアイコンの上部にある小さい黄色のバーは、条件モードがオンであることを示しています。)
4. **条件の列の数** が 1 に設定されていることを確認してください。



5. **2D プロットパレットのカラープロットボタン**  をクリックします。Trellis グラフは、生産地で条件付けた各品種の大麦収穫高を示しています。1931年と1932年の収穫高は、異なる色で表されます。得られたグラフを図 3.41 に示します。

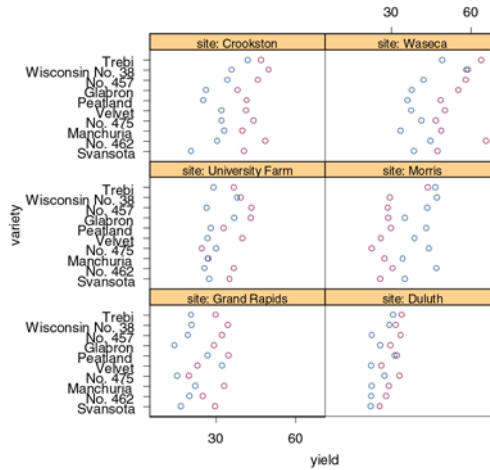



図 3.41 : 1931 年と 1932 年の大麦収穫高のデフォルトで表示される Trellis プロット

6. 条件付けモードをオフにし (条件付けモードボタン  をもう 1 回クリックする)、新しいグラフシートを最大化し、**2D プロットパレット** を閉じます。


グラフのレイアウトを変える

生産地による収穫高を比較しやすくするために、パネルのレイアウトに 3 つの変更を加えます。まずパネルを 1 列に積み重ね、各パネルに示された収穫高データの中央値によってパネルを並べ替え、各パネルの縦横比を 0.5 に設定します。

1. グラフ内の空白部分を右クリックして、グラフのショートカットメニューを表示させ、**Multipanel** を選択します。
2. **Conditioning Columns** グループで、**Order Type** を **Median of X** に設定します。
3. **Layout** グループで、**# of Columns** を **1** に設定します。
4. ダイアログの **Position/Size** タブをクリックします。**Aspect Ratio** を **0.5** に設定し、**OK** をクリックします。

凡例を追加する

次に、グラフに最終的にいくつか手を加えます。2つの年を凡例で示し、記号のスタイルならびに記号の色を変更します。

- グラフツールバーの**凡例の自動作成**ボタン  をクリックします。凡例が自動的に作成され、グラフに表示されます。凡例を選択し（凡例の境界のすぐ内側でクリック）希望の位置までドラッグして、凡例を位置決めすることができます。

記号を変更する

1. プロット上の任意の点を右クリックして、そのショートカットメニューを表示させ、**Vary Symbols** を選択します。
2. **Vary Symbols** ページで、**Vary Style By** を **z Column** に設定し、色と点の両方のスタイルが、年ごとに異なるものになります。（点のスタイルと色に対するデフォルトを変更することもできます。詳細は、第13章「S-PLUS セッションのカスタマイズ」を参照してください。）
3. **OK** をクリックします。新しい点のスタイルが反映され凡例が更新されたことに注目してください。最終的な Trellis プロットは、図 3.42 に示したものと似ています。

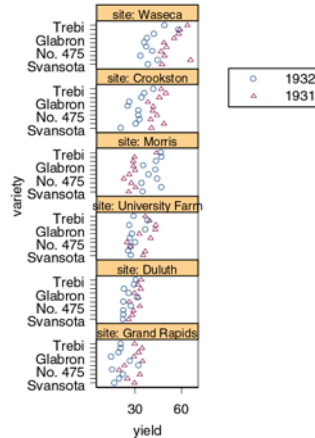


図 3.42: 1931 年と 1932 の大麦収穫高のレイアウトを変えた Trellis プロット

次に、これまで長い間検出されなかった矛盾点がグラフに現れていることを調べます。それは、**Morris** パネルに現われています。他のすべての生産地では、1931 年の全体の収穫高が 1932 年よりもかなり高くなっていますが、**Morris** では逆になっています。さらに最も重要なことは、**Morris** で 1931 年から 1932 年の増加分が、他の生産地で 1931 年から 1932 年に減少した分と、

ほぼ一致していることです。病気や局地的な異常気象などの自然現象により奇妙な偶然の一致が起こったか、Morris での年度が間違っただけになったかのいずれかです。さらに多くの Trellis グラフ、データの統計モデル、実験に関するいくつかの背景調査を行って、データが間違っているという結論が得られました。しかし、そのような結果につながる驚くべき発見ができたのは、今回の図 3.42 に作成したような Trellis グラフのおかげです。

3 次元グラフ

これまでの節で説明した散布図、対散布図、Trellis プロットはすべて、3次元データを表示するために使用することができます。さらに、3次元散布図、鳥瞰図、等高線図を使用することができます。これらのグラフは、特に1つの変量を他の2つの変量の関数としてプロットすることも含め、点を3次元空間でプロットするのに適しています。

S-PLUS では、これら3変量のグラフのいくつかは2次元グラフ、それ以外は3次元グラフとみなしています。

- 3次元散布図
- 鳥瞰図
- 等高線図

最初の2つのグラフは、点を3次元空間で表すため、3次元の軸を使用する3次元グラフと考えられます。等高線図は、第3の変量を等高線または色の塗りつぶしで表す2次元グラフです。

3次元散布図


3次元散布図は、変量値を3次元空間の位置として表します。

スライスボール・データ

最初に、データセット例 `sliced.ball` に含まれる一組のシミュレーションによるデータを調べます。これから発見しようとする構造的特徴の他は、球内に点が均一に分散しています。

まず、このデータで対散布図と Trellis プロットを作成し、次に3次元散布図を作成します。

対散布図

1. **Select Data** ダイアログを使って、**データ・ウィンドウ**に sliced.ball データを表示させます。
2. 3つの列をすべて選択します。
3. **2D プロットパレット**の**対散布図**ボタン  をクリックします。図 3.43 に示すように、各パネルに一見ランダムなデータのボールが表示されます。

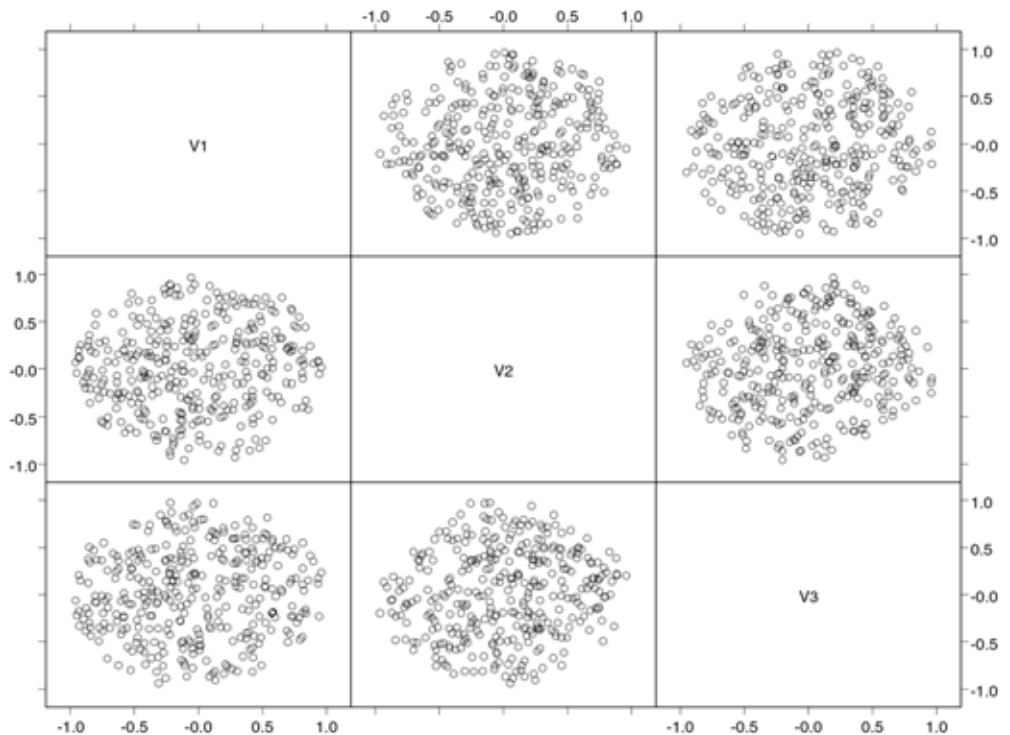




図 3.43 : スライスボール・データの対散布図

データのどの2つの列の間にも明らかな関係はないので、2変量プロットは、データが球内にランダムに分散していることを示しています。他に関係は見うけられません。

Trellis プロット

2変量プロットの限界は、変量間の交互作用を見ることができないことです。たとえば、V1 と V2 との関係は V3 によってどのように変化するでしょうか。Trellis プロットを使用すると、そのような交互作用を調べることができます。

1. 標準ツールバーの**条件モードのオン・オフボタン**  をクリックして、条件付けモードにします。(2D プロットパレットのそれぞれのアイコンの上部にある小さい黄色のバーは、条件付けモードがオンになっていることを示しています。)
2. **条件の列の数**が 1 に設定されていることを確認してください。こうすると、最後に選択された列の値でグループ化されたデータで、複数のパネルが作成されます。
3. V1 を選択し、次に CTRL-クリックして列 V2 および V3 を選択します。
4. 2D プロットパレットの**散布図ボタン**  をクリックします。得られたグラフを、図 3.44 に示します。

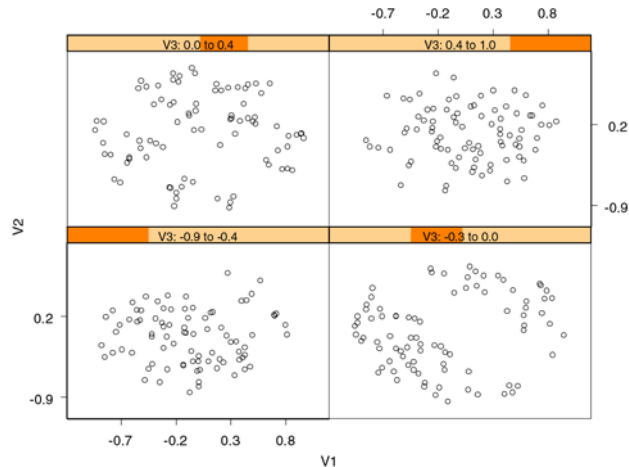




図 3.44 : スライスボール・データの Trellis プロット

5. 標準ツールバーの**条件モードのオン・オフボタン**  をもう 1 回クリックして条件付けモードをオフにします。(この操作を省略すると、後のプロットで条件付けモードが意図せずオンになっていることがあります。)

データ点は、4つのグループに分けられます。左下のパネルには、V3の値が-0.9~-0.4のデータ点が示されます。次に高い値を有するV3のデータ点が、そのパネルの右側に示されます。下2つのグラフで対角線方向に隙間があることに注意してください。これは、左上のパネルでも明らかです。これは、データが実際には球全体にランダムになっていないことを示します。

3次元散布図

さらに調べるために、3次元散布図を作成することができます。これは、データが3次元空間中の点を表すときに使用する最も一般的なグラフです。

1. データ・ウィンドウの3つの列をすべて選択します。
2. 3Dプロットパレットの**散布図**ボタン  をクリックします。得られたグラフを、図 3.45 に示します。

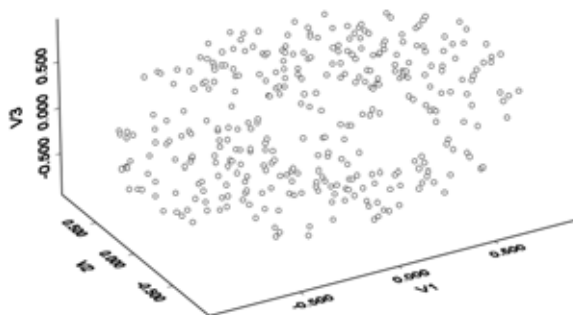


図 3.45 : スライスボール・データの3次元散布図

このグラフは、前のグラフと違い、3次元空間内で点の位置を表すのに、3次元の軸を使用することに注意してください。

3次元グラフを回転させる


1つの3次元散布図は、雲状の点で1つのグラフを表します。全面的な構造を理解するためには、多数の方向から雲状の点を見ることが有効です。軸を回転させて、点を様々な角度から見るすることができます。

1. グラフ領域内の空白の場所をクリックして、3次元作業ボックスを選択します。4つの緑色の円と1つの緑色の三角形が現われます。(1つの緑色のノブしか現われない場合は、作業ボックスではなく点をクリックした状態です。もう1回試してください。)

2. 緑色の円の1つを画面上横にドラッグします。境界ボックスが現れるので、マウスを回転する方向にドラッグします。マウスをはなすと、グラフは新しい角度から再描画されます。緑色の三角形を上下にドラッグすると、グラフを垂直方向に回転させることができます。グラフを回転させてみてください。

複数のパネル

異なる角度から雲状の点を同時に見ることが有効な場合があります。これは、各パネルにおいてデータを様々に回転させた複数のグラフを使って行うことができます。

1. グラフの内側をクリックしてグラフ領域を選択します。グラフが選択されたことを示す8つの緑色のボックスがグラフシートのまわりに現われます。
2. **3Dプロットパレットの6つのパネルに回転角で分割ボタン**  **をクリックします。得られたグラフを、図 3.46 に示します。**

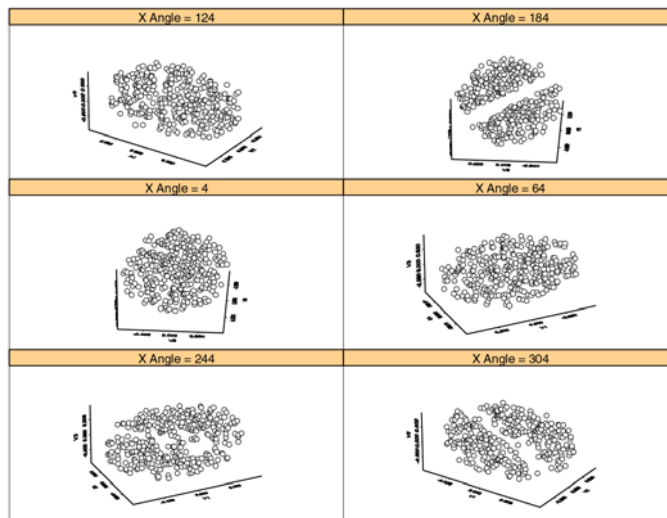


図 3.46 : スライスボール・データの複数パネル 3次元散布図

複数パネル 3次元散布図を使用すると、スライスボール・データ内の不連続部分をはっきり見つけることができます。データを 3次元で表示すると、2変量プロットでは表示されない角度で見ることができます。これは、より高い次元の構造を発見するのに役立ちます。

このようなデータを調べるもう 1 つの方法は、ブラッシングとスピンを使用することです。これについては、118 ページの「ダイナミックグラフ」で説明します。

鳥瞰図


鳥瞰図は、ある変量の値を他の 2 つの変量の組み合わせにより表示する 3 次元の曲面です。鳥瞰図で使用されるデータは、2 次元グリッド上の値のことがあります。使用できるデータがグリッド上にない場合は、補間してグリッド上に面をあてはめてからプロットします。

Surface データの例

データセット例 `exsurf` は、グリッドの座標を表す 2 つの列と、それらの関数として考えられる第 3 の列を含みます。`exsurf` のヘルプファイルに、表現される関数式の説明があります。

鳥瞰図を作成する

最初に、これらのデータの鳥瞰図を作成します。

1. **Select Data** ダイアログを使ってデータ・ウィンドウに `exsurf` データを表示させます。
2. データ・ウィンドウ内の 3 つの列をすべて選択します。
3. **3D プロットパレット** の鳥瞰図ボタン  をクリックします。得られたグラフを、図 3.47 に示します。

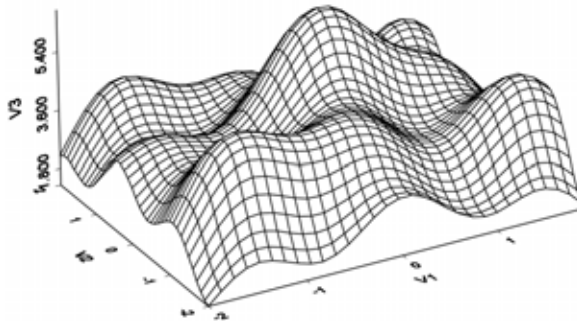




図 3.47 : `exsurf` データの鳥瞰図

4. 3 次元散布図と同じ手順で、グラフを回転させてあらゆる角度から調べてください。

塗りつぶし面を作成する

次に、塗りつぶし面を作成します。

1. 3次元メッシュ内の線をクリックしてグラフを選択します。
2. 3Dプロットパレットの16色イメージ鳥瞰図ボタン  をクリックして、鳥瞰図をz値で変化する色付きの塗りつぶし鳥瞰図に変換させます。
3. 凡例を追加するときは、グラフツールバーの色凡例の自動作成ボタン  をクリックします。得られたグラフを、図 3.48 に示します。

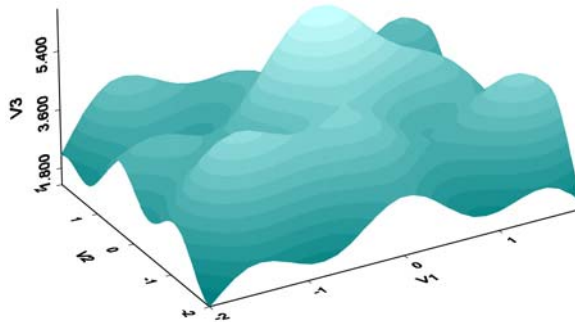



図 3.48 : exsurf データの塗りつぶし鳥瞰図

等高線図

等高線図は、3次元データを平面で表した2次元表現です。これは、しばしばグリッド上で収集されたデータに使用されます。使用できるデータがグリッド上にない場合、補間してグリッド上に等高線をあてはめてからプロットされます。

等高線図を作成する

- 1 exsurf データ・ウィンドウ内の3つの列をすべて選択します。

2. **2D** プロットパレットの**等高線**ボタン  をクリックします。得られたグラフを、図 3.49 に示します。

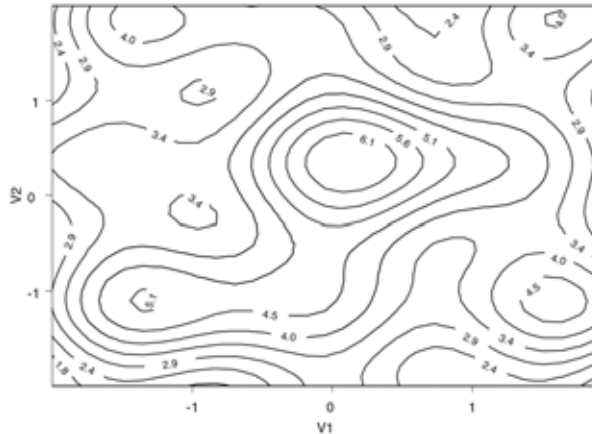



図 3.49 : exsurf データの等高線図

等高線図上の線は、同じ値の位置を示します。等高線図は、等高線地図と同じように読みます。線は、様々な領域における表面の勾配と同様に最小値と最大値を導き出します。

カラー等高線図

等高線図を理解しやすくするために、それぞれの位置に z 変量の大きさを示す色を追加します。

1. 任意の等高線をクリックしてグラフを選択します。
2. **2D** プロットパレットの**カラー等高線**ボタン  をクリックします。グラフが、カラー等高線図に変化します。

3. グラフを右クリックしてショートカットメニューを表示させ、**Contour/Fills** を選択して **Contour Plot** ダイアログを開きます。**Fill Type** が **2 Coilor Range** であることを確認します。色の設定は同じ場所でも変更可能です。デフォルトと等高線図には、図 3.50 に示すように **Blue** から **Lt Cyan** までの色が塗られます。

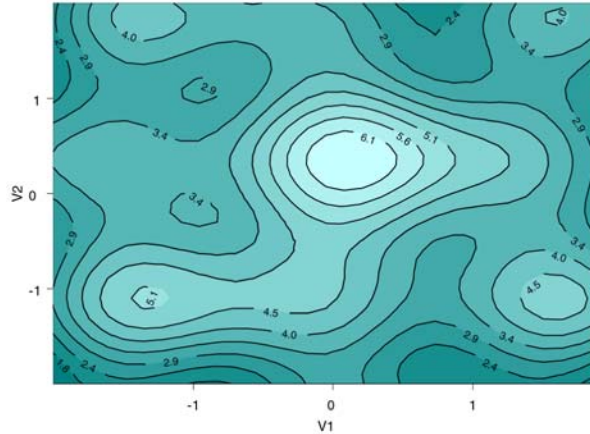



図 3.50 : exsurf データのカラー等高線図

4. 凡例を追加するときは、**グラフツールバーの色凡例の自動作成ボタン**  をクリックします。

ダイナミック グラフ

対散布図と 3 次元散布図が有効なデータに対しては、対話式のダイナミックグラフを生成することができます。これらの図では、データの対散布図を表示させ、点グループを選択してラベルをつけ、データを対話式に回転させて、多次元データをダイナミックに見せることができます。


スライスボール のデータ

データ例 `sliced.ball` を使って、S-PLUS のダイナミックグラフの使用方法について説明します。このデータセットは、球内の点を表す 3 次元データセットです。また、2 次元の対散布図でははっきり現れない構造を持っているようでした。

1. **Select Data** ダイアログを使ってデータ・ウィンドウに `sliced.ball` データを表示させます。
2. データを眺めます。何かおかしな点はありませんか？

- データの対散布図を作成します。今度は、何か通常と異なるものや、何かはっきりしたパターンまたはその他の特徴的なことはありますか。ほとんどの人には、通常と異なることは何も見えず、データがランダムなデータ点の雲のように見えるでしょう。
- 次に、データの 3 次元散布図を作成します。

データ点の選択

- グラフツールパレットの矩形内のデータ選択ボタンを  クリックします。
- ポインタをクリックし、3 次元散布図内のいくつかの点にドラッグします。ドラッグすると、四角形の“枠”が現れます。マウスボタンをはなすと、枠内のすべての点がハイライトされます。
- 次に、対散布図グラフシートのタイトルバーをクリックしてフォーカスを当てます。このとき 3 次元散布図内でハイライトさせた点が、それぞれの対散布図パネルでもハイライトされていることに注目してください。
- 対散布図パネルの空白部分でポインタをクリック アンド ドラッグして枠で囲みます。前に選択された点のハイライト表示が取り消されます。
- 点を選択したり選択解除したりして、矩形内のデータ選択ツールを試してください。

これらのデータにブラッシングを使用して、意味のあるパターンを見つけることは難しいでしょう。データセットによっては、ブラッシングがきわめて有効な探索的データ解析ツールになることもあります。

データのスピンの

次に、ダイナミックな 3 次元散布図を用います。これは、データを対話的に回転させることができます。これにより、データをあらゆる角度から見ることができ、データをさらに良く理解することができます。

- グラフ領域内の空白部分をクリックして、3 次元作業ボックスを選択します。4 つの緑色の円と緑色の三角形が現われます。(1 つの緑色のノブしか現われない場合は、作業ボックスではなくグラフをクリックしています。もう 1 回試してみてください。)
- 次に、円 (縦軸に関する回転) または三角形 (画面と平行な横軸に関する回転) のどれかをドラッグすることによって、雲状の点を回転させます。回転させながら、その雲を調べてください。何回か回転させると、何かお気づきになるでしょう。図 3.51 に示したように、このランダムに見える雲状の点の中に、抜けているデータの断層があります。

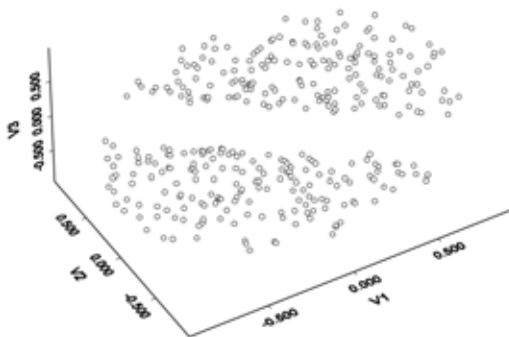


図 3.51 : 抜けている点を明らかにするために回転させた sliced.ball データセット

点を 3 次元で表示させることにより、2 変量プロットでははっきりしない構造を見つけることができます。

第4章

グラフの作成

はじめに	123
プロットパレット	123
Insert Graph ダイアログ	124
Plot Properties ダイアログ	124
データをグラフに合った形に構築する	124
1 次元データのプロット	126
箱型図 (箱ひげ図)	126
QQ プロット	127
確率プロット	128
確率分布つきヒストグラム	129
円グラフ	130
点グラフ	131
棒グラフ	132
パレートグラフ	134
2 次元データのプロット	135
散布図と線グラフ	135
曲線あてはめプロット	137
非線形曲線あてはめプロット	138
平滑化プロット	139
シンボル付き線グラフ	141
Y 系列プロット	141
XY ペア線グラフ	142
類別箱型図	143
類別棒グラフ	143
積み上げ棒グラフ	145
極座標プロット	146
多次元データのプロット	147
3 次元の散布図と線グラフ	147
バブルプロットとカラープロット	148

第4章 グラフの作成

バブルカラー・プロット	149
High-Low プロット	149
ローソク足プロット	150
エラーバー・プロット	151
ベクトル図	153
エリアチャート	153
対散布図（散布図 行列）	154
等高線図とイメージ	155
鳥瞰図／3次元棒グラフ	156
コメント図	158
スミスチャート	159
投射プロット	161
Trellis グラフ	162

はじめに

データの視覚的な探索やモデルの評価のために使用する“とりあえずの”グラフや、社内で非公式に共有するためのグラフ、論文やマーケティング材料を出版するためのある程度見栄えのするグラフなど、様々な目的でグラフを作成しなければならないことがあります。S-PLUS は、そのようなすべての用途のため様々な種類のグラフを提供します。この章では、S-PLUS で作成することができる様々なグラフについて説明します。

プロット パレット

2D プロットパレット、**3D プロットパレット**および**エクストラプロットパレット**には、プロットを素早く作成するためのボタンがあります（下の図 4.1 を参照）。グラフを作成するときは、**データ・ウィンドウ**または**オブジェクト・エクスプローラ**からデータの列を選択してから、パレットボタンをクリックするだけです。

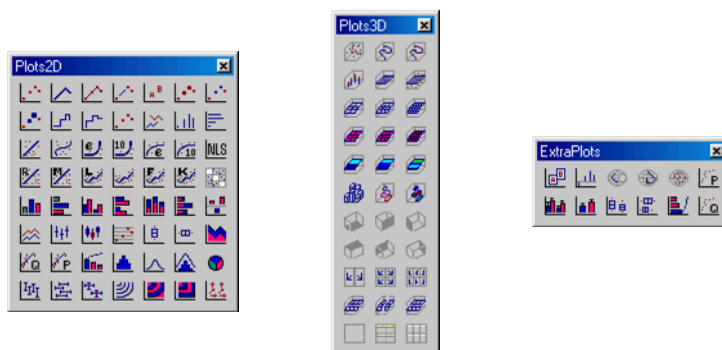


図 4.1 : 2D プロットパレット、3D プロットパレットおよび
エクストラプロットパレット

2D パレットと 3D パレットの違いは、グラフが 2 つの軸で作成されるか 3 つの軸で作成されるかによります。対散布図、バブルカラー・プロット、等高線図などの多くの 2 次元グラフは、2 次元以上のデータを表現することができます。この章では、主にグラフをデータの次元ごとにまとめて説明します。

Insert Graph ダイアログ

図 4.2 に示すように、**Insert Graph** ダイアログでグラフの種類を選択することによって、任意のグラフを作成することができます。

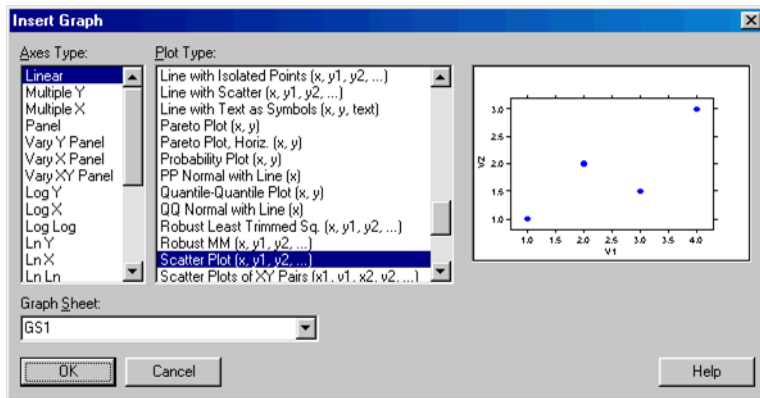


図 4.2 : Insert Graph ダイアログ

Insert Graph ダイアログを開くには、以下の操作のいずれかを行います。

- メインメニューから、**グラフ**を選択し、**2D プロット・グラフ**、**3D プロット・グラフ**または**マルチパネルグラフ**を選択します。
- **データ・ウィンドウ**を開き、メインメニューから**挿入 ▶ グラフ**を選択します。

Plot Properties ダイアログ

既存のグラフをダブルクリックするか、最初にデータを選択せずに **Insert Graph** ダイアログでグラフを作成すると、グラフの種類によって固有のプロットプロパティ・ダイアログが現れます。このダイアログによってグラフを作成したり修正したりすることもできます。プロットプロパティ・ダイアログの詳細は、オンラインヘルプを参照してください。

データを グラフに 合った形に 構築する

いくつかのグラフでは、データを特定の方法で構築する必要があります。以下の節では、様々なグラフごとにサンプルのデータセットを示します。しかし、多くのグラフでは、データをいくつかの異なる方法で持つことができます。

たとえば、類別箱型図を作成するには、データをグループごとに積み上げても、横に並べても、どちらの形式でもグラフ化可能です。いずれのデータ形式でも同じ類別箱型図が作成されます。

グループ積み上げデータ（図 4.3 を参照）では、 x 列は整数列であり、それぞれの y 値をグループに割り当て、 x 軸方向の箱の配置を決定します。データがどちらの方法で構築されているときでも、類別箱型図を作成するには x 列と y 列を選択するだけです。

	1	2	3
	x	y	
1	1.00	10.91	
2	1.00	16.45	
3	1.00	17.18	
4	1.00	12.20	
5	1.00	9.22	
6	2.00	10.37	
7	2.00	16.20	
8	2.00	14.00	
9	2.00	9.00	
10	2.00	12.51	
11	3.00	13.00	
12	3.00	10.26	
13	3.00	8.71	
14	3.00	12.09	
15	3.00	7.70	

図 4.3 : 類別箱型図を作成するためのグループ積み上げデータ

グループを横に並べたデータ（図 4.4 を参照）では、 x 列が、複数の y 列のデータの類別水準を決定します。データがこの方法で構築されているとき、類別箱型図を作成するには x 、 $y1$ 、 $y2$ 、 $y3$ 、 $y4$ および $y5$ 列を選択してください。

	1	2	3	4	5	6	7
	x	y1	y2	y3	y4	y5	
1	1.00	10.91	15.45	17.18	12.20	9.22	
2	2.00	10.37	16.20	14.00	9.00	12.51	
3	3.00	13.00	10.26	8.71	12.09	7.70	

図 4.4 : 類別箱型図を作成するための複数 y 形式データ

所定のグラフに適したデータ構造を確認するには、**Insert Graph** ダイアログを開きます（図 4.2 を参照）。以降の節では、特定のグラフを作成するための様々な方法をご紹介します。

1次元データのプロット


箱型図 (箱ひげ図)

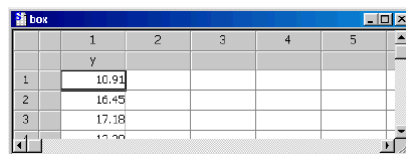
箱型図は、中央値、上下側の四分位数、最小値、最大値、および外れ値（存在すれば）などの1次元データ分布の基本統計量の位置を示します。

箱型図を使用すると、データ分布の位置、規模（幅）、および大体の形を素早く把握することができます。たとえば、箱型図の上下の四分位位置が中央値からほぼ同じ距離にある場合、データの分布は、中心に対してほぼ対称的です。中央値は、四角形内の横の線分で表され、四角形の上と下の辺は、上下の四分位数を表します。


箱型図には、単一と類別の2つのタイプがあります（類別箱型図は、この章の後の方で説明します）。単一箱型図は、1列のデータを表す1つの箱型図からなります。


一組のデータの縦箱型図を作成するには、

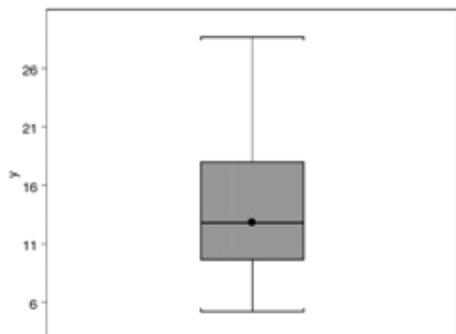
1. y列を選択します。
2. 2Dプロットパレットの  ボタンをクリックします。




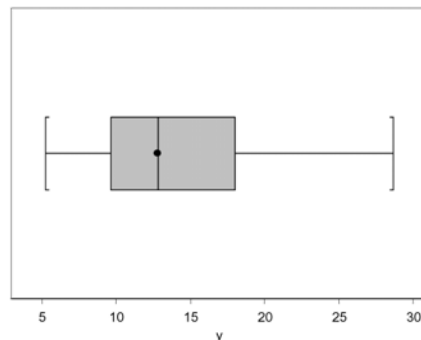
	1	2	3	4	5
	y				
1	10.91				
2	16.45				
3	17.18				
	17.20				

横箱型図を作成するときは、代わりに  ボタンをクリックします。

縦箱型図 



横箱型図 




QQ プロット

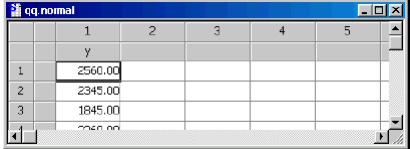
確率点-確率点プロット、略して **QQ** プロットは、データの分布を別のデータと比較したり、適用できると思われるモデルの分布関数の確率点と比較するのに有効です。後者の場合、グラフは、縦軸方向のソートされたデータ値と横軸方向に指定した分布関数の対応する確率点との関係を示します。グラフの解釈は、次のようにきわめて簡単です。

- 点の集合が直線に近い場合、推測される分布は、データにとってかなり良いモデルです。
- 点の集合が直線に近くない場合、推測される分布は良いモデルではなく、さらに良いモデルの分布を探す必要があります。


線付き正規 **QQ** プロットは、一組のデータを分布関数（デフォルトでは、正規分布）の確率点と比較するためのものです。**QQ** プロットは、2組のデータと比較するためのものであり、分布線を自動的に表示しません。

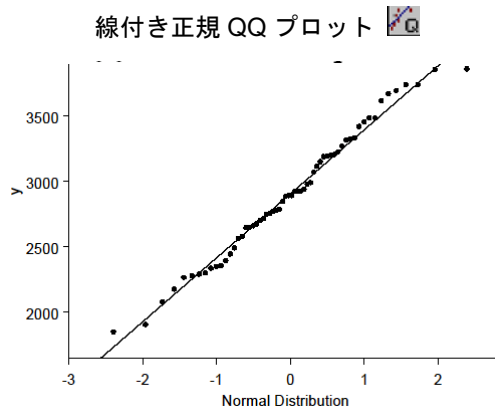
一組のデータの線付き正規 **QQ** プロットを作成するには、

1. y 列を選択します。
2. **2D** プロットパレットの  ボタンをクリックします。



	1	2	3	4	5
	y				
1	2560.00				
2	2345.00				
3	1845.00				


同じグラフを分布線なしに作成するときは、**エクストラプロット**パレットの  ボタンをクリックします。




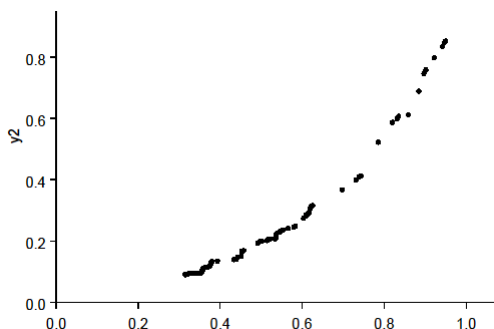
2組のデータを比較するQQプロットを作成するには、

1. y_1 に対して y_2 をプロットするときには y_1 列と y_2 列を選択します。

	1	2	3	4	5
	y_1	y_2			
1	0.07	0.59			
2	0.73	0.40			
3	0.62	0.29			

2. エクストラプロットパレットの  ボタンをクリックします。


QQプロット 




確率プロット

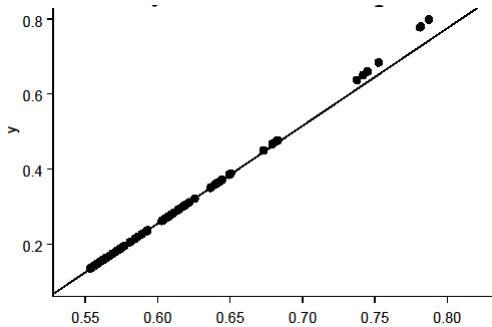

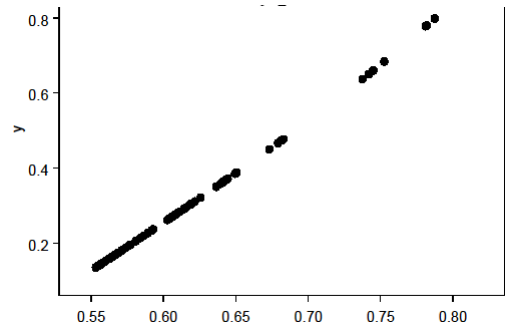

確率プロットは、QQプロットと似ていますが、データを累積確率分布関数の確率点と比較する点が異なります。確率プロットは、分布線付きでも分布線なしでも作成することができます。

一組のデータの分布線付き確率プロットを作成するには、

1. y 列を選択します。
2. 2Dプロットパレットの  ボタンをクリックします。

	1	2	3	4	5
	y				
1	0.66				
2	0.47				
3	0.35				

同じプロットを分布線なしで作成するには、エクストラプロットパレットの  ボタンをクリックします。

線付き正規確率プロット 確率プロット 




確率分布つき ヒストグラム

ヒストグラムは、一連の棒から成り立っています。各バーの幅は、値の範囲を表し、バーの高さは、その範囲内の観測値の個数を表します。ノンパラメトリック確率分布推定は、平均パラメータ μ と分散パラメータ σ^2 を有する正規確率分布のようなパラメトリックな確率分布に対して、そのような仮定をしないデータの確率密度関数（略して、確率分布）の推定値です。


確率分布を推定するのに用いるグラフは、ヒストグラムと似ているので、ノンパラメトリック確率分布密度グラフはヒストグラムの平滑化された代替とみなすことができます。確率分布つきヒストグラムは、箱型図のように大量のデータ省略を行わなくてもよいので、強力な視覚化ツールです。

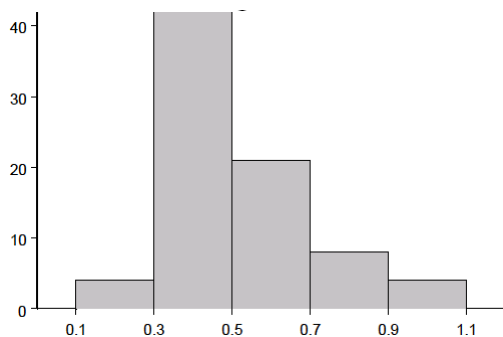
一組のデータの確率分布つきヒストグラムを作成するには、 x 列を選択します。次に、

	1	2	3	4	5
x					
1	0.64				
2	0.70				
3	0.50				

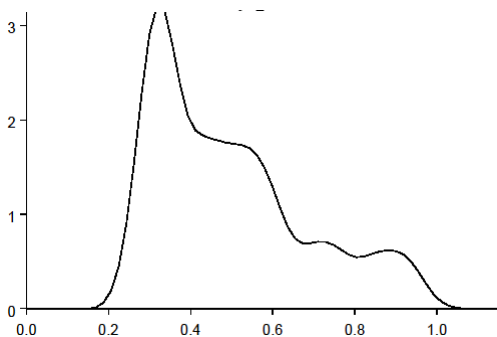
- ヒストグラムを作成するときは、**2D プロットパレット**の  ボタンをクリックします。
- 確率分布プロットを作成するときは、**2D プロットパレット**の  ボタンをクリックします。
- 確率分布つきヒストグラムを作成するときは、**2D プロットパレット**の  ボタンをクリックします。


第4章 グラフの作成

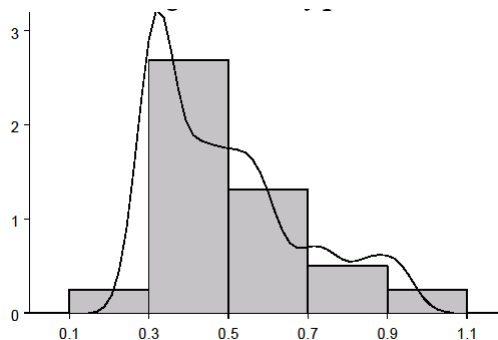
ヒストグラム 



確率分布プロット 




確率分布つきヒストグラム 

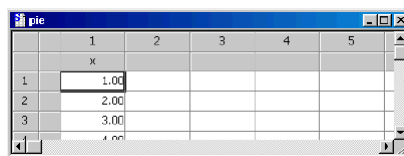


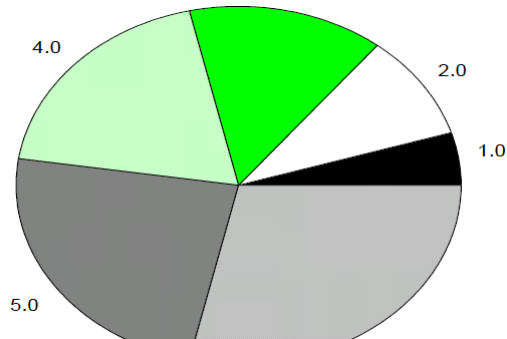

円グラフ

円グラフは、列の合計に対してその列の個々の値の構成比を示します。

円グラフを作成するには、

1. x 列を選択します。
2. **2D プロットパレット**の  ボタンをクリックします。




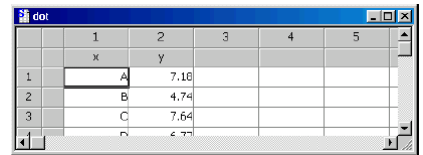
円グラフ 

点グラフ

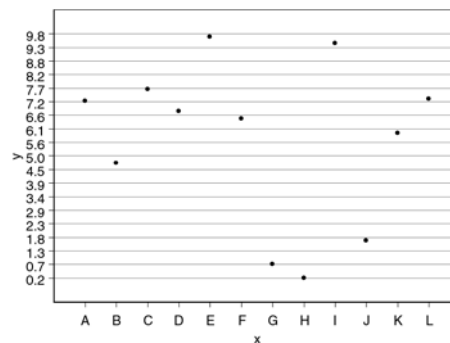

点グラフは、格子線を使ってカテゴリカルな目的変数を表し、その線の上に説明変数をプロットして、個々のカテゴリと変数の関係をグラフに表します。

点グラフを作成するには、

- 縦軸方向の整数列に対して横軸方向に値をプロットする1つのx列を選択するか、カテゴリカルデータxと対応するyを両方選択して、xに対してyをプロットします。
- 2Dプロットパレットのボタン  をクリックします。



	1	2	3	4	5
x	y				
1	A	7.10			
2	B	4.74			
3	C	7.64			
	D				



点グラフ 


棒グラフ


棒グラフは、対応するデータ値によって決まる高さ（横棒グラフの場合は幅）のバーを表したものです。統計量付き棒グラフは、各バーの最上部（または端部）にエラーバーを表します。

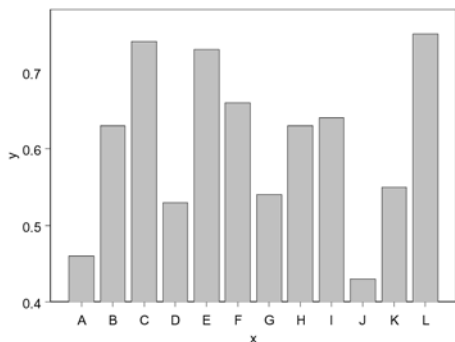
縦棒（Y の最小値を基準にした）グラフを作成するには、

	1	2	3	4	5
	x	y			
1	A	0.46			
2	B	0.63			
3	C	0.74			
	D	0.55			

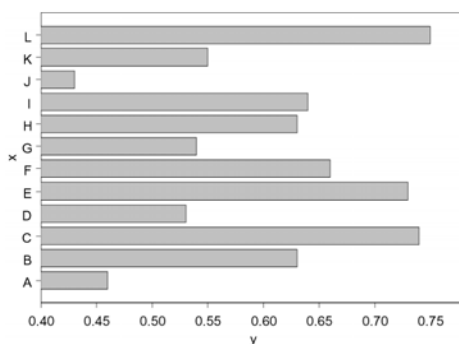
1. 棒グラフを作成する1つの列を選択し、x 軸ラベルを整数値にするか、x、y 両方を選択し、x 軸ラベルを付けるか、いずれかの選択をします。
2. 2D プロットパレットの  ボタンをクリックします（列のどれかの値が負の場合は、代わりに  ボタンをクリックします）。

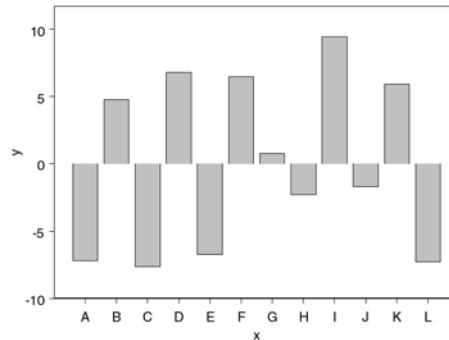
横棒グラフを作成するときは、列を逆の順序で選択し、 ボタンをクリックします。

Y の最小値を基準とした縦棒グラフ 




横棒グラフ 



基準が0の棒グラフ 

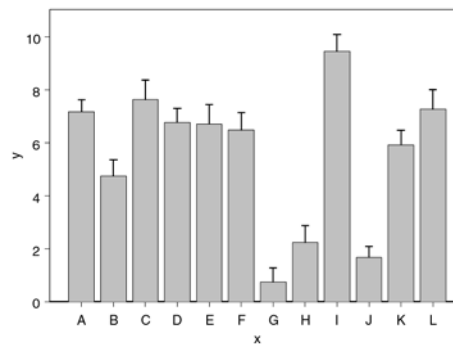

データセットが、誤差値を表す z 列を含む場合は、z データを使用して y の棒グラフを作成し、誤差範囲を表示させることができます。

統計量付き棒グラフを作成するには、

1. x、y および z 列を選択します。
2. エクストラプロットパレットの  ボタンをクリックします。

	1	2	3	4	5
	x	y	z		
1	A	7.10	0.46		
2	B	4.74	0.63		
3	C	7.64	0.74		


データが複数の y 列で配列されている場合は、S-PLUS が、自動的に統計量を計算して表示します。詳細は、オンラインヘルプを参照してください。

統計量付き棒グラフ 


パレート グラフ


パレートグラフは、独立変数に関してソートして並べられた棒グラフをそのカテゴリ（棒）の累積率を示す線グラフと組み合わせて表示したものです。各カテゴリの降順のヒストグラムが、累積率を示す線グラフと一緒にプロットされます。パレートグラフは、基本的に、棒グラフと線グラフの特性が組み合わさったものです。

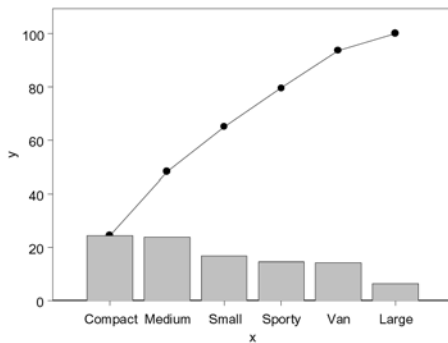
縦パレートグラフを作成するには、


1. カテゴリカルデータの x 列と数値の y 列を選択します。
2. 2D プロットパレットの  ボタンをクリックします。

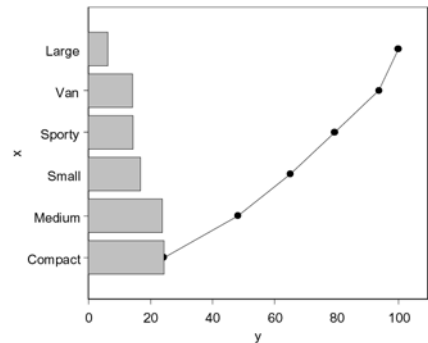
	1	2	3	4	5
	x	y			
1	Small	2550.00			
2	Small	2345.00			
3	Small	1845.00			
	Small	2250.00			

横パレートグラフを作成するには、列を逆の順序で選択し、エクストラプロットパレットの  ボタンをクリックします。

縦パレートグラフ 



横パレートグラフ 




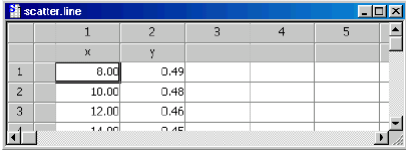
2次元データのプロット

散布図と 線グラフ


散布図と線グラフは、データを表示するための最も基本的なグラフです。これらのグラフを使用して、1つの列のデータをプロットしたり、あるデータ列を別のデータ列に対応してプロットしたりすることができます。

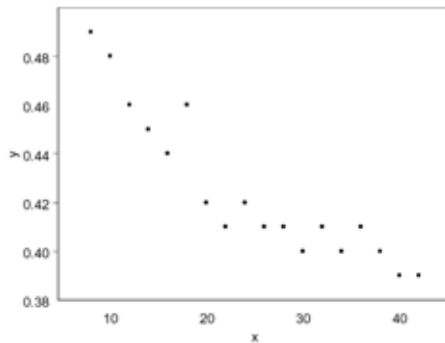
散布図／線グラフを作成するには、


1. x または y の列を選択して、縦軸方向の値を横軸方向の整数列に対してプロットするか、x と y を両方選択して、x に対して y をプロットします。
2. 希望のプロットに対応する **2D プロット** パレットボタンをクリックします（高密度棒グラフ（Y 基準）を作成するには、**エクストラプロット** パレットの  ボタンをクリックします）。

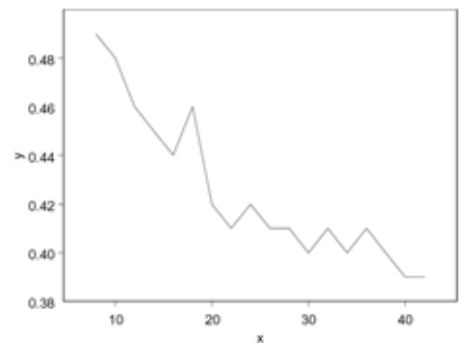


	1	2	3	4	5
	x	y			
1	0.00	0.49			
2	10.00	0.48			
3	12.00	0.46			
	14.00	0.45			


散布図 

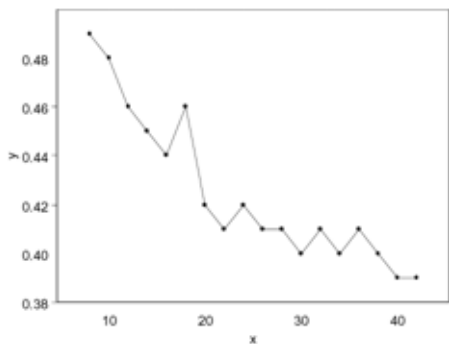



線グラフ 

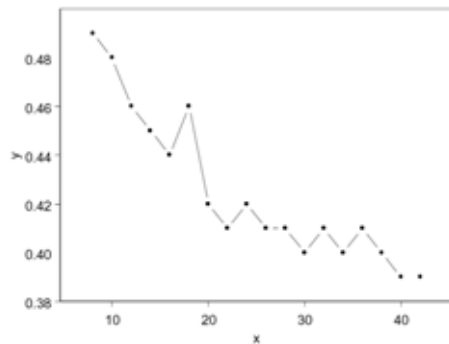



第4章 グラフの作成

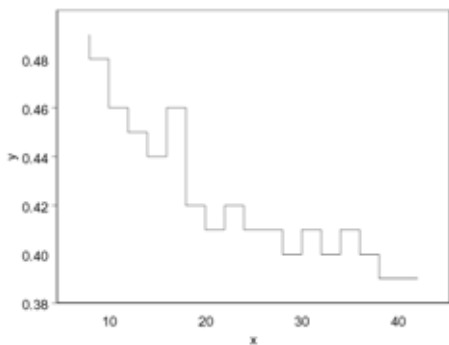
線付き散布図 



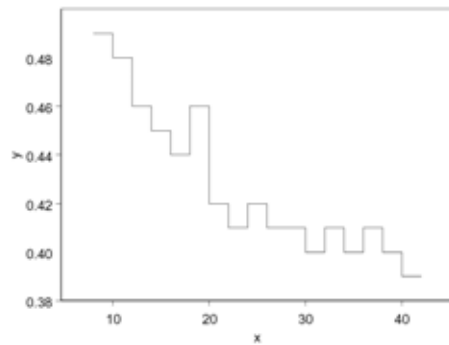
飛び線付き散布図 




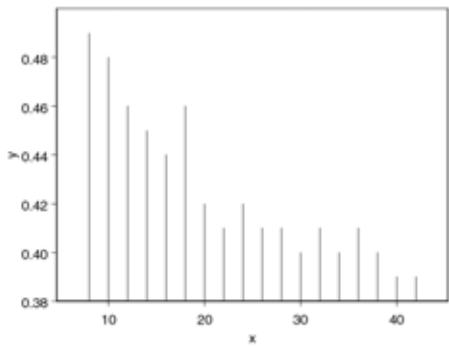
縦ステップグラフ 




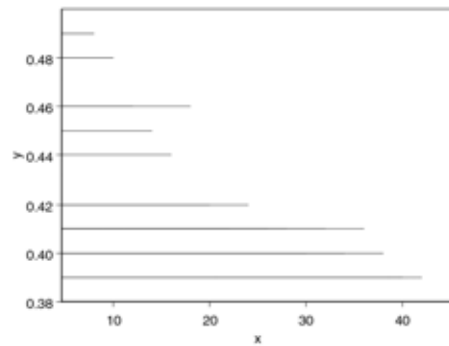
横ステップグラフ 



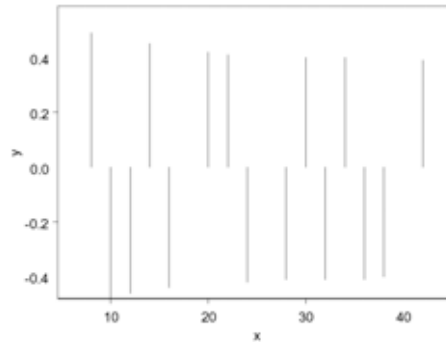
高密度縦棒グラフ 



高密度横棒グラフ 



高密度棒グラフ (Y ゼロ基準)



曲線あてはめ プロット

曲線あてはめプロットは、データ点の散布図とその回帰線を表します。回帰線は、適切な変換モデルを使用して与えられた x に対する y 値を計算するために、通常の最小二乗法を用いて生成されます。

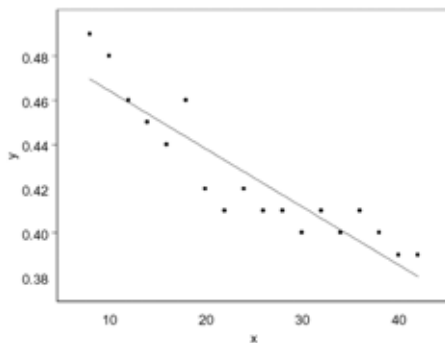
曲線あてはめプロットを作成するには、

1. x と y の列を選択します。
2. 希望のプロットに対応する **2D プロット** パレットのボタンをクリックします。

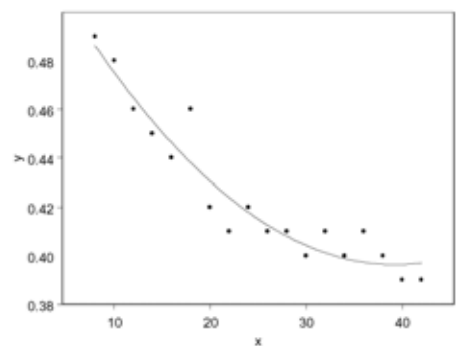
A screenshot of a software interface showing a data table. The table has columns labeled 'x' and 'y'. The data points are:

	x	y
1	0.00	0.49
2	10.00	0.48
3	12.00	0.46
4	14.00	0.46
5	16.00	0.44
6	18.00	0.42
7	20.00	0.41
8	22.00	0.41
9	24.00	0.41
10	26.00	0.41
11	28.00	0.40
12	30.00	0.40
13	32.00	0.40
14	34.00	0.40
15	36.00	0.39
16	38.00	0.39
17	40.00	0.39

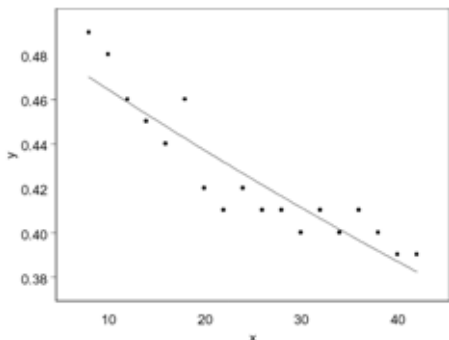
回帰直線プロット



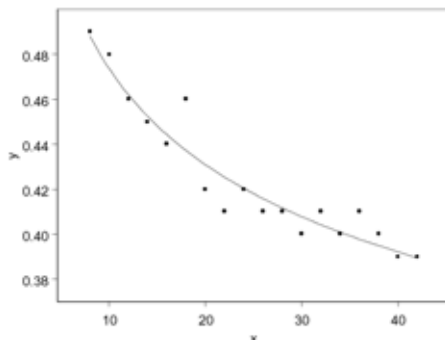
多項式回帰プロット




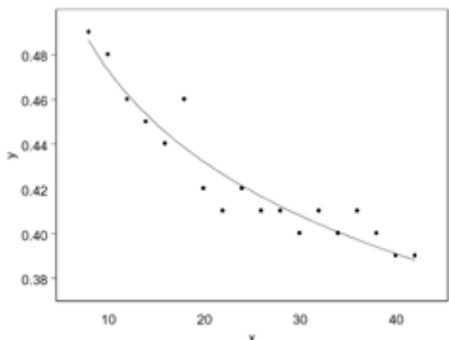
指数回帰プロット 



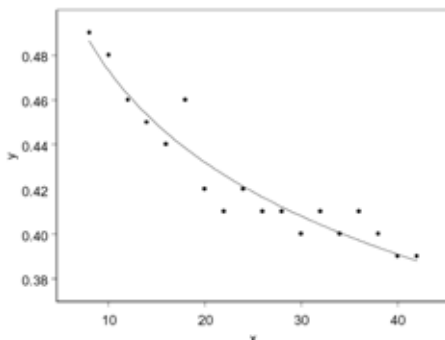
べき乗回帰プロット 



常用対数回帰プロット 



自然対数回帰プロット 



非線形曲線 あてはめ プロット

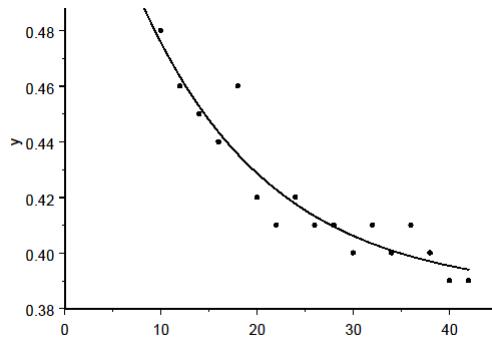
非線形曲線あてはめは、ユーザが定義したモデル式を一組のデータ点にあてはめます。モデル指定と、すべてのパラメータの初期値指定が必要なので、データを選択してプロットボタンをクリックするだけでは、グラフは自動的に生成されません。代わりに新しい**グラフシート**が開かれ、プロットアイコンが左上角に現れます。グラフを生成するためには、プロットアイコンをダブルクリックして **Nonlinear Curve Fitting** ダイアログを表示させ、適切なフィールドに必要な情報を指定してください。このグラフ作成の詳細は、オンラインヘルプを参照してください。

非線形回帰プロットを作成するには、

1. 独立変数として x 列を選択し、従属変数として y 列を選択します。
2. **2D プロット**パレットの **NLS** ボタンをクリックします。オンラインヘルプを参照してください。

	1	2	3	4	5
	x	y	x_{Nonlin}	y_{Nonlin}	
1	0.00	0.49	0.00	0.49	
2	10.00	0.48	0.34	0.49	
3	12.00	0.46	0.69	0.46	
	14.00	0.45	0.99	0.46	

非線形回帰プロット **NLS**



平滑化 プロット

散布図平滑化法は、任意の平滑化関数を散布図にあてはめるのに役立ちます。

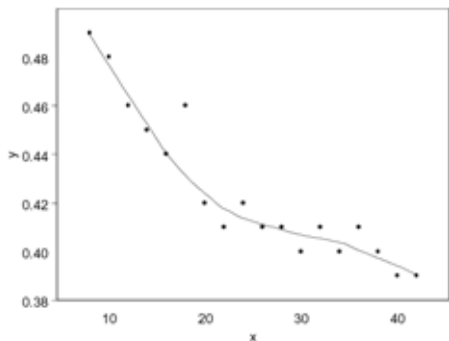
平滑化プロットを作成するには、

1. x と y の列を選択します。
2. 希望のプロットに対応する **2D プロット**パレットのボタンをクリックします。

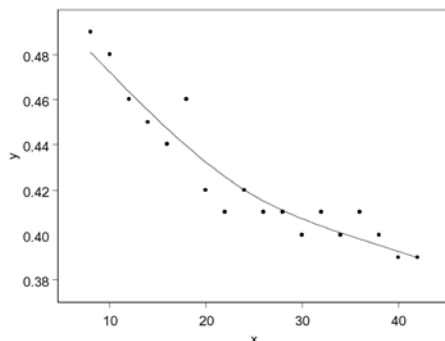
	1	2	3	4	5
	x	y			
1	0.00	0.49			
2	10.00	0.48			
3	12.00	0.46			
	14.00	0.45			


第4章 グラフの作成

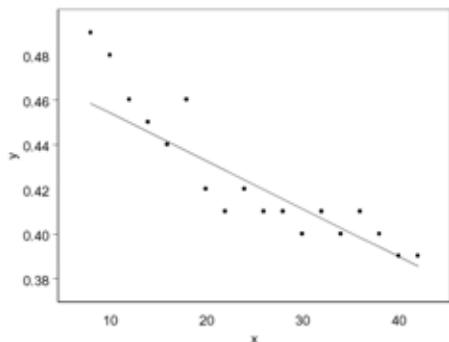
局所重み付き平滑化プロット 




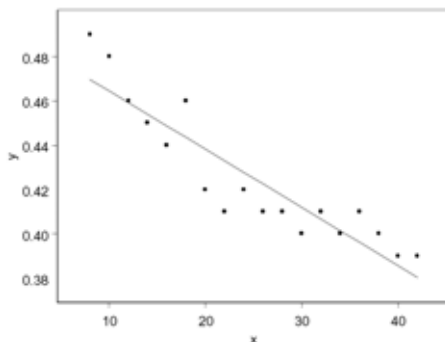
スプライン平滑化プロット 




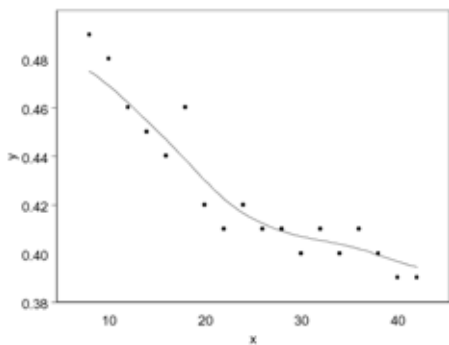
ロバスト LTS 平滑化プロット 




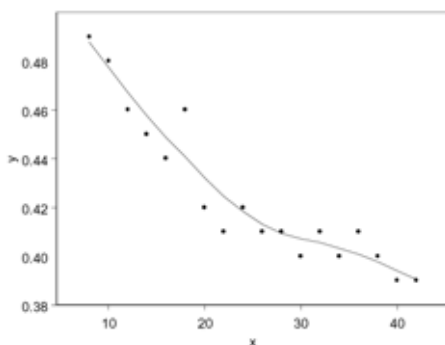
ロバスト MM 平滑化プロット 



核平滑化プロット 



フリードマン・スーパー平滑化プロット 



シンボル付き 線グラフ


シンボル付き線グラフは、単にテキスト文字列をプロット記号として使用した線グラフ／散布図の特殊な形です。

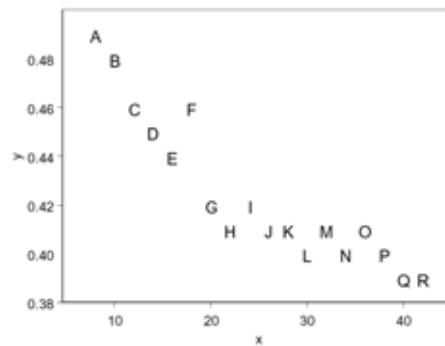
シンボル付き線グラフを作成するには、

1. x、y および z 列を選択します、z 列のテキストが、プロットシンボルとして使用されます。

	1	2	3	4	5
	x	y	z		
1	0.00	0.49	A		
2	10.00	0.48	B		
3	12.00	0.46	C		

2. 2D プロットパレットの  ボタンをクリックします。


シンボル付き線グラフ 



Y 系列 プロット

Y 系列プロットは、同じグラフに複数の系列をプロットする特殊な線グラフです。横軸方向の共通に自動生成された整数列に対して、データが縦軸方向にプロットされます。**X Axis Labels** ダイアログを使用して、整数列、時間や日付などのより適切なラベルと置き換えることができます。詳細は、オンラインヘルプを参照してください。

Y 系列プロットを作成するには、

1. y1 と y2 の列を選択します。
2. 2D プロットパレットの  ボタンをクリックします。

	1	2	3	4	5
	y1	y2			
1	0.00	1.14			
2	0.13	1.08			
3	0.14	1.50			


Y 系列プロット 




XY ペア線 グラフ

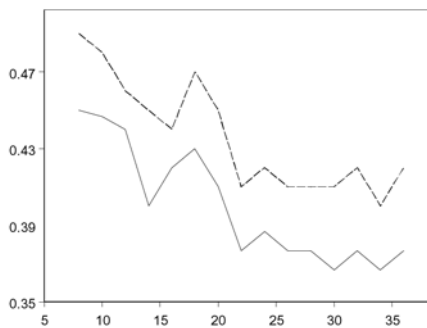
XY ペア線グラフを使用して、共通組の軸に複数組の x と y のペアをプロットすることができます。

XY ペア線グラフを作成するには、

1. $x1$, $y1$, $x2$ および $y2$ の列を選択します。
2. 2D プロットパレットの  ボタンをクリックします。

	1	2	3	4	5
	x1	y1	x2	y2	
1	8.00	0.49	6.00	0.45	
2	10.00	0.48	10.00	0.45	
3	12.00	0.46	12.00	0.44	


XY ペア線グラフ 



類別箱型図


類別箱型図は、データの複数の列を表す並んだ箱型図からなります。 x 列の行の数が箱の数を決定し、 y 列の行の数が、 x 列の行の数で割り切れる数でなければなりません。


類別縦箱型図を作成するには、

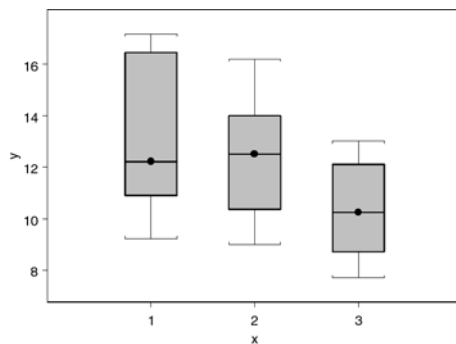
1. x と y の列を選択します。
2. エクストラプロットパレット  ボタンをクリックします。




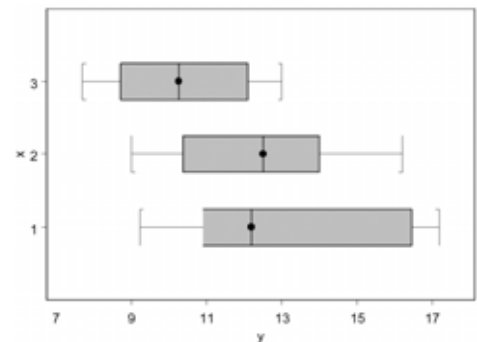
	1	2	3	4	5
x	1.00	10.91			
2	2.00	10.37			
3	3.00	13.00			

類別横箱型図を作成するには、代わりに  ボタンをクリックします。

類別縦箱型図 




類別横箱型図 



類別棒グラフ

類別棒グラフは、データを類別されたバーで表します。 x 値は、ラベルです。複数の y 列が、バーの高さを決定します。たとえば、各グループの最初のバーの高さは、 y の最初の列の値によって決定され、各グループの2番目のバーの高さは、 y の2番目の列の値によって決定されるという具合です。

類別縦棒グラフを作成するには、

1. x および複数の y 列を選択します。
2. 2D プロットパレットの  ボタンをクリックします。




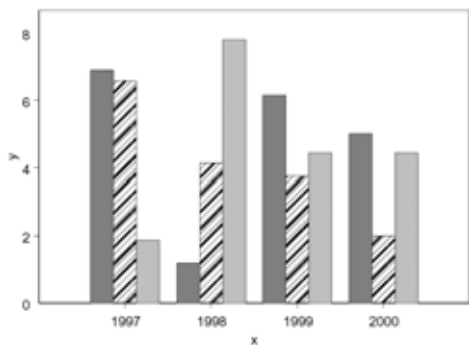
	1	2	3	4	5
x	1997	5.09	6.57	1.07	
2	1998	1.19	4.13	7.79	
3	1999	5.14	3.74	4.42	

類別横棒プロットを作成するには、

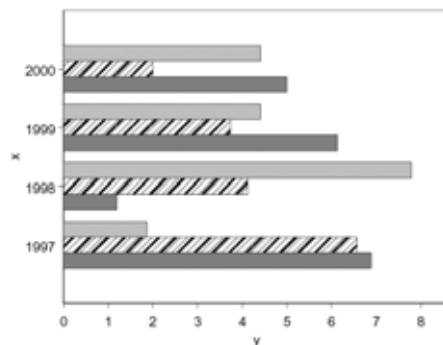
1. まず、複数の y 列を選択し、次に CTRL-クリックで x 列を最後に選択します。

2. 2Dプロットパレットの  ボタンをクリックします。

類別棒グラフ 




類別横棒グラフ 




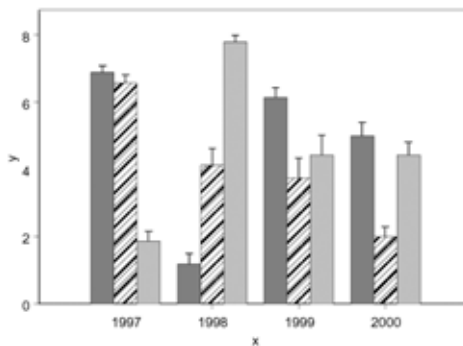
類別統計量付き棒グラフを作成するためには、1つの列にすべての y データが入っている必要があります。誤差として用いる値を含む同じ長さの z 列を作成してください。類別棒グラフは統計量を自動的に計算することができないことに注意してください。

類別統計量付き棒グラフを作成するには、

1. x 、 y および z 列を選択します。
2. エクストラプロットパレットの  ボタンをクリックします。

	1	2	3	4	5
	x	y	z		
1	1997	6.09	0.20		
2	1998	1.19	0.30		
3	1999	6.14	0.30		
	2000	5.00	0.20		


類別統計量付き棒グラフ 



積み上げ棒グラフ


積み上げ棒グラフは、積み上げたバーでデータを表します。 x の値がラベルになります。複数の y 列が、それぞれのバー部分の高さを決定します。たとえば、積み上げた一番下のバー部分の高さは、最初の y の値によって決定され、真ん中のバー部分の高さは、2番目の列の値によって決定されます。積み上げ棒グラフでは、エラーバーを表示することができないことに注意してください。


積み上げ縦棒グラフを作成するには、

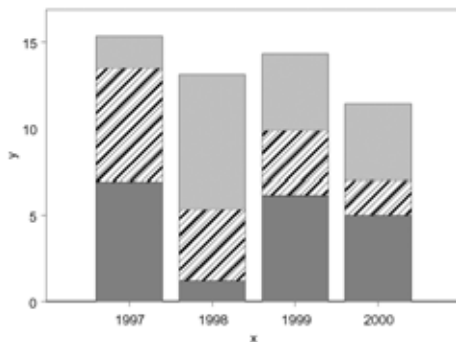
1. x および複数の y 列を選択します。
2. 2D プロットパレットの  ボタンをクリックします。


	1	2	3	4	5
x	y1	y2	y3		
1	1997	6.09	6.57	1.07	
2	1998	1.19	4.13	7.79	
3	1999	6.14	3.74	4.42	
4	2000	5.00	2.00	4.42	

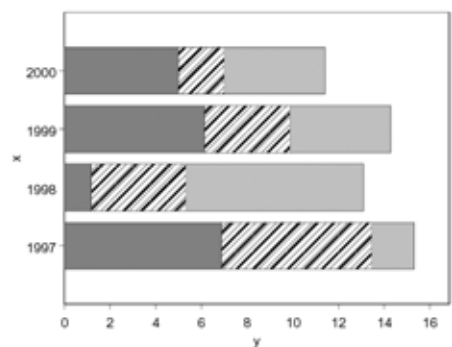
積み上げ横棒グラフを作成するには、

1. まず、複数の y 列を選択し、次に CTRL-クリックで x 列を最後に選択します。
2. 2D プロットパレットの  ボタンをクリックします。

積み上げ棒グラフ 




積み上げ横棒グラフ 




極座標 プロット


極座標プロットは、データを極座標で表示します。

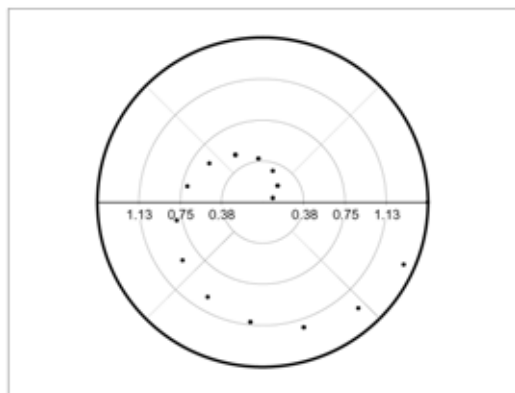
極座標散布図を作成するには、


1. x (半径値) と y (角度) の列を選択します。
2. エクストラプロットパレットの  ボタンをクリックします。

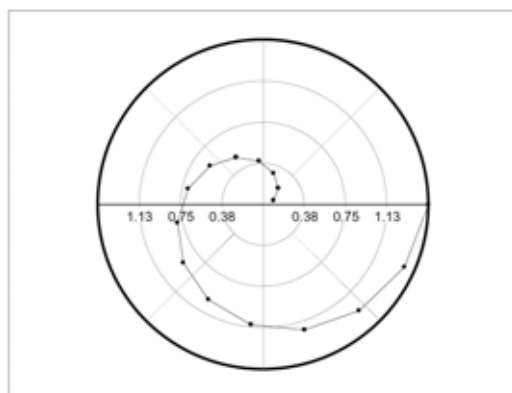
	1	2	3	4	5
	x	y			
1	0.10	0.42			
2	0.20	0.84			
3	0.30	1.26			

極座標線グラフを作成するときは、代わりに  ボタンをクリックします。

極座標散布図 



極座標線グラフ 



多次元データのプロット


3次元の散布図と線グラフ

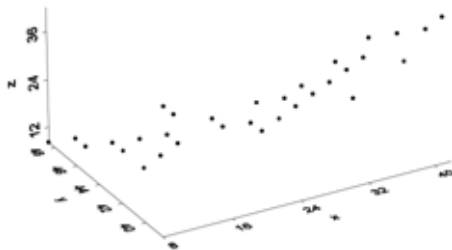
3次元散布図と3次元線グラフは、多次元データを3次元空間で表示します。3次元回帰プロットは、3次元散布図と3次元線グラフの一種であり、データ点により回帰面を描画します。


3次元散布図／線グラフを作成するには、

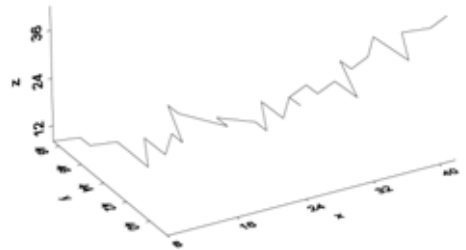
1. x、yおよびz列を選択します。
2. 希望のプロットに対応する **3D** プロットパレットのボタンをクリックします。


	1	2	3	4	5
	x	y	z		
1	6	49	8		
2	8	49	8		
3	10	48	10		

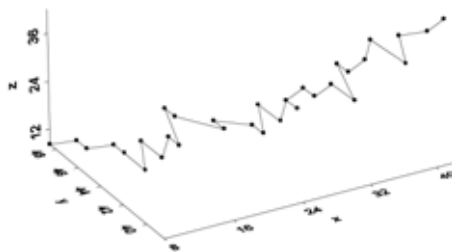
3次元散布図 




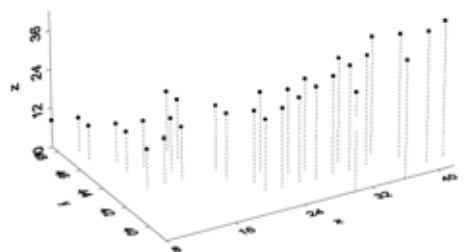
3次元線グラフ 



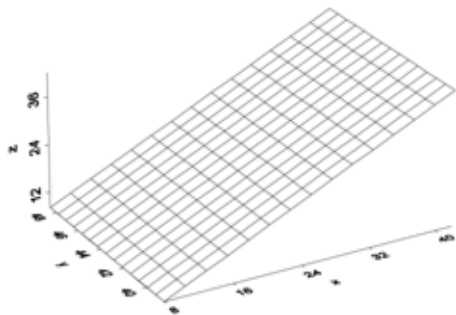
3次元線付き散布図 




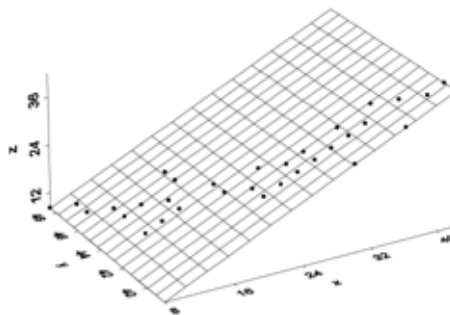
3次元射影付き散布図 



3次元回帰プロット 



3次元散布付き回帰プロット 





バブルプロットとカラープロット

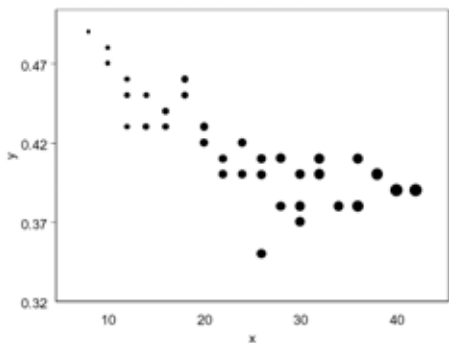
バブルプロットとカラープロットは、プロット記号の大きさや色を変化させることで3次元以上のデータを表すことができる散布図です。

様々な大きさや色の円として表されたzデータで、xに対するyの散布図を生成するには、

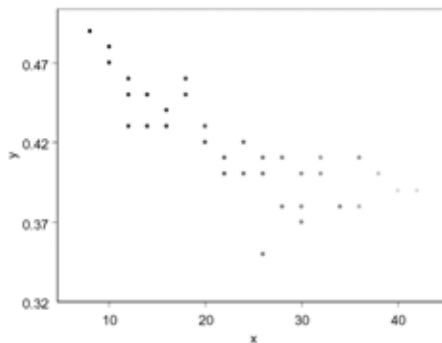
	1	2	3	4	5
	x	y	z		
1	6.00	0.49	6.00		
2	8.00	0.49	6.50		
3	10.00	0.48	10.00		
4	10.00	0.47	6.00		

1. x、yおよびz列を選択します。
2. バブルプロットの場合は、**2Dプロットパレット**の  ボタンをクリックします。カラープロットの場合は、ボタン  をクリックします。

バブルプロット 



カラープロット 



バブル カラー・ プロット


バブルカラー・プロットは、単にバブルプロットとカラープロットを組み合わせたものです。

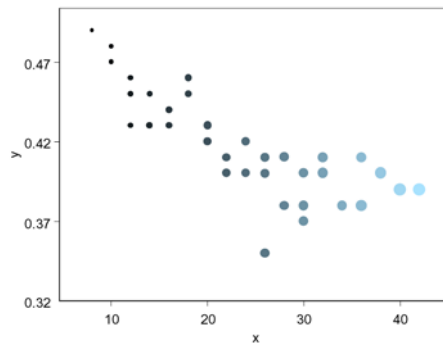
円の大きさを表す z データと、色として表す w データで、 x に対する y の散布図を生成するには、

	1	2	3	4	5
	x	y	z	w	
1	6.00	0.49	0.00	0.00	
2	8.00	0.49	8.50	8.00	
3	10.00	0.48	10.00	10.00	
4	10.00	0.47	0.00	10.00	

1. x 、 y 、 z および w 列を選択します。

2. 2D プロットパレットの  ボタンをクリックします。

バブルカラー・プロット 



High-Low プロット


High-Low プロットは通常、日ごと、月ごとまたは年ごとの単位時間内での高い値と低い値を、平均値、終値あるいは始値と共に表示します。したがって、意味のある High-Low プロットは、3~5 のデータ列からなるものです。 x データを含む選択した最初の列は、 x 軸にラベルを付けるために使用されます。最後の 2 つの列は、高いデータ値と低いデータ値を表します。平均データ、または始めと終わりのデータを、それぞれ y 、または y と z 列として選択してください。

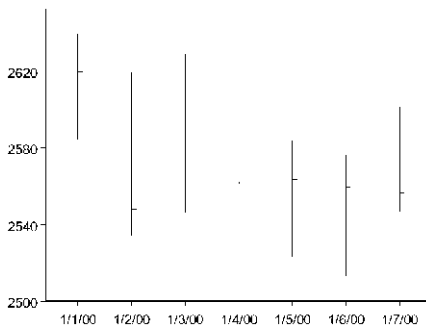
High-Low-Close または High-Low-Average プロットを作成するには、

	1	2	3	4	5
	x	close	high	low	
1	1/1/00	2619.37	2639.64	2504.01	
2	1/2/00	2547.75	2619.39	2534.23	
3	1/3/00	2615.77	2628.83	2546.17	
4	1/4/00	2551.74	2591.67	2516.00	


1. x 、close または average、high、および low の列を選択します。

- 2Dプロットパレットの  ボタンをクリックします。


High-Low プロット 

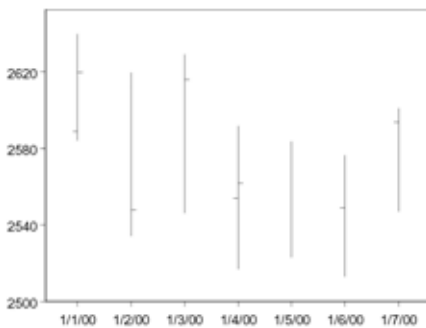


High-Low-Open-Close プロットを作成するには、

1. x、open、close、high、および low のデータを選択します（この順序で）。
2. 2Dプロットパレットの  ボタンをクリックします。

	1	2	3	4	5
x	1/1/00	open	close	high	low
1	1/1/00	2590.81	2619.37	2639.64	2584.01
2	1/2/00	2591.39	2547.73	2619.59	2534.23
3	1/3/00	2596.81	2615.77	2628.83	2546.17

High-Low-Open-Close プロット 



ローソク足 プロット


ローソク足プロットは、High-Low-Open-Close プロットを変形させたもので、始値と終値の間の差を塗りつぶした四角形として表示します。四角形の色は、差が正か負かを示します。

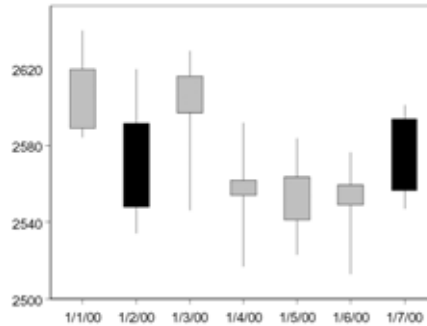
ローソク足プロットを作成するには、

1. x, open, close, high, および low のデータを選択します (この順序で)。

	1	2	3	4	5
	x	open	close	high	low
1	1/1/00	2500.01	2519.37	2639.64	2504.01
2	1/2/00	2591.39	2547.75	2619.56	2534.23
3	1/3/00	2595.81	2615.77	2628.83	2545.17

2. 2Dプロットパレットの  ボタンをクリックします。

ローソク足プロット 



エラーバー・プロット


エラーバー・プロットは、プロットしたデータ点のまわりの誤差の範囲を表します。x値は、x軸方向のバーの位置を決定します。データセットに1つのx列と複数のy列が含まれる場合、S-PLUSは、統計量を自動的に計算し表示します。詳細は、オンラインヘルプを参照してください。

エラーバー縦棒グラフを作成するには、

1. zのデータを使ってyのエラーバー・プロットを作成し、エラーバーを表示するために、x、y、およびz列を選択します。

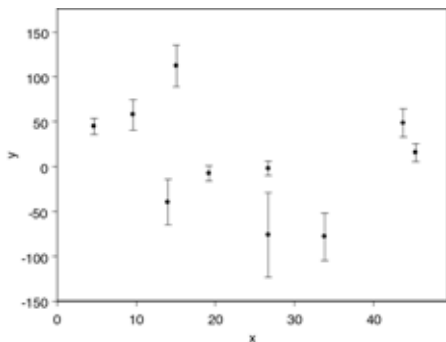
	1	2	3	4	5
	x	y	z		
1	4.67	44.73	-9.00		
2	9.51	57.65	15.96		
3	14.04	-99.66	-25.28		

2. 2Dプロットパレットの  ボタンをクリックします。

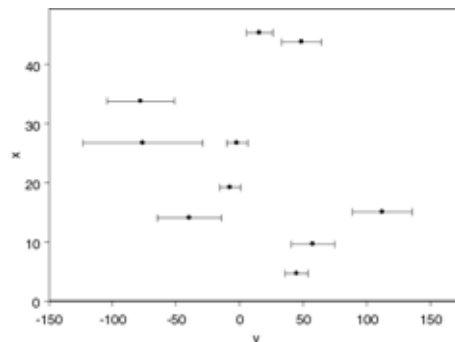
エラーバー横棒グラフを作成するには、xとyの列を逆の順序で選択し、 ボタンをクリックします。

第4章 グラフの作成


エラー棒グラフ

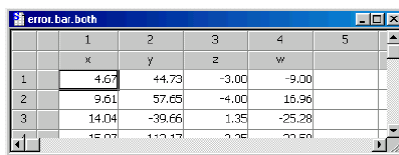


エラー棒横棒グラフ



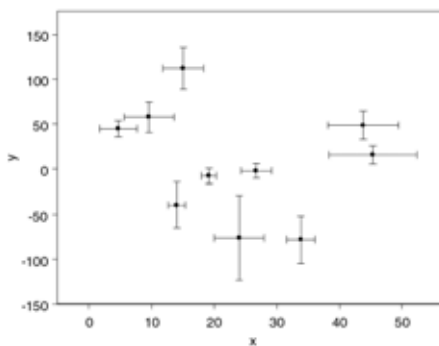
両方向のエラー棒縦横グラフを作成するには、

1. z データを使ってエラー棒横棒グラフを表示し、w データを使ってエラー棒縦棒グラフを表示することによって、エラー棒・プロットを作成するために、x、y、z および w 列を選択します。
2. 2D プロットパレットの  ボタンをクリックします。



	1	2	3	4	5
	x	y	z	w	
1	4.67	44.73	-3.00	-9.00	
2	9.61	57.65	-4.00	16.36	
3	14.04	-39.66	1.35	-25.28	
	15.07	110.17	0.26	20.50	


エラー棒両方向グラフ




ベクトル図


ベクトル図は、 x - y 平面におけるある位置での流れの向きと速度を表します。また、ベクトル図を使って、データセット内のデータで矢印のグループを描くことができます。

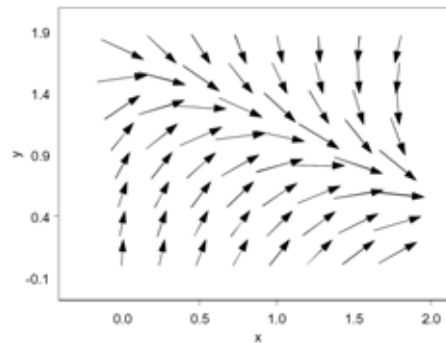
ベクトル図を作成するには、

1. x 、 y 、 z (角度) と w (大きさ) の列を選択します。
2. 2D プロットパレットの  ボタンをクリックします。



	1	2	3	4	5
	x	y	z	w	
1	0.00	0.10	87.21	0.21	
2	0.20	0.10	80.07	0.22	
3	0.51	0.10	72.88	0.22	


ベクトル図 

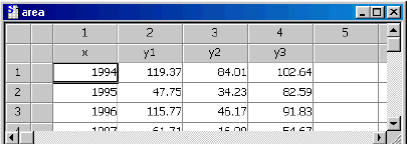


エリア チャート


エリアチャートは、一組のデータの各系列が時間全体にどのように影響するかを示すのに便利です。

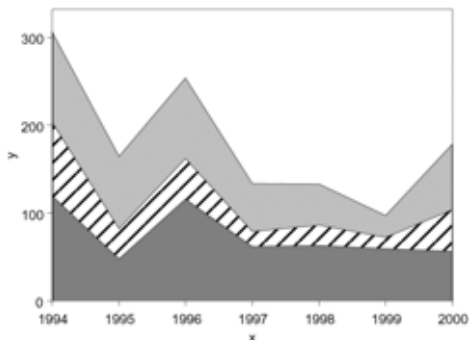
エリアチャートを作成するには、

1. x 、 y の折れ線を描き、その折れ線の下エリアを塗りつぶすときは、 x 列と 1 つの y 列を選択します。各組の値の折れ線を描き、各折れ線の下エリアを塗りつぶすときは、 x と複数の y 列を選択してください。
2. 2D プロットパレットの  ボタンをクリックします。



	1	2	3	4	5
	x	y1	y2	y3	
1	1994	119.37	84.01	102.64	
2	1995	47.75	34.23	82.59	
3	1996	115.77	45.17	91.83	


エリアチャート 

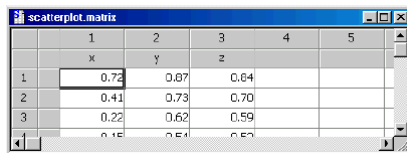



対散布図 (散布図 行列)

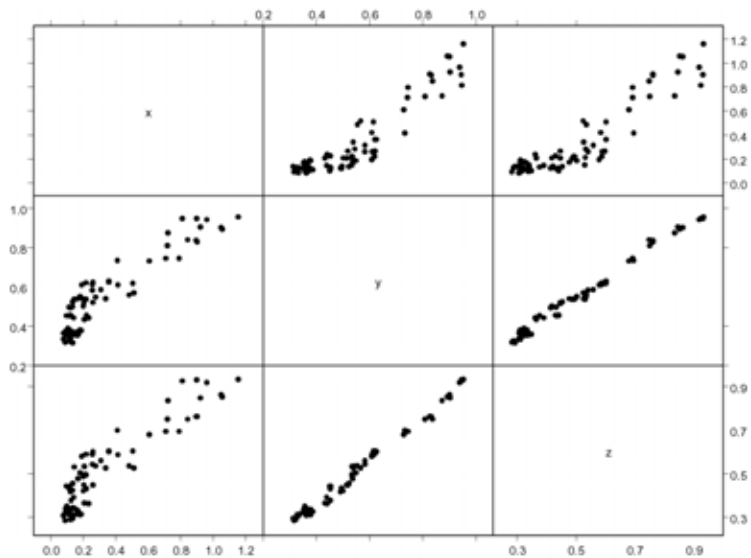
対散布図は、多変量データの対になった変数の関係を示す対になった散布図の行列です。

対散布図を作成するには、

1. x, y および z 列を選択します。
2. 2D プロットパレットの  ボタンをクリックします。



対散布図 



等高線図とイメージ

2次元の等高線図とイメージは、3次元データを2次元平面に表わすものです。それぞれの等高線は、対応する3次元面によるレベル、または高さを表します。カラー等高線図は、等高線間に色を塗ってレベルを区別します。3次元等高線図は、等高線が3次元空間に描かれる点以外は2次元等高線図と同じです。

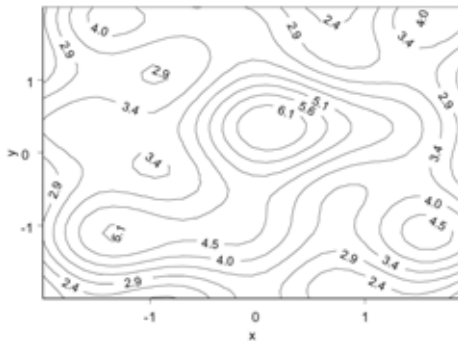
グリッドデータからでも不規則なデータからでも、2次元と3次元の等高線図を作成することができます。詳細は、オンラインヘルプを参照してください。


等高線図／イメージを作成するには、

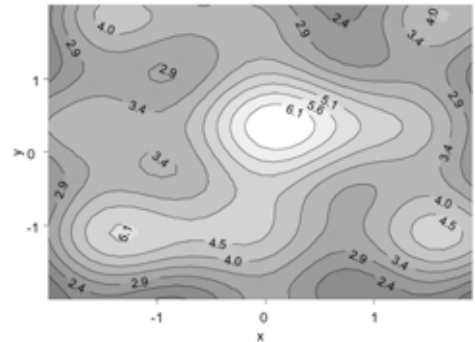
1. xおよび複数のy列を選択します。
2. 希望のプロットに対応する **2D** プロットパレットまたは **3D** プロットパレットをクリックします。


contour levels					
	1	2	3	4	5
	x	y	z		
1	-2.00	-2.00	1.20		
2	-1.87	-2.00	1.53		
3	-1.74	-2.00	1.94		
4	1.61	2.00	2.15		

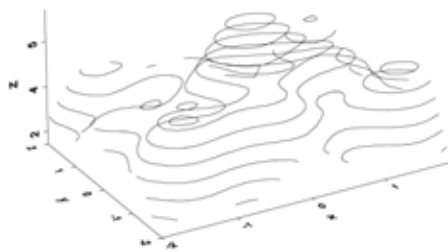
等高線図 




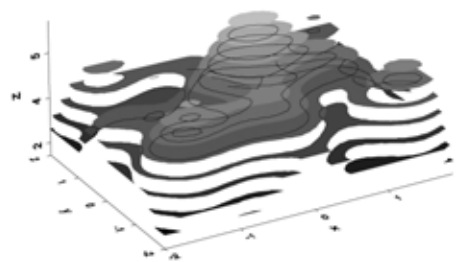
カラー等高線図 




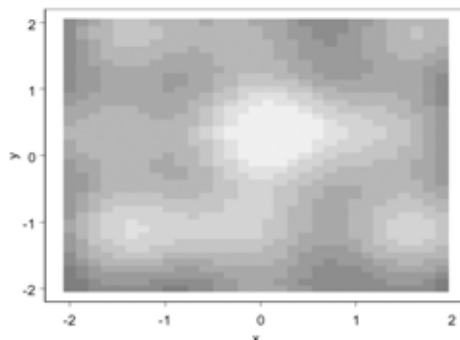
3次元等高線図 



3次元カラー等高線図 



イメージ 



鳥瞰図／3次元棒グラフ

鳥瞰図は、メッシュまたはグリッド上のデータを3次元空間に描きます。スプラインプロットは、グリッドデータを平滑化した面で、3次元棒グラフは、グリッド上の値をバーで表したものです。2変量データに対し3次元棒グラフはデータの同時分布を示す2次元ヒストグラムを作成します。カラーイメージオプションによって、鳥瞰図の帯またはグリッドを塗りつぶす色を指定することができます。


グリッドデータと不規則データのどちらからも鳥瞰図と3次元棒グラフを作成することができます。詳細は、オンラインヘルプを参照してください。

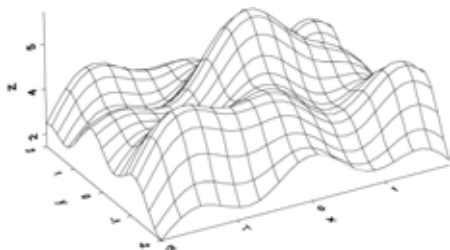
鳥瞰図、3次元棒グラフを作成するには、

1. x、y および z 列を選択します。
2. 希望のプロットに対応する

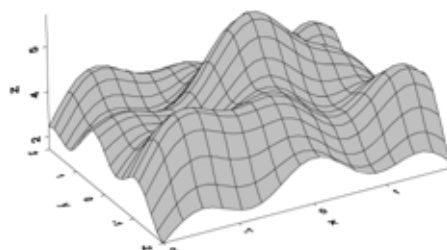
three.d.surface					
	1	2	3	4	5
	x	y	z		
1	-2.00	-2.00	1.20		
2	-1.87	-2.00	1.53		
3	-1.74	-2.00	1.94		


3Dプロットパレットのボタンをクリックします。

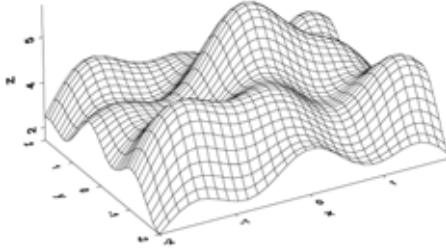
鳥瞰図 (粗い) 



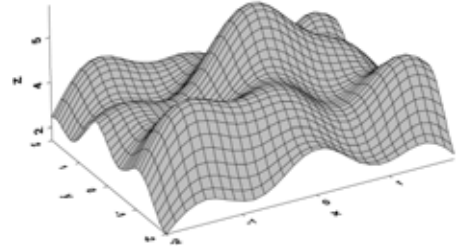
塗りつぶし鳥瞰図 (粗い) 




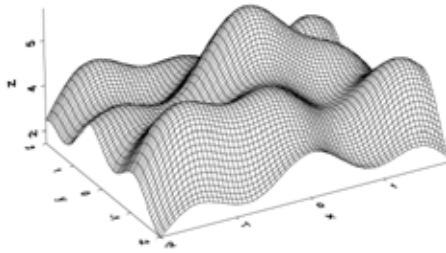
鳥瞰図 




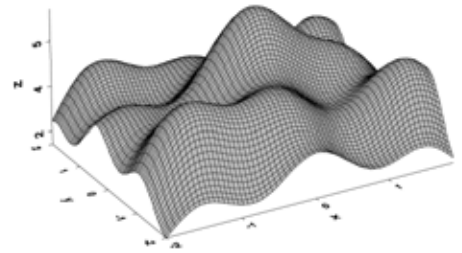
塗りつぶし鳥瞰図 



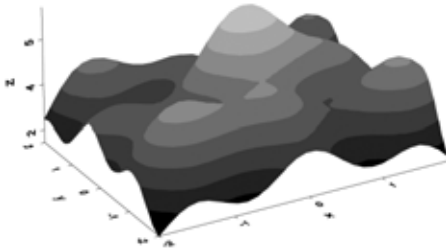
鳥瞰図 (スプライン) 




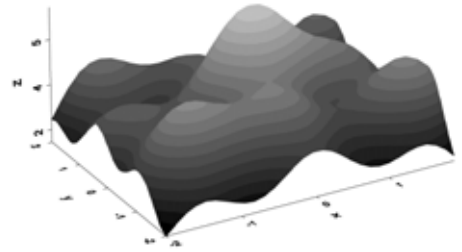
塗りつぶし鳥瞰図 (スプライン) 



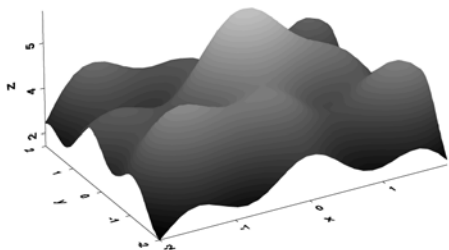
8色鳥瞰図 



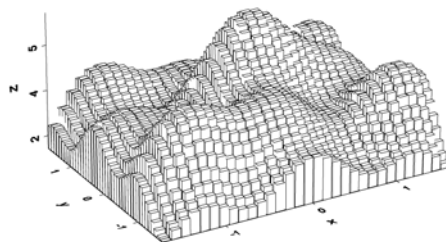
16色鳥瞰図 



32色鳥瞰図 



3次元棒グラフ 




コメント図


コメント図は、グラフ上に文字データをプロットします。すべての軸タイプを使用することができます。2次元コメント図では、 x 値と y 値が、各コメントの x 、 y 位置を指定し、 z 値は、コメントテキストです。 z 値が指定されない場合は、グラフに x 、 y 座標が表示されます。

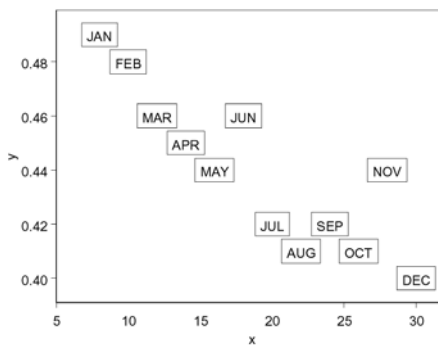
コメント図を使用して、任意の文字または文字の組合せを記号としてプロットしたり、ラベル付き散布図を作成したり、文字データを自動的にプロットしたり表を作ることができます。

コメント図を作成するには、

1. x 、 y および z 列を選択します。
2. エクストラプロットパレットの  ボタンをクリックします。

	1	2	3	4	5
	x	y	z		
1	0.00	0.49	JAN		
2	10.00	0.48	FEB		
3	12.00	0.46	MAR		
4	11.00	0.45	APR		
5	18.00	0.46	MAY		
6	20.00	0.42	JUN		
7	23.00	0.41	JUL		
8	24.00	0.41	AUG		
9	25.00	0.42	SEP		
10	26.00	0.41	OCT		
11	28.00	0.44	NOV		
12	30.00	0.40	DEC		

コメント図 




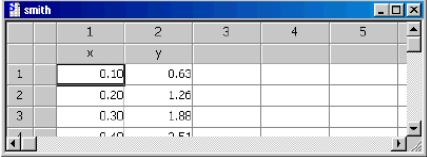
スミス チャート

極座標で描かれたスミスチャートは、マイクロ波工学でインピーダンス特性を示すためにしばしば使用されます。スミスチャートには、反射、インピーダンス、および円の3つの種類があります。パレットボタンをクリックして、反射プロットだけを自動的に作成することができます。


スミス反射チャートでは、 x 値は、振幅であり、0 から 1 の範囲でなければなりません。 y 値は、水平線から時計回りに測定した角度です。

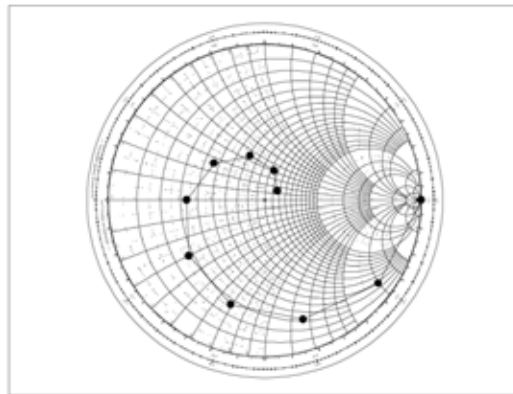
スミス反射チャートを作成するには、

1. x と y の列を選択します。
2. エクストラプロットパレットの  ボタンをクリックします。



	1	2	3	4	5
	x	y			
1	0.10	0.65			
2	0.20	1.26			
3	0.30	1.88			

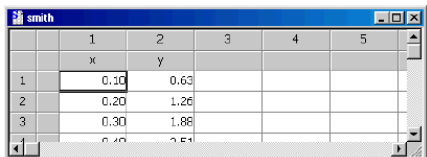
スミス反射チャート 




スミス・インピーダンスチャートでは、 x 値は、抵抗データであり、 y 値は、リアクタンスデータです。

スミス・インピーダンスチャートを作成するには、

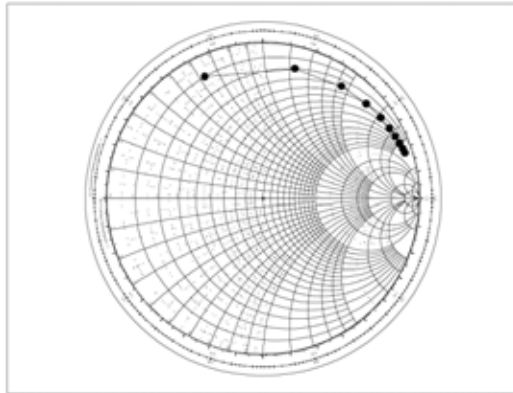
1. x と y の列を選択します。



	1	2	3	4	5
	x	y			
1	0.10	0.65			
2	0.20	1.26			
3	0.30	1.88			


2. エクストラプロットパレットの  ボタンをクリックします。
3. プロット要素（点など）を右クリックし、ショートカットメニューから **Options** を選択します。
4. **Data Options** グループで、**Data Type** フィールドの **Impedance** を選択します。
5. **OK** をクリックします。

スミス・インピーダンスチャート



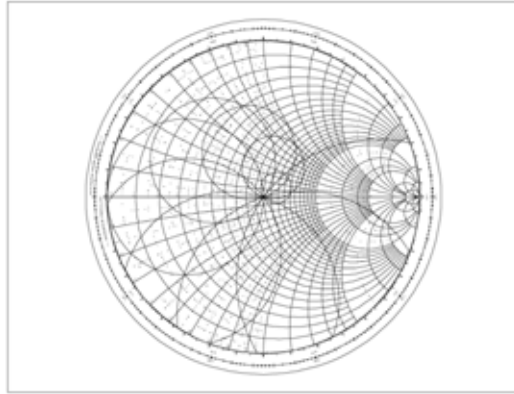
スミス円チャートでは、 x 値は、正の値でなければならず、スミスチャートの中心から描きたい円の中心までの距離を指定します。 y 値は、水平線から時計回りに測定した角度です。 z 値は、半径であり、正の値でなければなりません。

スミス円チャートを作成するには、

1. x 、 y および z 列を選択します。
2. エクストラプロットパレットの  ボタンをクリックします。
3. プロット要素（点など）を右クリックし、ショートカットメニューから **Options** を選択します。
4. **Data Options** グループで、**Data Type** フィールドの **Circle** を選択します。
5. **OK** をクリックします。

	1	2	3	4	5
	x	y	z		
1	0.10	0.60	0.10		
2	0.20	1.20	0.20		
3	0.30	1.80	0.30		

スミス円チャート

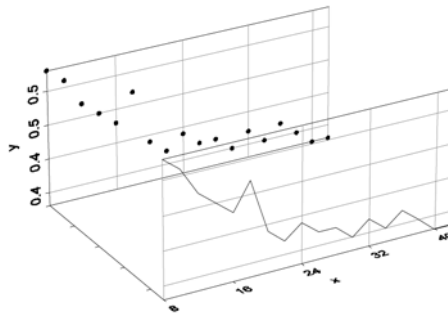


投射プロット

2次元グラフタイプのほとんどは、3次元平面に投影することができます。投射プロットを使って、複数の2次元グラフを3次元空間で結合し、次にその結果を回転させることができます。

投射プロットを作成するときは、メニューまたはドラッグ アンド ドロップを使用することができます。投影プロットを作成する詳細は、オンラインヘルプを参照してください。

投射プロット





Trellis グラフ

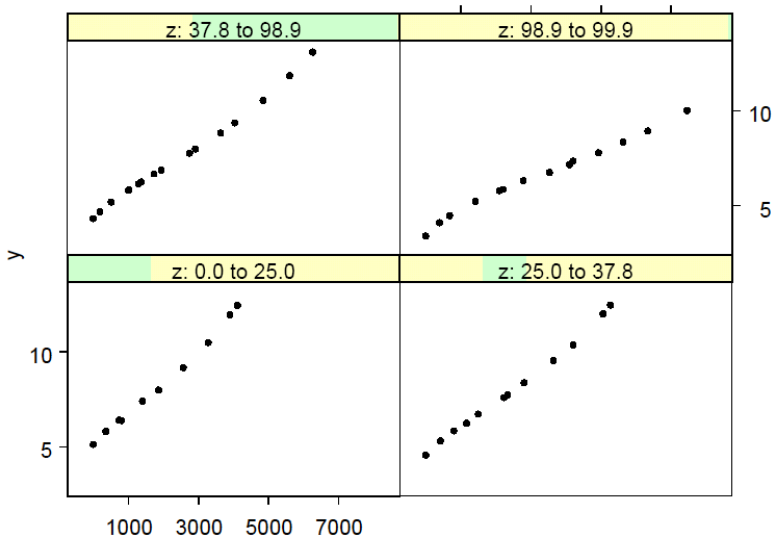
Trellis グラフを使用すると、“条件付け”によりデータ内の様々な変数の関係を表示することができます。条件付け変数の間隔にあわせて分割されたデータの各サブセットからなる、一連のパネルが表示されます。

z に関して条件が付けられた x に対する y の散布図を作成するには、

	1	2	3	4	5
x		y	z		
1	1.00	5.11	0.00		
2	740.80	6.39	0.00		
3	1407.47	7.39	0.00		

1. **2D プロットパレット**を開き、次に、**標準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 プロットパレットの各プロットボタンの上部に黄色のバーが現われます。
2. x 、 y および z 列を選択します。
3. **2D プロットパレット**の  ボタンをクリックします。

Trellis グラフ



Trellis グラフの他の例は、102 ページの「Trellis グラフ」を参照してください。

第 5 章

データのインポートと エクスポート

はじめに	164
インポートとエクスポートが可能なファイルタイプ	165
データファイルからのインポートとデータファイルへのエクスポート	169
データファイルからインポートする	169
データファイルへエクスポートする	176
ODBC テーブルからのインポートと ODBC テーブルへのエクスポート	180
ODBC Data Source Administrator	180
ODBC ドライバ	181
データソースを定義する	181
ODBC テーブルからインポートする	182
ODBC テーブルへエクスポートする	186
フィルタ式	189
変量式	189
特定タイプのファイルのインポートとエクスポートに関する注意	191
ASCII (可変長 ASCII) ファイル	191
dBASE ファイル	191
複数のテーブルを含むファイル	191
FASCII (固定長 ASCII) ファイル	191
Informix ファイル	194
Lotus ファイル	194
Microsoft Access ファイル	194
Microsoft Excel ファイル	194
Oracle ファイル	194
SYBASE ファイル	195

はじめに

作図や分析を行うために S-PLUS にデータを取り込む簡単な方法の 1 つは、別のソースからデータをインポートすることです。

S-PLUS は、様々なデータフォーマットの読み書きができます。また S-PLUS は、Open DataBase Connectivity (ODBC) 規格をサポートしており、S-PLUS と ODBC 対応データベース間のインポートとエクスポートが可能です。

さらに、主要な金融データベースである Bloomberg、FAME、および MIM からデータをインポートすることができます。これを行うときは、メインメニューから **ファイル ▶ ライブラリのロード** でライブラリ **financedb** を選択するか、**コマンドウィンドウ** から次のコマンドで **financedb** をロードしてください。

```
>library (financedb)
```

このライブラリには、ダイアログをカスタマイズする機能を含む、データのインポートとエクスポートに関連する機能が含まれています。ヘルプは、すべての金融関連の関数について、**コマンドウィンドウ** から利用でき、またメインメニューから **ヘルプ ▶ Available Help ▶ financedb** を選択することも可能です。

この章では、これらのインポートおよびエクスポート・オプションについて順を追って解説します。

インポートとエクスポートが可能な ファイルタイプ

表 5.1 は、インポートとエクスポートが可能なすべてのファイルフォーマットのリストです。S-PLUS は、リストにあるタイプに対してインポートとエクスポートの両方を行うことができますが、2 つだけ例外があり、SigmaPlot (.jnb) ファイルはインポートのみ、HTML (.htm*) のテーブルはエクスポートのみが可能です。

表 5.1 : データのインポートとエクスポートが可能なファイルタイプ

フォーマット	タイプ	デフォルト拡張子	備考
ASCII ファイル	"ASCII"	.csv .asc、.csv、.txt、.prn .asc、.dat、.txt、.prn	カンマで区切られる。 区切り文字あり。 空白区切り、スペース区切り、タブ区切り、ユーザ指定の区切り記号による。
dBASE ファイル	"DBASE"	.dbf	II、II+、III、IV ファイル。
DIRECT-DB2	"DIRECT-DB2"		DB2 データベース接続。ファイル引数は指定できません。
DIRECT-ORACLE	"DIRECT-ORACLE"	.ora	Oracle データベース接続。ファイル引数は指定できません。
DIRECT-SQL	"DIRECT-SQL"		Microsoft SQL Server データベース接続。ファイル引数は指定できません。このオプションは S-PLUS for Windows でのみ使用可能です。
DIRECT-SYBASE	"DIRECT-SYBASE"		Sybase データベース接続。ファイル引数は指定できません。
Epi Info ファイル	"EPI"	.rec	

第5章 データのインポートとエクスポート

表 5.1 : データのインポートとエクスポートが可能なファイルタイプ (続き)

フォーマット	タイプ	デフォルト拡張子	備考
固定フォーマット ASCII ファイル	"FASCII"	.fix 、 .fsc	
FoxPro ファイル	"FOXPRO"	.dbf	
Gauss データ ファイル	"GAUSS"、 "GAUSS96"	.dat	関連する DHT ファイルがあれば、自動的に GAUSS 89 として読み込みます。 DHT ファイルが見つからない場合は、 .DAT ファイルを GAUSS96 として読み込みます。GAUSS96 は Unix でのみ使用可能です。
HTML テーブル	"HTML"	.htm*	エクスポートのみ。
Lotus 1-2-3 ワークシート	"LOTUS"	.wk* 、 .wr*	
Matlab Matrix	"MATLAB"	.mat	ファイルには行列が 1 つ含まれていなければなりません。 バージョン 4 以下 — インポート/エクスポート。 バージョン 5 — インポートのみ。
Minitab ワークブック	"MINITAB"	.mtw	バージョン 8 からバージョン 12 まで。
Microsoft Access ファイル	"ACCESS"	.mdb	Microsoft Access ファイル。このファイルタイプは S-PLUS for Windows でのみ使用可能です。
Microsoft Excel ワークシート	"EXCEL"	.xl?	バージョン 2.1 から 2003 まで。
ODBC	"ODBC"	適用されない。	Informix (.ifx)、Oracle (.ora)、Sybase (.syb) データベース用。このファイルタイプは、S-PLUS for Windows でのみ使用可能です。

表 5.1 : データのインポートとエクスポートが可能なファイルタイプ (続き)

フォーマット	タイプ	デフォルト拡張子	備考
Oracle	"Oracle"	.ora	"DIRECT-ORACLE"と同じ。Oracle データベース接続。ファイル引数は指定できません。
Paradox データ ファイル	"PARADOX"	.db	
QuattroPro ワークシート	"QUATTRO"	.wq?、.wb?	
S-PLUS ファイル	"SPLUS"	.sdd	Windows、DEC UNIX。ファイルのインポートには、 <code>data.restore()</code> を使用する。
SAS ファイル	"SAS"、"SASV6"	.sd2	SAS バージョン 6 ファイル、Windows。
	"SAS1"、 "SAS6UX32"	.ssd01	SAS バージョン 6 ファイル、HP、IBM、Sun、UNIX。
	"SAS4"、 "SAS6UX64"	.ssd04	SAS バージョン 6 ファイル、Digital UNIX。
	"SAS7"	.sas7bdat、.sd7	SAS バージョン 7 または 8 ファイル、 現行のプラットフォーム。
	"SAS7WIN"	.sas7bdat、.sd7	SAS バージョン 7 以降のデータファイル、 Windows。
	"SAS7UX32"	.sas7bdat、.sd7	SAS バージョン 7 以降のデータファイル、 Solaris (SPARC)、HP-UX、IBM AIX。
	"SAS7UX64"	.sas7bdat、.sd7	SAS バージョン 7 以上のデータファイル、 Digital/Compaq UNIX。

第5章 データのインポートとエクスポート

表 5.1 : データのインポートとエクスポートが可能なファイルタイプ (続き)

フォーマット	タイプ	デフォルト拡張子	備考
SAS Transport ファイル	"SAS_TPT"	.xpt 、 .tpt	バージョン 6.x。SAS プログラムでは、特別なエクスポート・オプションを指定しなければならない場合があります。これらのファイルの読み書きには、SAS Xport エンジン (PROC CPORT ではなく) を使用してください。
SigmaPlot ファイル	"SIGMAPLOT"	.jnb	インポートのみ。
SPSS データ ファイル	"SPSS"	.sav	OS/2 ; Windows ; HP、IBM、Sun、DEC UNIX。
SPSS Portable ファイル	"SPSSP"	.por	
Stata データ ファイル	"STATA"	.dta	バージョン 2.0 以上。
Sybase	"SYBASE"		"DIRECT-SYBASE"と同じ。Sybase データベース接続。ファイル引数は指定できません。
SYSTAT ファイル	"SYSTAT"	.syd 、 .sys	倍精度または単精度の.SYS ファイル。

注意

Informix、Oracle、SYBASE の各データベースからのインポート、およびこれらのデータベースへのエクスポートは、すべて ODBC を使用して行われます。したがって、様々な ODBC コンポーネントが適切にインストールされていなければなりません。180 ページの「ODBC テーブルからのインポートと ODBC テーブルへのエクスポート」を参照してください。

データファイルからのインポートと データファイルへのエクスポート

S-PLUS では、データのインポートとエクスポートのために便利なダイアログを用意しています。インポートするときは、ほとんどの種類のファイルでは、一般にファイルの名前と種類を指定するだけです。デフォルトでは、S-PLUS がデータを新しいデータセットにインポートし、そのデータをデータ・ウィンドウに表示します。

注意

インポートしたデータを自動的にデータ・ウィンドウに表示したくない場合は、メインメニューから、オプション ▶ 設定を選択し、Data タブをクリックし、**Show Imported Data in View** チェックボックスのチェックを外します。

エクスポートするときは、ほとんどの種類のファイルで、一般にエクスポートしたいデータセットと新しいファイルの名前と種類を指定するだけです。デフォルト設定により S-PLUS がデータを新しいデータファイルにエクスポートします。

データ ファイルから インポート する

下の図 5.1 に示す **Import From File** ダイアログは、それぞれ **Data Specs**、**Options**、**Rows**、および **Columns** というタブの付いた 4 つのページから構成されています。このダイアログを開くには、以下のいずれかを実行します。

- メインメニューから **ファイル ▶ インポート ▶ ファイルから読み込み** を選択します。
- メインメニューから **データ ▶ データの選択** を選択します。 **Select Data** ダイアログの **Source** グループで、**Import File** ラジオボタンをクリックし、**OK** をクリックします。

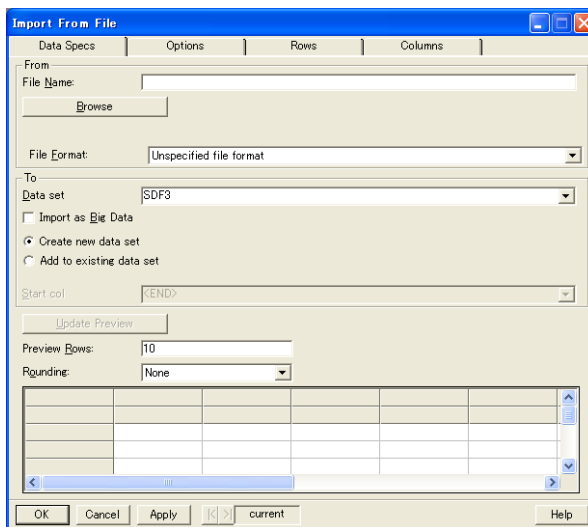


図 5.1 : Import From File ダイアログの Data Specs ページ

Data Specs ページ、From グループ

File Name 以下のいずれかを実行して、インポートしたいファイルを指定します。

- **Browse** をクリックし、目的のファイルに移動します。
- そのファイルが現在のプロジェクトフォルダ内にある場合は、ファイル名だけを拡張子も含めて入力します。
- そのファイルが現在のプロジェクトフォルダ内がない場合は、ファイルのフルパス名を入力します。

File Format このフィールドのドロップダウンリストからファイルの種類を選択します。

ヒント

ほとんどの場合、**Browse** を使ってインポートするファイルを指定した場合は、S-PLUS が自動的に正しいフォーマットタイプを選択します。

データファイルからのインポートとデータファイルへのエクスポート

Data Specs ページ、To グループ

Data set デフォルトでは、S-PLUS は、インポートしたファイルと同じ名前で新しい S-PLUS データセットを作成します。デフォルト名を無効にするには、このフィールドに異なる名前を入力します。既存のデータセットにデータをインポートするには、そのデータセットの名前を入力または選択します。

注意

Data set フィールドで指定した名前が既存のオブジェクトと同じ名前である場合、デフォルトでは、上書きの確認メッセージが表示されます。この機能をオフにするには、メインメニューから、**オプション ▶ 設定**を選択し、**Data** タブをクリックし、**Prompt on overwrite** チェックボックスの選択を解除してください。

Create new data set このオプションを選択すると、**Data set** フィールドで指定した名前で新しいデータセットが作成されます。

Add to existing data set このオプションを選択すると、**Data set** フィールドで指定した既存のデータセットにデータがインポートされます。

Start col デフォルトでは、**Add to existing data set** を選択すると、インポートしたデータがデータセットの末尾に追加されます。異なる開始列を指定するときは、このフィールドでその列の名前を入力または選択します。

注意

Import From File ダイアログの **Data Specs** ページで選択を行った後は、S-PLUS がデータファイルのインポートに必要な基本情報を記憶しています。インポート用パラメータをさらに調整するには、以下で説明する **Options** および **Filter** ページを使用してください。

Import From File ダイアログの **Options** ページを、図 5.2 に示します。

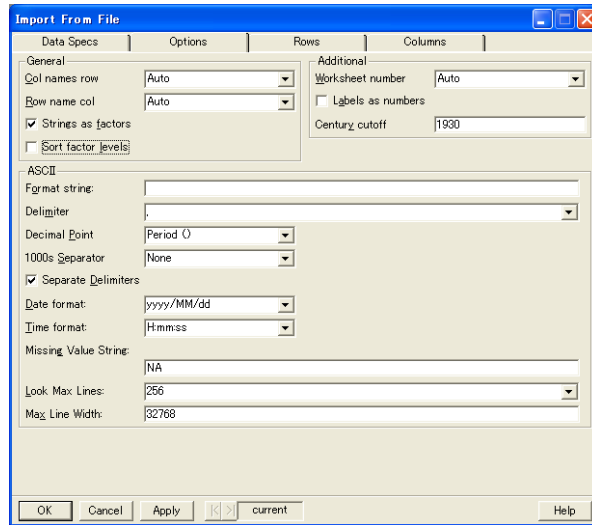


図 5.2 : Import From File ダイアログの Options ページ

Options ページ、General グループ

Col names row インポートするファイルのある行に列名が含まれている場合は、その列名の入った行番号に対応する数字を入力してください。デフォルトでは、インポートした先頭行から、S-PLUS が適切な列名の作成を試みます。

Row name col インポートするファイルのある列に行名が含まれている場合は、その行名の入った列番号に対応する数字を入力してください。

Strings as factors このオプションが選択されると、データファイルをインポートするときに、すべての文字列が因子変数に変換されます。選択しないときは、文字列はクラス `character` でインポートされます。

Sort factor levels このオプションが選択されると、文字列から作成されたすべての因子変数の水準が（アルファベット順に）ソートされます。選択しないときは、データファイルから読み込んだ順番で水準が定義されます。

注意

S-PLUS ではアンダースコア文字が予約文字になっているため、行または列の名前をインポートするとき、アンダースコア（“_”）はピリオド（“.”）に置き換えられます。

データファイルからのインポートとデータファイルへのエクスポート

Options ページ、Additional グループ

Worksheet number Excel や Lotus のようなスプレッドシートからデータをインポートするときは、インポートしたいワークシートの番号を指定します。

Labels as numbers このオプションが選択されると、ラベルを持つ SAS および SPSS の変数が、数字としてインポートされます。選択しないときは、値ラベルがデータとしてインポートされます。

Century cutoff ASCII テキストファイルをインポートするときは、このフィールドに日付の起点を 2 桁で指定します。この数値で始まる 100 年間に対して、年数を 2 桁で表した日付が割り当てられます。デフォルト値は 1930 であるため、6/15/30 は 1930 年 6 月 15 日と読まれ、また 12/29/29 という日付は、2029 年 12 月 29 日と解釈されます。

Options ページ、ASCII グループ

Format string このフィールドは、固定長 ASCII (FASCII) テキストファイルをインポートするときに必要となります。フォーマット文字列が、データ形式とインポートした列のフォーマットを指定します。このフィールドが受け入れるシンタックスの詳細については、192 ページの「フォーマット文字列」を参照してください。

Delimiter ASCII ファイル内で要素の区切りに使用するすべての文字列を指定します。デフォルトでは、カンマ (“,”) を使用します。

Decimal Point 小数点に使う文字を指定します。

1000s Separator 1000 ごとの区切り文字を指定します。

Separate Delimiters このオプションが選択されると、区切り記号が厳密に 1 つの文字として扱われます。選択しないときは、連続した区切り文字列は 1 つの区切り記号として扱われます。

Date format 日付データをインポートするときに使用したいフォーマットを選択します。利用可能な選択肢は、Windows の「地域のプロパティ」(Windows Regional Options) と同じ内容です。つまり、このフィールドのデフォルト値は Windows の現在のデフォルトです。

第5章 データのインポートとエクスポート

Time format 時間データをインポートするときに使用したいフォーマットを選択します。利用可能な選択肢は、Windows の「地域のプロパティ」と同じ内容です。つまり、このフィールドのデフォルト値は Windows の現在のデフォルトです。

Missing Value String 欠損値としたい文字列を指定します。

Look Max Lines 各列のデータ形式を決めるのに、何行目までを見るかを指定します。

Max Line Width 1 行の最大幅を指定します。

注意

ASCII ファイルをインポートするためのデフォルトの区切り記号を変更するには、メインメニューから、オプション ► 設定を選択し、Data タブをクリックします。ASCII Import/Export Options グループで、Import Delimiter フィールドに、デフォルトとして使用したい区切り記号を入力してください。

Import From File ダイアログの Rows ページを、図 5.3 に示します。

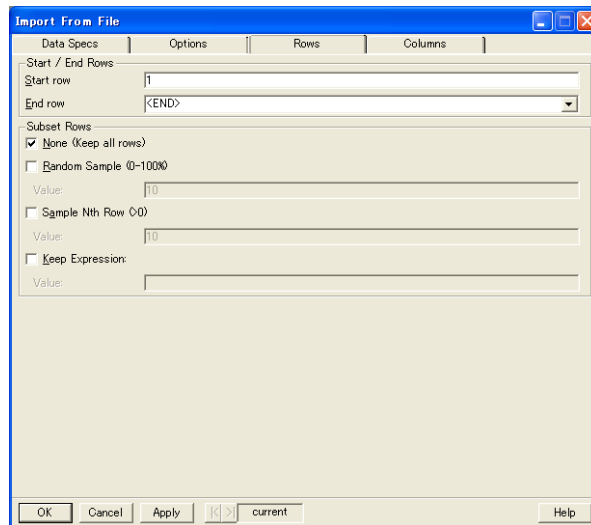


図 5.3 : Import From File ダイアログの Rows ページ

Start row ファイルからインポートする最初のデータ行に対応する数字を入力します。デフォルトでは、ファイルの第 1 行から読み込みを開始します。

End row ファイルからインポートする最後のデータ行に対応する数字を入力します。デフォルトでは、ファイルの最後の行になっています。

データファイルからのインポートとデータファイルへのエクスポート

Random Sample データから全体の何パーセントをランダムに抽出するかを指定することができます。パーセンテージを数値で入力します。

Sample Nth Row データを何個おきに抽出するかを指定することができます。正の整数を入力します。

Keep Expression データを抽出する条件式を指定します。デフォルトでは、このフィールドは空白になっており、すべてのデータがインポートされます。このフィールドで指定可能なシンタックスについては、189 ページの「フィルタ式」を参照してください。

Import From File ダイアログの **Columns** ページを、図 5.4 に示します。

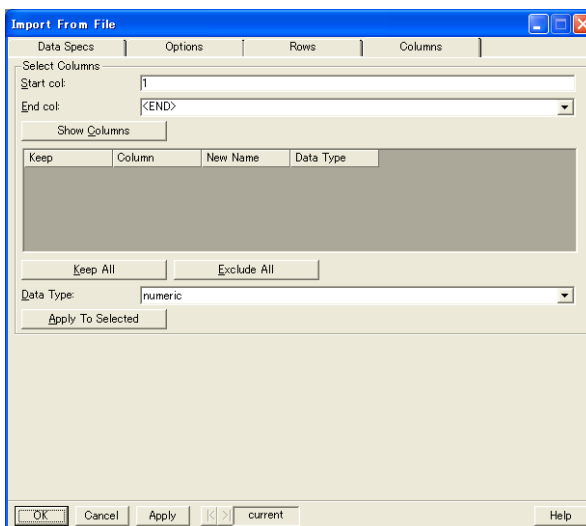


図 5.4 : Import From File ダイアログの Columns ページ

Start col ファイルからインポートする最初のデータ列に対応する数字を入力します。デフォルトでは、ファイルの第 1 列から読み込みを開始します。

End col ファイルからインポートする最後のデータ列に対応する数字を入力します。デフォルトでは、ファイルの最後の列になっています。

Show Column ボタンをクリックすると、現在読み込まれるデータの列名、データタイプが表示されます。一番左に表示されるチェックボックスのチェックをはずすと、その列はインポートされません。

Apply To Selected ボタンをクリックすると、すべての列がすぐ上の Data Type フィールドで指定された形式でインポートされます。

データ ファイルへ エクスポート する

下の図 5.4 に示す **Export To File** ダイアログは、**Data Specs**、**Options**、および **Filter** というラベルの付いた、3 つのタブ付きページから構成されています。

- メインメニューから、ファイル ▶ エクスポート ▶ ファイルに書き込みを選択します。

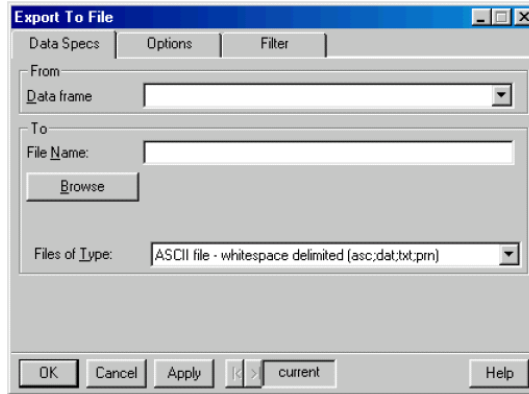


図 5.5 : Export To File ダイアログの Data Specs ページ

Data Specs ページ、From グループ

Data frame エクスポートしたいデータセットの名前を入力または選択します。

Data Specs ページ、To グループ

File Name デフォルトでは、S-PLUS がエクスポートされたデータセットと同じ名前で新しいデータファイルを作成し、そのファイルを現在使用中のプロジェクトフォルダに保存します。他の場所にファイルを保存するには、**Browse** をクリックして希望のフォルダに移動します。また、デフォルトのファイル名を他の名前に変更することもできます。

注意

File Name フィールドで指定したファイル名が既存ファイルの名前である場合は、上書きしてよいかどうか確認するメッセージが表示されます。

Files of Type このフィールドのドロップダウンリストから、エクスポートしたファイルに適用したいフォーマットを選択します。ファイルの拡張子は、自動的に **File Name** フィールドに反映されます。

注意

Export To File ダイアログの **Data Specs** ページ上で選択を行った後は、S-PLUS がデータセットのエクスポートに必要な基本情報を記憶しています。エクスポート用のパラメータをさらに調整するには、以下に説明する **Options** および **Filter** ページを使用してください。

Export To File ダイアログの **Options** ページを、図 5.5 に示します。

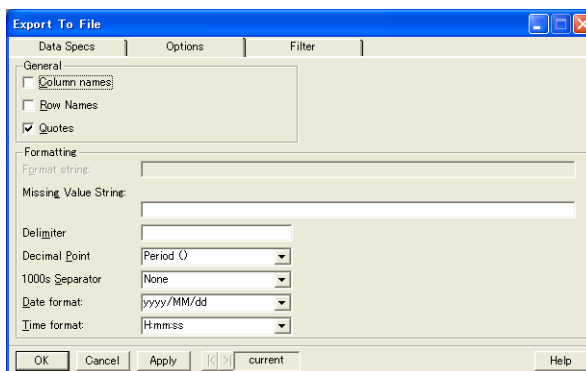


図 5.6 : Export To File ダイアログの Options ページ

Options ページ、General グループ

Column names このオプションを選択すると、データセットの列名が、ファイルの最初の行として使用されます。

Row Names このオプションを選択すると、データセットの行名が、ファイルの最初の列として使用されます。

Quotes このオプションを選択すると、因子および文字変数が、引用符付きの文字列としてエクスポートされます。

Options ページ、Formatting グループ

Missing Value String NA を置き換えたい文字を指定します。

Format string ASCII テキストファイルにエクスポートするとき、エクスポートする列のデータ形式とフォーマットを指定します。このフィールドで使用できるシンタックスについては、191 ページの「特定タイプのファイルのインポートとエクスポートに関する注意」を参照してください。

Delimiter ASCII テキストファイルにエクスポートするとき、ファイル内の各要素を区切るための区切り記号を指定します。デフォルトでは、カンマ(“,”)を使用します。

Decimal Point 小数点に使う文字を指定します。

1000s Separator 1000 ごとの区切り文字を指定します。

Date format 日付データをエクスポートするときを使用したいフォーマットを選択します。利用可能な選択肢は、Windows の「地域のプロパティ」と同じ内容です。つまり、このフィールドのデフォルト値は、Windows の現在のデフォルトです。

Time format 時間データをエクスポートするときを使用したいフォーマットを選択します。利用可能な選択肢は、Windows の「地域のプロパティ」と同じ内容です。つまり、このフィールドのデフォルト値は、Windows の現在のデフォルトです。

Export To File ダイアログの **Filter** ページを、図 5.7 に示します。

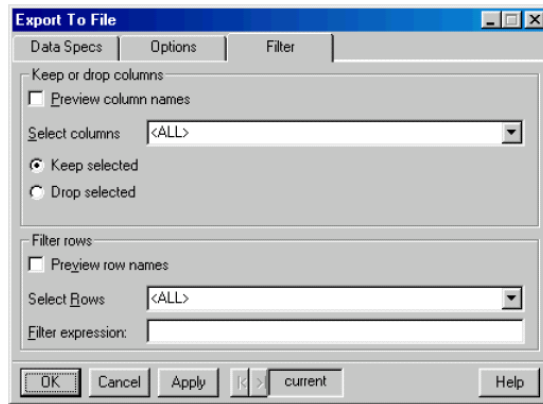


図 5.7 : Export To File ダイアログの Filter ページ

Filter ページ、Keep or drop columns グループ

Preview column names このオプションを選択すると、**Select columns** ドロップダウンリストに、エクスポート可能なデータセットの列名が表示されます。

Select columns **Keep selected** または **Drop selected** の選択により、それぞれエクスポートしたい列またはエクスポートしたくない列を指定します。**Preview column names** が選択されていると、このフィールドはドロップダウンリストとなり、希望する列を選択することができます。**Preview column names** が選択されていないときは、このフィールドに名前前列リスト (36 ページの「列リストを作成する」を参照) を入力してください。

Keep selected このオプションを選択すると、**Select columns** ドロップダウンリストで指定した列だけをエクスポートします。

データファイルからのインポートとデータファイルへのエクスポート

Drop selected このオプションを選択すると、**Select columns** ドロップダウンリストで指定した列を除くすべての列をエクスポートします。

Filter ページ、Filter rows グループ

Preview row names このオプションを選択すると、**Select Rows** ドロップダウンリストに、エクスポート可能なデータセットの行名が表示されます。

Select Rows このフィールドにエクスポートしたい行を選択します。

Preview row names が選択されていると、このフィールドは、ドロップダウンリストになり、希望する行を選択することができます。**Preview row names** が選択されていないときは、このフィールドに数字の行リスト（40 ページの「行リストを作成する」を参照）を入力してください。

Filter expression クエリまたはフィルタ式を指定することにより、エクスポートするデータのサブセットを作成することができます。デフォルトでは、このフィールドは空白になっており、すべてのデータがエクスポートされます。このフィールドで使用できるシンタックスについての詳細は、188 ページの「フィルタ式」を参照してください。

ODBC テーブルからのインポートと ODBC テーブルへのエクスポート

Microsoft Access や Excel などのアプリケーションと（一般にデータソースとして認知されている）ほとんどの商用データベースは Open DataBase Connectivity (ODBC) 規格をサポートしています。ODBC は、データベース間でデータを交換するための標準的な方法を提供しており、幅広くサポートされています。各アプリケーションは、一般に、ODBC インタフェースを介したデータの受入れまたは配布を可能にする ODBC ドライバを備えています。S-PLUS は、ODBC のバージョン 2.0 と 3.0 をサポートしています。

S-PLUS では、ODBC データベースのインポートとエクスポート、および ODBC 規格をサポートしているアプリケーションのために、便利なダイアログを用意しています。デフォルトでは、テーブル全体が S-PLUS のデータフレームにインポートされ、そのデータがデータ・ウィンドウに表示されます。同様に、特に指定がない限り、データセット全体が ODBC テーブルにエクスポートされます。

注意

インポートしたデータを自動的にデータ・ウィンドウに表示したくない場合は、メインメニューから、**オプション ▶ 設定**を選択し、**Data** タブをクリックし、**Show Imported Data in View** チェックボックスの選択を解除してください。

ODBC Data Source Administrator

ODBC Data Source Administrator は、データベースドライバおよびデータソースの管理を行います。以下の説明に進む前に、ご使用のコンピュータに Administrator をインストールしてください。

ご使用のコンピュータには、ODBC Data Source Administrator がすでにインストールされているかも知れません。これを確認するには、**スタートメニューから設定 ▶ コントロールパネル ▶ 管理ツール**を選択し、**データソース (ODBC)** という名前のアイコンがあるか確認してください。Windows98 または NT で使用している場合は、**設定 ▶ コントロールパネル**を選択し、**32bit ODBC** または **ODBC** があるか確認してください。

データソースがすでにインストールされている場合は、バージョンを 2.0 から 3.0 にアップグレードしたい場合を除き、この節の以下の部分はスキップすることができます。（バージョンを確認するには、データソースを起動し、**バージョン情報**タブを選択し、アドミニストレータを見ます。）

ODBC テーブルからのインポートと ODBC テーブルへのエクスポート

ODBC データソースがインストールされていない場合は、<ftp://ftp.insightful.com/outgoing/techsup/odbc> にアクセスし、**Install.exe** を実行します。

ODBC ドライバ

ODBC「ドライバ」は、アプリケーションまたはデータベースを ODBC インタフェースと接続するダイナミック リンク ライブラリ (DLL) です。アプリケーションが ODBC インタフェース内の関数を呼び出し、その関数が、データベース固有のドライバによって実行されます。プリンタのドライバがプリンタに固有のコマンドから文書処理プログラムを分離する場合と同じように、ドライバを使用することによってデータベース固有の呼出しからアプリケーションを分離します。

S-PLUS をインストールすると、S-PLUS に接続する ODBC ドライバも自動的にインストールされますが、ご使用のデータベースに対応した適切な ODBC ドライバをユーザが準備する必要があります。ご不明な点は、データベースの販売元またはサードパーティーの ODBC ドライバ販売元にお尋ねください。

ご使用のコンピュータにどのドライバがインストールされているかを確認するには、データソースを起動し、**ドライバタブ**を選択してください。コンピュータにインストールされている各 ODBC ドライバの、名前、バージョン、会社名、ファイル名、およびファイルの作成日付などが表示されます。新しいドライバの追加、またはすでにインストールされているドライバの削除を行うには、そのドライバのセットアッププログラムを使用します。

データソース を定義する

「データソース」は、データの保存場所またはデータベースの論理名です。このデータソースは、アクセスしたいデータ、データを使用しているアプリケーション、およびデータに到達するのに必要なコンピュータやネットワーク接続を指定します。データソースの追加または構成作業は、データソースを使用するか、または S-PLUS 内で行うことができます (オンラインヘルプを参照してください)。

データソースを追加するには、**コントロールパネル**内にあるデータソースアイコンをダブルクリックしてアドミニストレータを起動します。アドミニストレータ 3.0 をご使用の場合は、作成したい DSN (データソース名) のタイプに対応するタブを選択します。この DSN のタイプによって、作成しているデータソースへのアクセスを、次のように制御します。

- **ユーザー DSN** は、DSN が作成されたときに有効になるログインアカウントに固有の DSN タイプです。これらの DSN は、それぞれのコンピュータに限定され、カレントユーザ専用となります。

- **システム DSN** は、そのコンピュータに限定されますが、特定のユーザ専用ではありません。ログイン特権を持つユーザであれば誰でも、システム DSN を使用してセットアップしたデータソースを使用することができます。
- **ファイル DSN** は、ファイルベースのデータソースであり、同じドライバをインストールしているすべてのユーザが共有できます。これらのデータソースは、1人のユーザ専用ではなく、また1つのコンピュータに限定されることもありません。

該当するタブを選択し、**追加**ボタンをクリックしてください。(アドミニストレータ2.0をご使用の場合は、最初のダイアログから**追加**ボタンをクリックすることによってユーザーDSNを作成することができます。また、システムDSNを作成するときは、同じボタンを選択してから、次に現れたダイアログで追加ボタンを選択します。ファイルDSNは、アドミニストレータ3.0でのみ作成可能です。) **データソースの新規作成**ダイアログが表示されます。

接続したいデータベース用の ODBC ドライバを選択し、**完了**をクリックします。**データソースの新規作成**ダイアログ内にあるドライバのリストが空の場合、またはご使用のデータベースあるいはアプリケーション用のドライバがリストに含まれていない場合は、データベースまたはその ODBC ドライバをインストールする必要があります。

この時点で、ドライバに固有のダイアログが表示され、データベースおよびそのデータベースに接続するのに必要なドライバ固有の情報を求めてくるはずですが、必要なフィールドに入力し、**OK** をクリックします。次回からは、S-PLUS から **Import From ODBC** または **Export to ODBC** ダイアログを起動すると、新しいデータソースが表示されるはずですが、

ODBC テーブルから インポート する

下の図 5.7 に示す **Import From Database** ダイアログは、ODBC および **Filter** ラベルの付いた 2 つのタブ付きページで構成されています。ダイアログを開くには、以下の操作を行います。

- メインメニューから、**ファイル ▶ インポート ▶ ODBC 接続からの読み込み**を選択します。

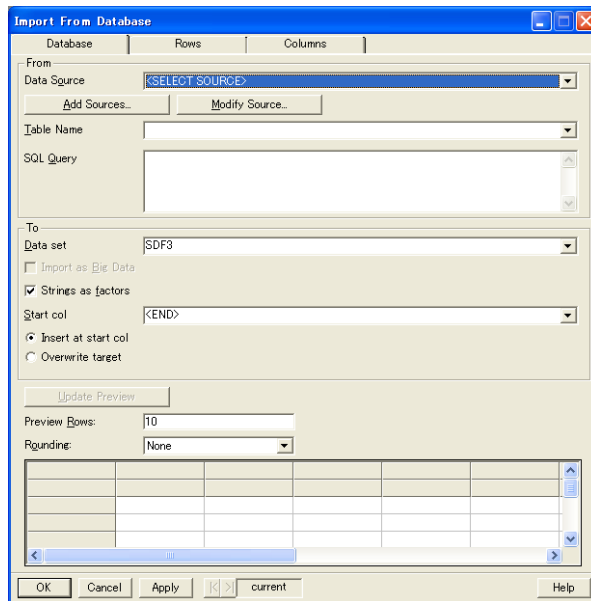


図 5.8 : Import From ODBC ダイアログの ODBC ページ

ODBC ページ、From グループ

Data Source このフィールドのドロップダウンリストから、希望のデータソースを選択します。

注意

データソースは、アクセスしたいデータ、データを使用しているアプリケーション、およびデータに到達するのに必要なコンピュータやネットワーク接続で構成されています。**Data Source** ドロップダウンリストに希望のデータソースが表示されていない場合やリストが空白の場合は、データソースを1つ以上構成しなければならない場合があります。構成を行うときは、コントロールパネル内の使用可能な ODBC アプレットを使用するか、または **Import From ODBC** ダイアログで **Add Sources** (または **Modify Source**) ボタンをクリックします。データソースの追加や変更については、オンラインヘルプを参照してください。

Table Name データソースを選択した後、このフィールドのドロップダウンリストにテーブル名が表示されます。インポートしたいテーブルを選択してください。デフォルトでは、データソースの最初のテーブルが選択されます。

SQL Query このフィールドに、任意の Structured Query Language (SQL) を指定してください。インポートするテーブルを選択すると、テーブル内のデータをすべてインポートするためのデフォルトの SQL クエリが生成されます。

ODBC ページ、To グループ

Data frame デフォルトでは、S-PLUS は、インポートしたテーブルと同じ名前の新規データセットを作成します。デフォルト名を変更するときは、このフィールドに違う名前を入力します。既存のデータセットにデータをインポートするときは、その名前を入力または選択してください。

Start col デフォルトでは、S-PLUS は、インポートしたデータを **Data frame** フィールドで指定したデータセットの末尾に追加します。異なる開始列を指定するには、このフィールドでその列名を入力または選択します。

Insert at start col このオプションを選択すると、S-PLUS は、インポートしたデータを **Start col** フィールドで指定した列に挿入します。

Overwrite target このオプションを選択すると、S-PLUS は、インポートするときに既存のデータを上書きします。

注意

Import From ODBC ダイアログの **ODBC** ページで選択を行った後は、S-PLUS が ODBC テーブルのインポートに必要な基本情報を記憶しています。インポート用パラメータをさらに調整するには、以下に説明する **Filter** ページを使用してください。

Import From ODBC ダイアログの **Rows** ページを図 5.9 に示します。

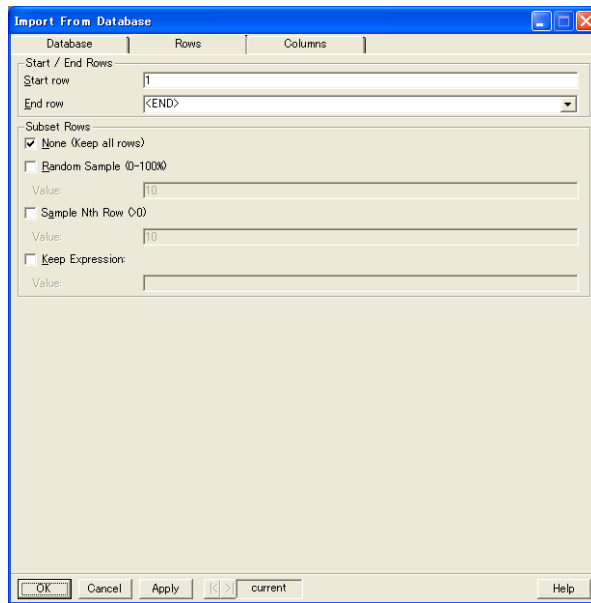


図 5.9 : Import From ODBC ダイアログの Filter ページ

Start row ファイルからインポートする最初のデータ行に対応する数字を入力します。デフォルトでは、ファイルの第 1 行から読み込みを開始します。

End row ファイルからインポートする最後のデータ行に対応する数字を入力します。デフォルトでは、ファイルの最後の行になっています。

Random Sample データから全体の何パーセントをランダムに抽出するかを指定することができます。パーセンテージを数値で入力します。

Sample Nth Row データを何個おきに抽出するかを指定することができます。正の整数を入力します。

Keep Expression データを抽出する条件式を指定します。デフォルトでは、このフィールドは空白になっており、すべてのデータがインポートされます。このフィールドで指定可能なシンタックスについては、189 ページの「フィルタ式」を参照してください。

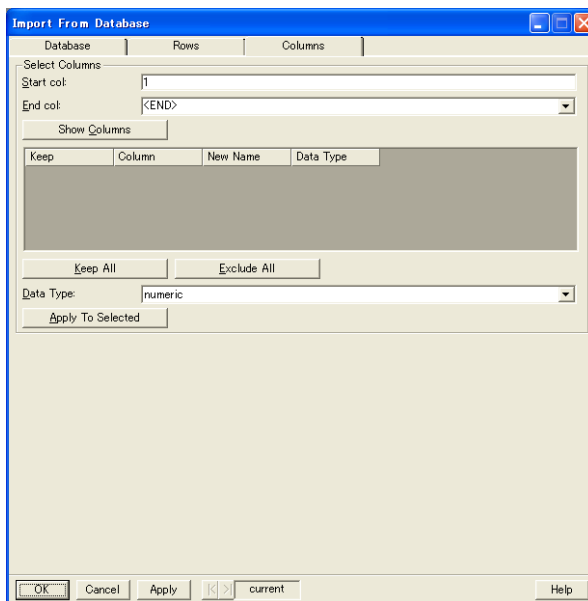


図 5.10 : Import from Database ダイアログの Columns ページ

ODBC テーブルへ エクスポート する

下の図 5.11 に示す **Export to Database** ダイアログは、**ODBC** および **Filter** ラベルの付いた 2 つのタブ付きページで構成されています。ダイアログを開くには、以下の操作を行います。

- メインメニューから、ファイル ▶ エクスポート ▶ ODBC 接続に書き込みを選択します。

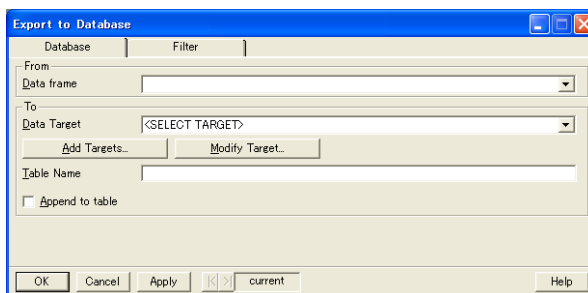


図 5.11 : Export to ODBC ダイアログの ODBC ページ

ODBC ページ、From グループ

Data frame エクスポートしたいデータセットの名前を、入力または選択します。

ODBC テーブルからのインポートと ODBC テーブルへのエクスポート

ODBC ページ、To グループ

Data Target このフィールドのドロップダウンリストから、希望するデータターゲットを選択します。

注意

データターゲットは、エクスポートするときに使用し、インポートするときのデータソースに相当するものです。**Data Target** ドロップダウンリストに希望のデータターゲットが表示されない場合、またはリストが空白の場合は、データターゲットを1つ以上構成しなければならない場合があります。この構成を行うときは、コントロールパネル内の使用可能な ODBC アプレットを使用するか、または **Export to Database** ダイアログで **Add Targets** (または **Modify Target**) ボタンをクリックします。データターゲットの追加や修正については、オンラインヘルプを参照してください。

Table Name デフォルトでは、エクスポートされたデータセットと同じ名前の新しい ODBC テーブルが作成されます。好みにより、デフォルトのテーブル名を他の名前に変更することができます。

注意

Export to Database ダイアログの **ODBC** ページで選択をした後は、S-PLUS が ODBC テーブルのエクスポートに必要な基本情報を記憶しています。エクスポート用のパラメータをさらに調整するには、以下で説明する **Filter** ページを使用します。

Export to ODBC ダイアログの **Filter** ページを図 5.10 に示します。

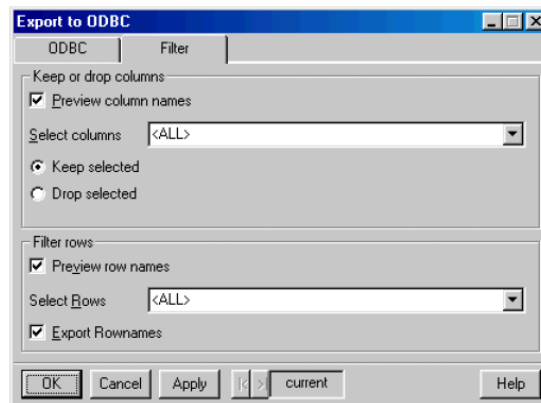


図 5.12 : Export to ODBC ダイアログの Filter ページ

Filter ページ、Keep or drop columns グループ

Preview column names このオプションを選択すると、**Select columns** ドロップダウンリストに、エクスポートするデータセットの列名が表示されます。

Select columns **Keep selected** または **Drop selected** を選択することにより、それぞれエクスポートしたい列またはエクスポートしたくない列を指定します。**Preview column names** が選択されているときは、このフィールドはドロップダウンリストとなり、希望する行を選択することができます。**Preview column names** が選択されていないときは、このフィールドに名前の列リスト（36 ページの「列リストを作成する」を参照）を入力してください。

Keep selected このオプションを選択すると、S-PLUS は、**Select columns** ドロップダウンリストで指定した列だけをエクスポートします。

Drop selected このオプションを選択すると、S-PLUS は、**Select columns** ドロップダウンリストで指定した列を除くすべての列をエクスポートします。

Filter ページ、Filter rows グループ

Preview row names このオプションを選択すると、**Select Rows** ドロップダウンリストに、エクスポートするデータセットの行名が表示されます。

Select Rows このフィールドに、エクスポートしたい行を指定します。**Preview row names** が選択されているとき、このフィールドは、ドロップダウンリストになり、希望する行を選択することができます。**Preview row names** が選択されていないときは、このフィールドに行リストの番号（40 ページの「行リストを作成する」を参照）を入力してください。

Export Rownames このオプションを選択すると、データセットの行名が、ODBC テーブルの最初の列として使用されます。

フィルタ式

インポートおよびエクスポートダイアログの **Filter expression** フィールドに論理式を含めることによって、インポートまたはエクスポートするデータのサブセットを作成することができます。フィルタ式は、ファイル内の元の列名の形で記述しなければならない、**Col names row** フィールドで指定された変数名の形で記述してはいけません。

また、フィルタは、**S-PLUS** によって評価されないことにも注意してください。つまり、**S-PLUS** に組み込まれている関数（たとえば `mean` など）を含む式は使用できません。この規則の 1 つの特別な例外は、欠損値を扱うもので、論理式の中の欠損値を示すために `NA` を使用することはできますが、たとえば `is.na` や `na.exclude` のような **S-PLUS** 固有の関数は使用することができません。

変数式

変数式を作成するには、単一の変数、または式を含む複数の変数を指定します。使用可能な通常の算術演算子 `[+ - * / ()]` に加えて、表 5.2 に、**Filter expression** フィールドで受け入れることができる関係演算子を示します。

表 5.2 : 関係演算子

演算子	説明
<code>==</code>	～と等しい
<code>!=</code>	～と等しくない
<code><</code>	～より小さい
<code>></code>	～より大きい
<code><=</code>	～以下
<code>>=</code>	～以上
<code>&</code>	および
<code> </code>	または
<code>!</code>	否定

第5章 データのインポートとエクスポート

たとえば、id 列内に欠損値を持たないすべての行を選択するときは、下のように入力します。

```
id != NA
```

10歳で体重が150ポンド未満の子供に対応するすべての行をインポートするには、下のように入力します。

```
Age == 10 & Weight < 150
```

注意

フィルタ式を作成するときは、必ず関係演算子の左側に変量名を入力するように注意してください。たとえば `12 < Age` ではなく、`Age > 12` と入力します。

また、`?`（単一文字用）および`*`（任意の長さを持つ文字列用）のワイルドカード文字を使用して、文字の部分抽出を選択することもできます。たとえば以下の文字列は、

```
account == ?????22
```

`account` 変量の長さが文字 6 個分で、`22` で終了しているすべての行を選択します。また、下の式は、

```
id == 3*
```

`id` が 3 で始まる行を、文字列の長さに関係なくすべて選択します。

特定の行数をインポートするには、組み込み変量`@rownum`を使用します。たとえば下の式は、

```
@rownum < 200
```

データファイルの先頭から 199 行までをインポートします。

特定タイプのファイルのインポートとエクスポートに関する注意

この節では、データを特定のファイルフォーマットでインポートおよびエクスポートするときに役立つ追加の説明を行います。

ASCII (可変長 ASCII) ファイル

データの行は、必ず復帰改行で終わっていなければなりません。複数の区切り記号はグループ化されず、1つの区切り記号と同じように扱われます。たとえば、カンマが区切り記号の場合、2つ続いたカンマは、欠損フィールドとして解釈されます。

二重引用符 ("") の扱いは特別です。これは常に 1 つの“囲み”マークとして扱われ、必ず対でなければなりません。二重引用符に囲まれたデータは、1つの単位の文字データとして読み込まれます。したがって、文字フィールドが二重引用符で囲まれていれば、その文字フィールド内でスペースとカンマを使用することができますので、スペースとカンマを区切り記号として使用することができます。なお、二重引用符は、通常の区切り記号としては使用できません。

ASCII にエクスポートするときは、列名が 34 文字を超えるとその分は切りつめられることに注意してください。

dBASE ファイル

dBASE および dBASE 互換ファイルをインポートするとき、指定しなければならないのは、多くの場合、ファイル名とファイルタイプだけです。(列名とデータ形式は dBASE ファイルから取得されます。) また、最初と最後それぞれの列と行を指定することによって、データの四角形のサブセットを選択することができます。

複数のテーブルを含むファイル

複数のテーブルまたはデータをサポートすることができるアプリケーション (たとえば、Informix、Microsoft Access、Microsoft SQL Server、Oracle、SAS、SigmaPlot、SYBASE など) は、複数のテーブルまたはデータを 1 つのファイルにエクスポートすることができます。S-PLUS は現在、ファイルの種類が ODBC の場合を除き、ファイルからの最初のテーブルのインポートだけをサポートしています。

FASCII (固定長 ASCII) ファイル

FASCII インポートを使用して、インポートしたファイル内のそれぞれの文字をどのように扱うかを指定することができます。たとえば、ファイル内の行が改行で区切られていない場合、またはファイル内の各データ行が複数の行に分かれている場合は、区切り記号で区切られていない固定幅の列に、FASCII を使用してください。

各行が復帰改行で終わる場合、復帰改行は、スキップすべき1つの文字幅変量として扱われます。

行の一部分だけをインポートしたい場合は、その最初と最後の行を指定してください。

フォーマット文字列

フォーマット文字列は、FASCII (固定長 ASCII) テキストファイルからデータをインポートするとき、または FASCII ファイルにデータをエクスポートするときに使います。また、フォーマット文字列は、インポートしたファイルの各文字をどのように扱うかを指定します。データファイルの各列が、区切り記号で区切られていない場合は、フォーマット文字列を、FASCII ファイルと一緒に使用してください。

データをインポートするためのフォーマット文字列

FASCII ファイルからデータをインポートする場合、**Import From File** の **Options** ページの **Format string** で有効なフォーマット文字列として、ファイル内の各列に、パーセント記号 (%) に続けてデータ形式を指定してください。使用できるデータ形式は、次の通りです。

- 文字列を示す **s**
- 数値を示す **f**
- スキップする列を示すアスタリスク (*)

文字列内のそれぞれの指定は、**Delimiter** フィールドで指定した文字のうちの1つで区切られなければなりません。たとえば、次のフォーマット文字列は、

```
%s, %f, %*, %f
```

データファイルの1番目の列を形式 **character** としてインポートし、2番目と4番目の列を形式 **numeric** としてインポートし、3番目の列をすべてスキップします。

特定タイプのファイルのインポートとエクスポートに関する注意

変量を形式 `numeric` として指定し、セルが数字と解釈できない場合、そのセルは、欠損値で埋められます。また、不完全な行も、欠損値で埋められます。

注意

テキストファイル内の日付が、自動的に数字としてインポートされる場合があります。日付を含むデータをインポートした後は、各列のクラスを確認し、必要に応じて適切なデータ形式に変更しなければなりません。

フォーマット文字列とフィールド幅指定は、通常の ASCII ファイルには無関係であり、したがって無視されます。しかし、FASCII ファイルでは、各フィールドの幅を定義する整数値を指定することができます。たとえば、次のフォーマット文字列は、

```
%4f, %6s, %3*, %6f
```

各行内の最初の 4 つのエントリを数値列としてインポートします。各列の次の 6 つのエントリは文字として読み取られ、次の 3 つのエントリはスキップされ、さらに次の 6 つのエントリは、もう 1 つの数値列としてインポートされます。

データをエクスポートするためのフォーマット文字列

FASCII ファイルにデータをエクスポートするとき、**Format string** フィールドに受け入れられるシンタックスは、データをインポートする場合と似ています。しかし、データ形式の他に、数値の精度も指定することができます。たとえば、次のフォーマット文字列は、

```
%3, %7.2, %4, %5.2
```

1 番目と 3 番目の列を、それぞれ 3 桁と 4 桁の整数としてエクスポートします。2 番目と 4 番目の列は、それぞれ小数点以下 2 桁の精度です。

精度値を指定しない場合は、ゼロと見なされます。また、文字の列に精度値を入れた場合は無視されます。行名をエクスポートするときは、フォーマット文字列の最初のエントリがその行の名前として用いられることに注意してください。

フォーマット文字列を指定することにより、多数の文字列を含むデータのエクスポートを短時間で行うことができます。ユーザがフォーマット文字列を含めない場合は、**S-PLUS** が、文字列または因子列内のすべてのエントリの幅をチェックし、その列内のすべての値に対して幅が十分かどうかを決定する必要があります。サポートされているファイルタイプの多くが固定幅を使用しています。したがって、短い値が多く長い値が少ない文字の列で幅を狭く指定することにより、かなり大きなスペースを節約することができます。この方法を使用すると、いくつかの長い値は切りつめられます。

Informix ファイル

Informix ファイルからのインポートおよび Informix ファイルへのエクスポートはすべて、ODBC を使用して行われるため、様々な ODBC コンポーネントが正しくインストールされていなければなりません。

Lotus ファイル

Lotus 型のワークシートが、ワークシートの第 1 の行と列から始まる四角形ブロックに数値データだけを含む場合は、ファイル名とファイルタイプのみを指定します。行が列名を含む場合は、**Col names row** フィールドにその行番号を指定してください（第 1 行でなくてもかまいません）。最初と最後の列と行を指定することによって、ワークシートの四角形サブセットを選択することができます。Lotus スタイルの列名（たとえば、**A**、**AB**）を使用して、最初と最後の列を指定することができます。

最初の行として指定された行は、列のデータ形式を確認するために必ず最初に読み込まれます。したがって、この行に空白のセルがあってははいけません。他の行では空白のセルは、欠損値で埋められます。

Microsoft Access ファイル

Access ファイルからのインポートおよび Access ファイルへのエクスポートはすべて、ODBC を使用して行われるため、様々な ODBC コンポーネントが適切にインストールされていなければなりません。

Microsoft Excel ファイル

Excel のワークシートが、ワークシートの第 1 の行と列から始まる四角形ブロックで数値データだけを含む場合は、ファイル名とファイルタイプのみを指定します。行が列名を含む場合は、**Col names row** フィールドにその行番号を指定してください（第 1 行でなくてもかまいません）。最初と最後の列と行を指定することによって、ワークシートの四角形サブセットを選択することができます。Excel スタイルの列名（たとえば、**A**、**AB**）を使用して、最初と最後の列を指定することができます。

Excel にエクスポートするときは、列名が 34 を超えるとその分は切りつめられることに注意してください。

Oracle ファイル

Oracle ファイルからのインポートおよび Oracle ファイルへのエクスポートはすべて、ODBC を使用して行われるため、様々な ODBC コンポーネントが適切にインストールされていなければなりません。

Oracle にエクスポートするときは、テーブル名と列名は「大文字」でなければなりません。

SYBASE ファイル

SYBASE ファイルからのインポートおよび SYBASE ファイルへのエクスポートはすべて、ODBC を使用して行われるため、様々な ODBC コンポーネントが適切にインストールされていなければなりません。

第5章 データのインポートとエクスポート

第 6 章

グラフの編集

グラフ	199
グラフシート	199
グラフの作成方法	201
グラフの種類を変更する	202
グラフにプロットを追加する	203
グラフシートに複数のグラフを描く	205
3次元平面に2次元グラフを投射する	207
Trellis グラフ	209
グラフのフォーマット	213
グラフをフォーマットする：例	214
グラフシートをフォーマットする	219
グラフをフォーマットする	220
2次元軸をフォーマットする	222
2次元軸ラベルをフォーマットする	224
複数行テキストを追加しフォーマットする	225
タイトルと凡例を追加する	227
点にラベルを追加する	229
近似式を追加する	230
線、形、記号を追加する	231
グラフオブジェクトでの作業	232
図表示	234
グラフをフォーマットする（続き）	234
グラフスタイルを使用し色をカスタマイズする	237
グラフシートへのデータの埋込みと抽出	239
オブジェクトのリンクと埋込み	240
他のアプリケーションに S-PLUS グラフを埋め込む	241

第 6 章 グラフの編集

グラフを印刷する	243
グラフをファイルにエクスポートする	244

グラフ

S-PLUS の最も強力なツールの 1 つは、グラフを素早く簡単に作成する機能です。もしまだこの機能をお使いになっていない場合は、この章をお読みになる前にチュートリアル pdf ファイル『Getting Started GUIDE』を参照されることをお勧めします。S-PLUS は、様々な 2 次元グラフと 3 次元グラフを生成することができます。この章では、それらのうちの一部だけを取り上げます。その他のグラフに関するさらに詳しい情報は、第 4 章「グラフの作成」とオンラインヘルプを参照してください。

この章を読むにあたって

この章を最初から最後まで読む必要はありません。この章の大部分は、図表示の変更や軸のフォーマットなど、グラフのカスタマイズに関するある一定の操作を行う手順を説明しています。そのような作業を行う必要がない限り、読む必要はありません。

グラフの作成方法を理解するために、次のステップを踏まれることをお勧めします。

1. グラフの作成方法と図表示例の見方に関する基本的なことについては、pdf ファイル『Getting Started GUIDE』をお読みください。
2. 第 3 章「データの探索」の例をお読みください。
3. 基本用語とグラフの作成情報についてはこの節をお読みください。
4. この章の残りの部分をざっと読み、必要に応じて詳しく参照してください。

これで、S-PLUS のグラフィックス機能の概要をよく知ることができるはずです。

グラフシート

S-PLUS では、グラフシートとグラフ領域とプロット領域を区別しています。グラフシートは、プロットを描く用紙と説明するのが最も適当でしょう。印刷するときは、グラフシートの 1 ページまたは複数のページを印刷します。グラフシートは、複数のグラフを含むことができます。グラフ領域は、データ点/軸/凡例/グラフタイトルなどを取り囲む四角形を指します。プロット領域は、データをプロットするグラフ内の四角形の領域です。例は、図 6.1 を参照してください。

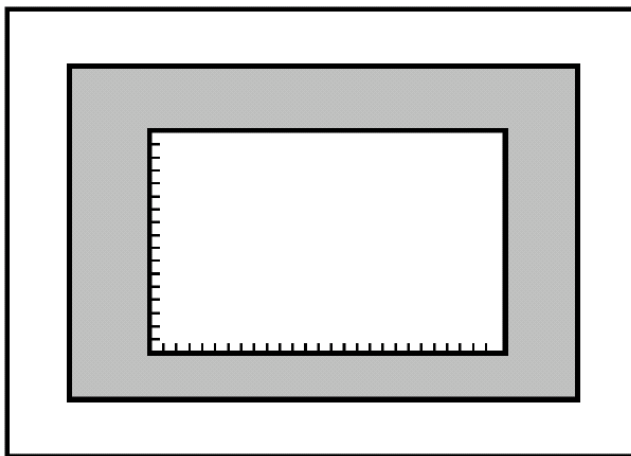






図 6.1 : グラフ領域（灰色）とプロット領域（中央）を含むグラフシート

作成する／
開く／
保存する／
印刷する

標準ツールバーの**新規ファイル**ボタン  を使って、新しい**グラフシート**を作成することができます。**グラフシート**を保存または印刷するときは、まずその**グラフシート**をクリックして選択します。次に、**保存**ボタン  を使って**グラフシート**を保存するか、**印刷**ボタン  を使ってプリンタで**グラフシート**を印刷します。前に保存した**グラフシート**を開くときは、**開く**ボタン  を使用してください（**グラフシート**には、ファイル拡張子**.SGR**が付いています）。また、これらの**グラフシート**の機能は**ファイルメニュー**から使用することもできます。

S-PLUS のグラフを、Windows のビットマップ（**.BMP**）、JPEG（**.JPG**）、Paintbrush（**.PCX**）などの画像ファイルのフォーマットにエクスポートしたいことがあるかもしれません。そのようなときは、メインメニューから**ファイル ▶ グラフのエクスポート**を選択してください。

グラフシートを
表示する



表示 ▶ ズームを選択することによって、**グラフシート**をズームして特定の領域に焦点を合わせることができます。メニューバー／ウィンドウタイトルバー／ツールバーを使わず、全画面でグラフを表示させるときは **F2** を押します。元の表示に戻すときはマウスをクリックするか、任意のキーボードボタンを押してください。**グラフシート**は、ズームされている場合でも常に 100% のサイズで印刷されることに注意してください。

グラフシートの表示速度を速めるために、ドラフトモードを使用することができます。**表示 ▶ ドラフトモードのオン・オフ**を選択することによって、このオプションをオンやオフに切り替えることができます。このオプションは、低速のコンピュータシステムを使用しているユーザに有効です。ドラフトモードは、グラフの印刷品質には影響しません。

グラフの作成方法



グラフを作成する方法にはいくつかあります。データを選択してプロットパレット・ボタンをクリックするか、プロットボタンをグラフ上にドラッグ アンド ドロップし、次にデータをプロットボタン上にドラッグするか、**挿入**メニューの**グラフ**オプションを使用することができます。まず、最初の方法を説明します。

プロットボタンを使用してグラフを作成する

グラフを素早く作成するために、**標準**ツールバーに **2D プロットボタン**  と **3D プロットボタン**  があります。**2D プロットボタン**か **3D プロットボタン**をクリックすると、プロットボタンのパレットが現われます。各プロットの説明が必要なときは、マウスカーソルをパレット内の各ボタン上に移動させてください。グラフの説明文が現われます。

プロットボタンを使って新しいグラフを作成するときは、新しい**グラフシート**が自動的に開かれます。



プロットボタンを使用してグラフを作成する

1. 利用可能な図表示のパレットを開くときは、**標準**ツールバーの **2D プロットボタン**  か **3D プロットボタン**  をクリックします。
2. プロットするデータを含む、**データ・ウィンドウ**または**オブジェクト・エクスプローラ**を開きます。
3. プロットしたいデータ列を選択します。連続していない列を選択するときは、**CTRL**-クリックを利用してください。列を選択した順序が、デフォルトのプロット順序となります。
4. パレット上の希望のプロットボタンをクリックします。マウスをプロットボタンの上に置くと、グラフの説明が現われます。
5. 新しい**グラフシート**が開き、グラフが**グラフシート**に描かれます。

ドラッグ アンド ドロップで グラフを作成 する

またグラフシートにグラフの各要素をドラッグ アンド ドロップすることによって、グラフを作成することができます。

ドラッグ アンド ドロップでグラフを作成する



1. プロットするデータを含む、**データ・ウィンドウ**または**オブジェクト・エクスプローラ**を開きます。
2. 新しい**グラフシート**を作成するか、または既存の**グラフシート**を開きます。
3. メインメニューから、**ウィンドウ ▶ 縦に並べる**を選択します。そうすると、データと**グラフシート**が同時に見えるようになります。**グラフシート・ウィンドウ**のタイトルバーをクリックして、ウィンドウを選択します。
4. **標準**ツールバーの**2D プロット**ボタン  か**3D プロット**ボタン  をクリックします。使用可能なプロットボタンのパレットが現われます。
5. パレットから希望のプロットボタンをドラッグし、**グラフシート**にドロップします。デフォルトの軸が描かれ、プロットアイコンがグラフに描かれます。
6. データウィンドウ上でプロットしたいデータ列を選択します。連続していない列を選択するときは、**CTRL**-クリックを使用してください。
7. 選択した領域（列見出しではなく）内で、カーソルが白い左上向き矢印に変化するまでマウスを移動させます。マウスを押したまま、グラフシート上のプロットアイコン上にドラッグします。プロットアイコンの色が変化したらマウスボタンを離してデータをドロップすると、グラフが生成されます。
8. プロットアイコンが消え、実際のプロットが現れます。

グラフの種類 を変更する

グラフを作成した後で、グラフの種類を変更することができます。たとえば、散布図を作成した後で、同じデータで回帰直線をプロットしたい場合を想定します。前述の手順にしたがって新しいプロットを作成することにより、回帰直線をプロットすることができます。新しいプロットを作成する代わりに、既存のプロットの図表示を変更することもできます。

プロットパレットを利用して図表示を変更する

1. 変更したいプロットを選択します。プロットが正確に選択されていると、**x** 軸に最も近いデータ点に緑色のノブが現われます。

- 標準ツールバーの **2D プロットボタン**  か **3D プロットボタン**  をクリックします。使用可能な 2 次元または 3 次元のグラフタイプのパレットが現われます。
- 希望のプロットボタンをクリックします。選択されたプロットが、新しい種類のグラフで再描画されます。適切なプロットボタンをクリックすることによって、違ったグラフに次々と変更することができます。

注意点は、選択するグラフが同じ種類のデータに対応するものでなければならないことだけです。たとえば、散布図から線形回帰プロットへは、両方が同じタイプのデータを必要とするため変更することができますが、2次元散布図から3次元散布図へは、2次元散布図では2列のデータしか必要なく、一方3次元散布図が3列のデータを必要とするため、変更することはできません。データが選択したグラフに適していない場合、プロットはグラフ上にアイコンの形で表示されます。プロットタイプを変更する別の方法として、プロットを選択して、メインメニューから**フォーマット ▶ プロットタイプ**を選択し、変更することができます。

グラフに プロットを 追加する

プロットボタンまたはメニューを使用して、既存のグラフにプロットを追加することができます。

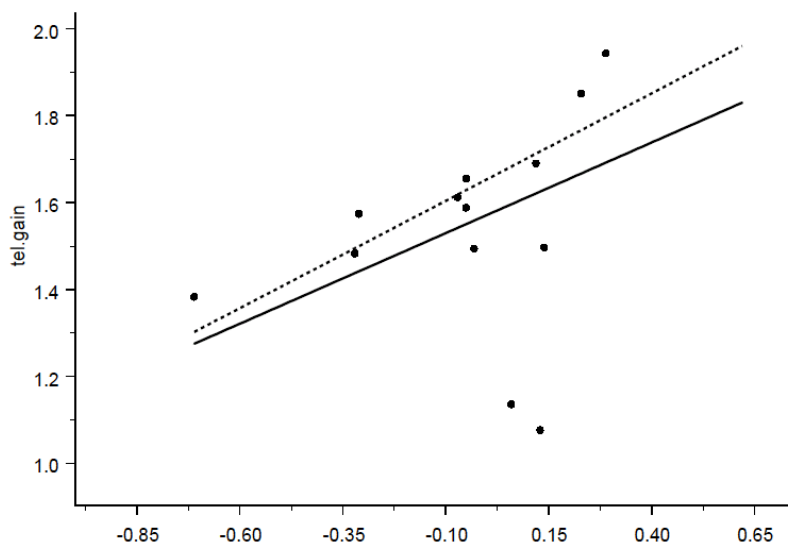




図 6.2 : 単一グラフ上に複数系列をプロット

グラフ上の各プロットは、1 つまたは複数の列のデータを表します。プロットはすべて同じグラフ（たとえば線グラフ）でも、異なるグラフの組み合わせ（たとえば線グラフと散布図と棒グラフ）でもかまいません。

組み合わせられたプロットの軸の「タイプ」は同じでなければなりません。たとえば、線グラフと棒グラフには両方とも **XY** 軸があり、1 つのグラフで組み合わせることができます。しかし、箱型図と鳥瞰図は軸のタイプが異なるため、同じグラフ上で組み合わせることはできません。2 次元グラフと 3 次元グラフは、両方とも同じ**グラフシート**に描くことができますが、同じグラフ上にのりません。

グラフとデータを選択し、プロットボタンを **SHIFT**-クリックすることによって、既存のグラフにプロットを簡単に追加することができます。

プロットボタンを利用してプロットを追加する

1. プロットを追加したいグラフを選択します。
2. プロットするデータを含む**データ・ウィンドウ**を開くか、右パネルに列が表示されるように、**オブジェクト・エクスプローラ**のデータを選択します。
3. メインメニューから、**ウィンドウ ▶ 縦に並べる**を選択します。データと**グラフシート**を同時に見ることができるようになります。
4. プロットしたいデータ列を選択します。連続していない列を選択するときは、**CTRL**-クリックを使用してください。
5. 標準ツールバーの **2D プロットボタン**  か **3D プロットボタン**  をクリックします。使用できるプロットタイプのパレットが現われます。
6. パレットの希望のプロットボタンを **SHIFT**-クリックします。

プロットは、選択されたデータ列を使用して選択されたグラフに追加されます。**SHIFT**-クリックする前にグラフが選択されていない場合は、現在の**グラフシート**に新しいグラフが追加されます。

注意

散布図に別の線または近似曲線を追加するときは、必ず以下の一般的な手順に従ってください。

1. グラフ領域を選択します。
2. データを選択します。
3. 散布図に追加したい線、または曲線のプロットパレット・ボタンを「SHIFT キーを押しながら」クリックします。

**グラフシート
に複数の
グラフを描く**

プロットボタン、メニューまたはドラッグ アンド ドロップを使用して、既存のグラフシートにグラフを追加することができます。プロットは、現在のページに追加することもできますし、新しいページを作成して追加のグラフをそこにおくこともできます。

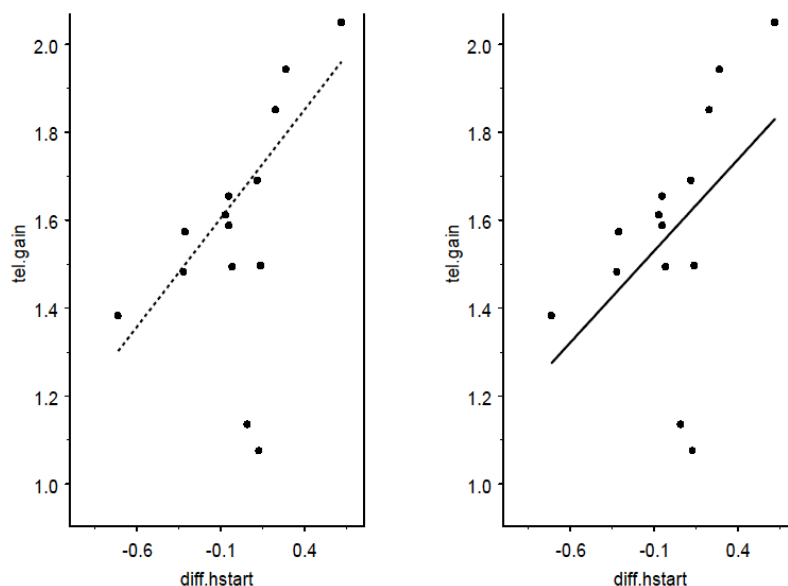




図 6.3 : 1 ページ上に複数のグラフ



プロットボタンの SHIFT-クリックによってグラフを追加する

1. グラフを追加したいグラフシートを開きます。グラフシート上で何も選択されていないことを確かめてください。グラフが選択されていると、新しいグラフが追加されずに選択されているグラフにプロットが追加されます。

2. プロットするデータを含む**データ・ウィンドウ**を開くか、**オブジェクト・エクスプローラ**にデータを表示させます。
3. メインメニューから、**ウィンドウ ▶ 縦に並べる**を選択します。データと**グラフシート**を同時に見ることができるようになります。
4. **データ・ウィンドウ**で、プロットしたいデータ列を選択します。連続していない列を選択するときは、**CTRL**-クリックを使用してください。
5. **標準**ツールバーの**2D プロット**ボタン  か**3D プロット**ボタン  をクリックします。使用できるプロットボタンのパレットが現われます。
6. パレットの希望するプロットボタンを**SHIFT**-クリックします。

グラフが現在の**グラフシート**に追加され、プロットが選択されたデータ列を使用してグラフに描かれます。

ドラッグ アンド ドロップでグラフを追加する

1. プロットするデータを含む**データ・ウィンドウ**を開くか、**オブジェクト・エクスプローラ**でデータを表示させます。
2. グラフを追加したい**グラフシート**を開きます。
3. メインメニューから、**ウィンドウ ▶ 縦に並べる**を選択します。データと**グラフシート**を同時に見ることができるようになります。タイトルバーをクリックし、**グラフシート・ウィンドウ**を選択します。
4. **標準**ツールバーの**2D プロット**ボタン  または**3D プロット**ボタン  をクリックします。使用できるプロットボタンのパレットが現れます。
5. パレットから希望のプロットボタンをドラッグし、**グラフシート**にドロップします。デフォルトの軸が描かれ、グラフ上にプロットアイコンが表示されます。
6. プロットしたいデータ列を選択します。連続していない列を選択するときは、**CTRL**-クリックを使用してください。
7. 選択した列の中で、カーソルが白い左上向き矢印に変わるまでマウスを移動させます。マウスの左ボタンを押したままデータをドラッグし、それをプロットアイコン上に移動させます。プロットアイコンの色が変化したら、マウスボタンを離しデータをドロップするとプロットが生成されます。

グラフシートに新しいページを追加する

1. グラフを追加したい**グラフシート**を開きます。

2. グラフシートの下にある **Page** タブを右クリックし、ショートカットメニューから **Insert Page** を選択します。
3. 前述のように、**SHIFT** クリックまたはドラッグ アンド ドロップして新しいページにプロットを追加します。

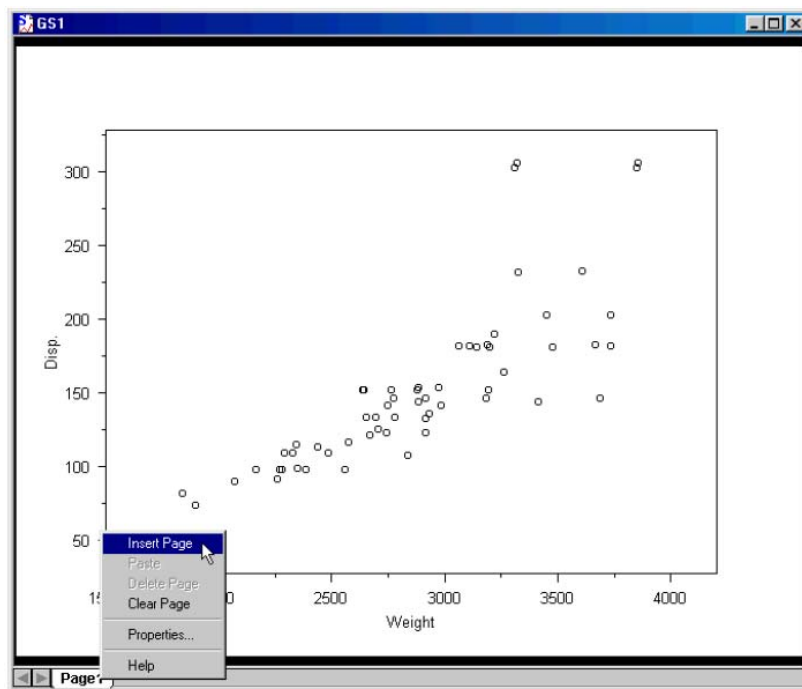


図 6.4 : グラフシートにページを追加する

3次元平面に 2次元グラフ を投映する

2次元グラフのほとんどのタイプは3次元平面に投映することができます。これは、3次元空間に複数の2次元グラフを重ねて、その結果を回転させて見たい場合に便利です。3次元平面に2次元グラフボタンをドラッグ アンド ドロップするか、挿入メニューからグラフを選択して、投映した2次元グラフを作成することができます。

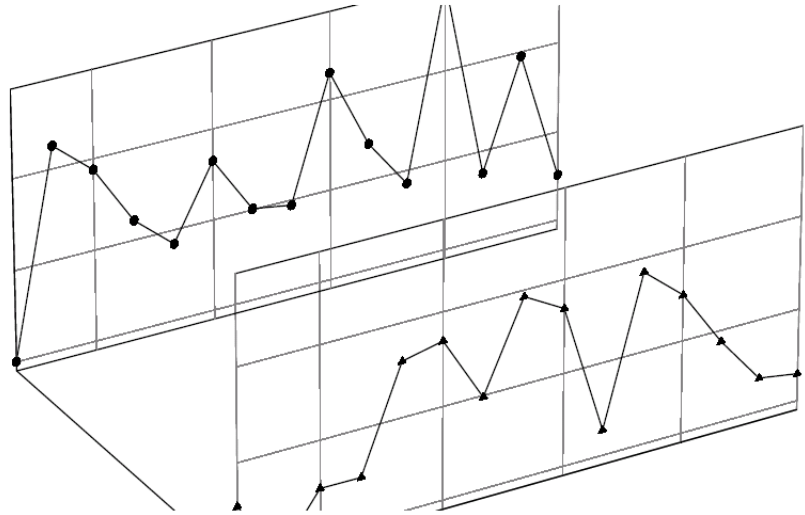




図 6.5 : 3次元空間での複数2次元グラフ

ドラッグ アンド ドロップで2次元グラフを投映する

1. 新しいグラフシートを作成します。
2. 標準ツールバーの**3Dプロット**ボタン  をクリックして、プロットパレットを表示させます。
3. **3Dプロット**パレットの上に、3次元平面が6つあります。



プロットパレットから3次元平面ボタンの1つをドラッグして、それを**グラフシート**にドロップします。3次元グラフが描かれ、1つの面が自動的にグラフに追加されます。面は選択した平面ボタンにより、最小または最大の位置に自動的に位置決めされます。必要に応じて、他の3次元平面をドラッグ アンド ドロップすることができます。

4. 標準ツールバーの**2Dプロット**ボタン  をクリックして、プロットパレットを表示させます。
5. 3次元平面に希望する**2Dプロット**ボタンのひとつをドラッグ アンド ドロップします。プロットを3次元平面にドラッグすると、その平面がハイライトされます（その平面がアクティブなドロップターゲットだからです）。
6. プロットアイコンが3次元平面とリンクされます。プロットアイコン

をダブルクリックしてデータ列を指定したり、データ列をデータから直接ドラッグ アンド ドロップすることができます。

データを指定すると指定した 3 次元平面に 2 次元グラフが描かれます。

Trellis グラフ

はじめに

Trellis グラフは、条件付けによってデータの様々な変数の関係を表わすことができます。第 3 章「データの探索」で、Trellis グラフを作成する例を説明しています。

複数の変数を含むデータセットがあり、2 つの変数のプロットが第 3 の“条件付け”変数の変化によってどのように変化するかを確認したい場合を想定します。Trellis グラフを使用することにより、条件付け変数により異なる区間に分割された元データの各サブセットをそれぞれ一連のパネルに表示させることができます。

たとえばデータセットが、50 のすべての州における 1 年ごとの高校卒業率と、年ごとの 1 世帯当たりの平均所得に関する情報を含んでいるとします。たとえば南部、北部、東部、西部などの米国の異なる地域に関して卒業率と所得の関係をプロットすれば、その関係が地域によって異なるかどうかを判断することができます。この場合、米国の地域が条件付け変数になります。

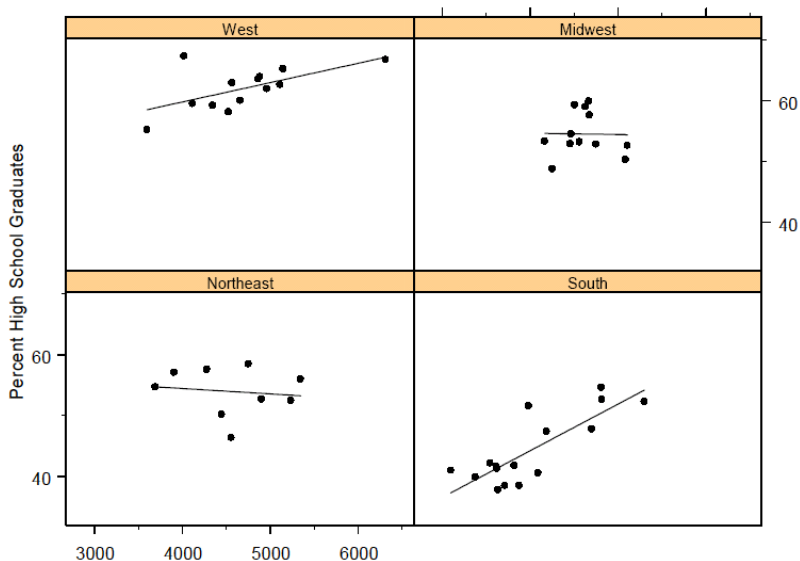


図 6.6 : Trellis グラフ

すべてのグラフは **Trellis** グラフを使用して条件付けすることができます。プロットと条件付け変量に使用されるデータ列は、長さが同じでなければなりません。軸の指定とパネルの表示属性（たとえば塗りつぶしの色）はそれぞれのパネルで同じですが、軸の範囲は異なってもかまいません。パネルの境界と塗りつぶし属性は、**グラフプロパティ**のダイアログにある **Fill/Border** ページで指定することができます。

Trellis グラフを作成する

ドラッグ アンド ドロップで Trellis グラフを作成する

1. グラフを作成します。

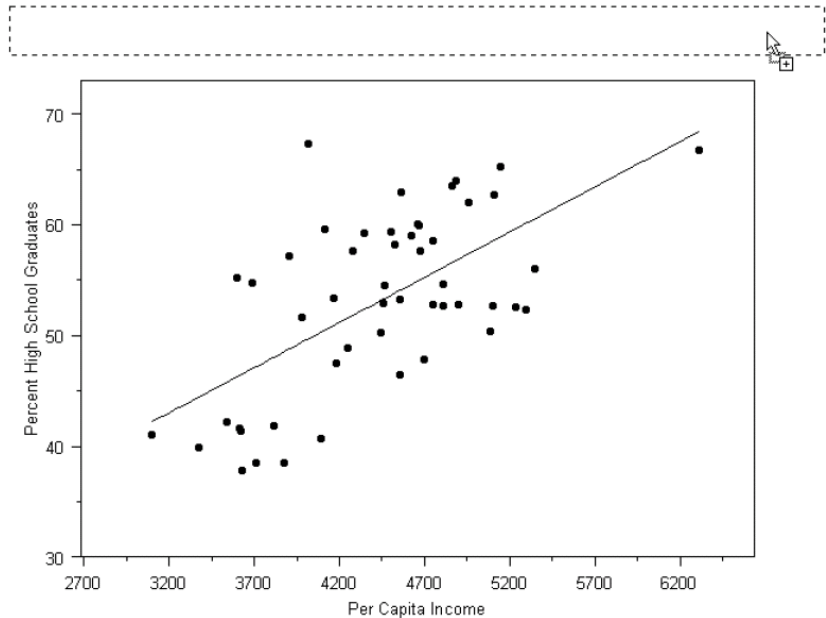


図 6.7 : 条件付けデータをドラッグ アンド ドロップする

2. 条件付けデータの列をグラフにドラッグし、グラフの一番上にあるハイライトされた四角形にドロップします（ハイライトした四角形は図 6.7 に示しています）。
3. 条件付けデータが連続値の場合は、グラフツールにある条件付けボタンを使ってパネルの数を変更することができます。

SHIFT-クリックを使用して Trellis グラフを作成する

1. グラフを作成し、グラフ領域を選択します。

2. データ・ウィンドウまたはオブジェクト・エクスプローラで条件付けするデータ列を選択します。
3. グラフツールパレットの条件付けボタンのどれか 1 つを **SHIFT**-クリックします。

Trellis グラフ上のプロット

Trellis グラフのプロットの操作は、標準のグラフ外での操作ととてもよく似ています。次の操作を行うことができます。


1. データ指定やその他の属性を変更するときは、ダブルクリックまたは右クリックしてください。プロットの指定が変わると、すべてのパネルで修正されます。
2. プロットを選択しプロットパレットをクリックして、グラフの種類を変更します。
3. その上に新しいデータをドラッグします。
4. 別のプロットを追加します。

デフォルトでは、それぞれのプロットはグラフのマルチパネルページで指定された条件付け変数を使用して、それぞれのパネルに使用するデータセットの行を決定します。これはグラフ上のすべての点と同じデータセットからのデータ上にある時、いいかえるとデータ列の長さがすべて等しい時に適しています。

Trellis プロットからパネルを抽出する


任意の Trellis グラフから 1 つのパネルを抽出することができます。たとえばそれぞれのパネルが多くの特徴を示しており、パネルを 1 つずつ慎重に調べたい場合に便利です。

Trellis プロットからパネルを抽出する

1. グラフツールパレットで、パネルの指定ボタンをクリックします。
2. 抽出したいパネル内の任意の場所をクリックします。

グラフの条件付けがオフになり、プロットの **Subset Rows with** 式がそのパネルの条件付けの式に設定されます。また、メインメニューから **書式 ▶ パネルの指定 ▶ 既存グラフシートに指定パネルのみ表示** を選択することによって、抽出したグラフのみを表示させることができます。別々のグラフシートにパネルを表示させるときは、メインメニューから **書式 ▶ パネルの指定 ▶ 新規グラフシートに指定パネルのみ表示** を選択します。

Trellis プロットを元に戻す

- グラフツールパレットで、すべてのパネルを表示ボタン  をクリックします。

あるいは

- メインメニューから、書式 ▶ すべてのパネルを表示を選択します。

グラフのフォーマット

S-PLUS は、データに則したデフォルト値を使ってすべてのプロットを生成します。たとえば、 x 軸と y 軸の長さはデータの最小値と最大値によって決まります。軸ラベルは、列名または `DataDescription` (もしあれば) が使われます。色、線のタイプ、プロットシンボルなどにはすべて、様々なデータで使うことができる適当なデフォルト値があります。S-PLUS はプロットを生成した後で、グラフのあらゆる要素を変更し、有用にカスタマイズすることができます。

この強力な機能は、“グラフのフォーマット” と呼ばれます。たとえば図 6.8 は、デフォルトで出力される回帰直線プロットです。

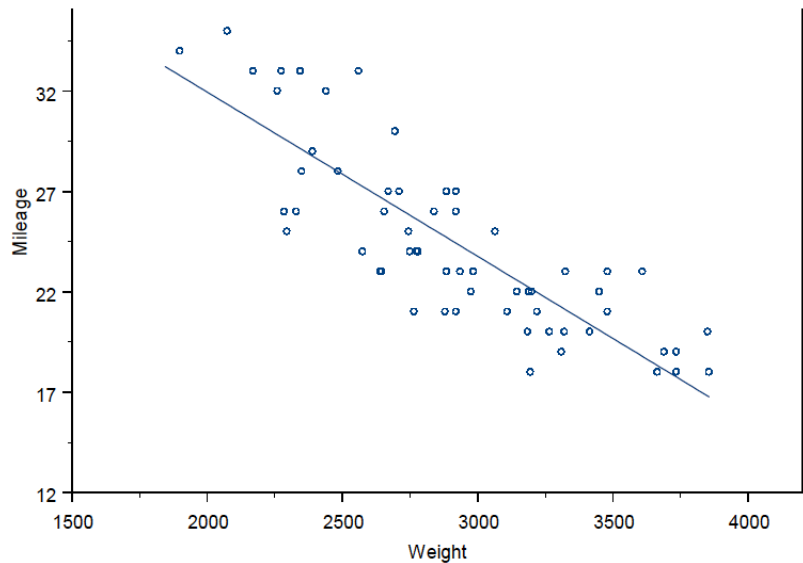


図 6.8 : フォーマットする前の回帰直線

タイトルを追加し、軸目盛りを変更し、いくつかグラフ要素を追加することによってプロットをフォーマットした後、プロットは図 6.9 に変換されます。この節ではフォーマット機能について説明し、図 6.8 がどのように図 6.9 に変換されたかを説明します。詳細は、オンラインヘルプを参照してください。

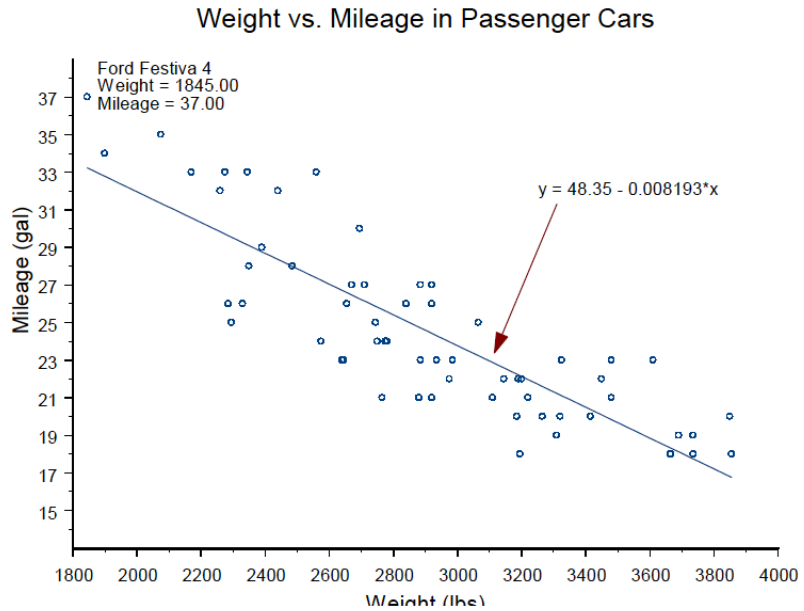


図 6.9 : フォーマット後の図 6.8

グラフの オブジェクト 指向的性質

すべてのグラフは、個別に編集することができるオブジェクトから構成されています。たとえば線グラフや散布図には、プロットオブジェクト（データをプロットするオブジェクト）、2つの軸オブジェクト（軸を定義して描画するオブジェクト）、2つの軸ラベルオブジェクト（軸ラベルを表示するオブジェクト）、タイトルオブジェクト（グラフのタイトルを表示するオブジェクト）、全体のレイアウトを定義する2次元グラフオブジェクトおよび2次元グラフのマルチパネルオプションがあります。これらの各オブジェクトは固有のプロパティダイアログがあり、編集することができます。これにより、グラフのあらゆる要素を実際にカスタマイズすることができます。グラフ内のすべての構成要素がオブジェクトであるため、他のオブジェクトを追加する（または削除する）こともできます。たとえばグラフに他の軸オブジェクトを追加することにより、プロットに第3の軸を加えることができます。

グラフを フォーマット する : 例


図 6.8 から図 6.9 への変換は、次のステップを使用しました。

プロットを生成する

最初に、フォーマットするプロットを生成します。

1. メインメニューから、**データ ▶ データの選択**を選択して **Select Data** ダイアログを開きます。 **Existing Data** グループの **Name** フィールドに **fuel.frame** と入力し、**OK** をクリックします。

データ・ウィンドウに `fuel.frame` データセットが現われます。データには、重量（ポンド）、エンジン排気量、燃費、燃料値および車種がそれぞれ異なる 60 台の車があります。ここでは、この 60 台の車の重量と燃費の関係をプロットし、車の重量が燃費に影響を及ぼすのか、またそれはどの程度か、という疑問の答えを導くプロットを生成したいと思います。

2. 2つのデータ列 `Weight` および `Mileage` を選択します。x軸を `Weight` にしたいので最初に `Weight` 列見出しをクリックし、次に `Mileage` を **CTRL**-クリックします。これで2つのデータ列がハイライトされるはずですが。
3. **2Dプロット**パレットを開き、**回帰直線** ボタン  を選択します (137 ページの「曲線あてはめプロット」の節を参照)。図 6.8 とよく似たプロットが表示されるはずですが。プロットがかなり異なって見える場合は、2次元グラフのデフォルトのプロパティが変更されている可能性があります。

軸タイトルをフォーマットする

y軸タイトルのテキストを変更しましょう。また、軸タイトルのサイズを大きくして、もう少し目立つようにしたいと思います。

4. y軸のタイトルの **Mileage** を2回クリックします。**@Auto** ラベルを **Mileage (gal)** に変更します。編集ボックスの外側をクリックします。
5. テキストを選択した状態で、ツールバーを使ってフォントサイズを **20** に大きくします。

これで y 軸タイトルはフォーマットされました。同様に x 軸タイトルをフォーマットし、タイトルを **Weight (lbs)** に、フォントを **20** に変更します。

軸をフォーマットする

引き続き軸ラベルをフォーマットします。軸のそれぞれに目盛りラベルの調整、細かい目盛りの追加、軸線の幅の拡張、プロットの右と上の枠線の除去など4つのフォーマット変更を行いたいと思います。

6. x 軸線をクリックして選択します。軸の中心に緑色の四角が現われるはずですが。もしその四角の代わりに緑の三角形が現われた場合は、軸ラベルが選択されています。x 軸線をダブルクリックすると **X Axis** ダイアログが開きます（あるいは軸を選択した後、**CTRL-1** を入力してください）。

7. 軸の幅を太くするために、**Display/Scale** ページの **Axis Display** グループ内の **Line Weight** には **1** を選択します。**Apply** をクリックすると変更が表示されます。

この変更がどのように見えるかを確認したいときは、**Apply** をクリックしてください。**Apply** をクリックしないと **OK** をクリックしてダイアログを閉じるまで、その変更はグラフに表示されません。このステップには注意してください。

8. 上と右の枠線をのぞくために、**Display/Scale** ページの **Options** グループの **Frame** に **None** を選択します。
9. **Range** タブをクリックしてダイアログの **Range** ページを表示します。**Axis Range** グループで、**Axis Minimum** に **1800** を、**Axis Maximum** に **4000** を入力します。
10. 目盛りが軸の始めから始まるようにしたいと思います。**First Tick** に **1800** と、**Last Tick** に **4000** と **Tick Range** を設定します。
11. 大きい目盛りが 200 ポンドごとになるようにします。**Major Tick Placement** の **Interval** として **200** を入力します。**Interval Type** を **Size** に設定します。
12. **Grids/Ticks** タブをクリックしてダイアログの **Grids/Ticks** ページを表示します。**Minor Ticks** グループで、**Length** に **0.05** を入力し、**OK** をクリックしてダイアログを終了します。

これで x 軸のフォーマットが完了です。ステップ 8~12 を繰り返し、 y 軸を同じようにフォーマットします。 y 軸の範囲は現在のままにしてください。すなわち y 軸をフォーマットするときに、ステップ 9 と 10 をスキップしてください。**Major Tick Placement** の **Interval** として **2** を入力し、**Interval Type** を **Size** に設定します。

グラフにタイトルを追加する

タイトルをフォーマットするときは、まずグラフにタイトルオブジェクトを挿入してください。

13. **グラフシート** にフォーカスが合っていることを確認し、メインメニューから **挿入 ▶ タイトル ▶ メインタイトル** を選択します。編集ボックスが現れます。**@Auto** を **Weight vs. Mileage in Passenger Cars** と変更します。これを行うときは RETURN を押さずに、タイトル領域の外側をクリックしてください。タイトルが自動的にセンタリングされます。
14. タイトルのフォントサイズを大きくするときは、タイトルをクリックしてタイトルオブジェクトを選択します。タイトルオブジェクトが選択されると、緑色のノブがタイトルを囲みます。ツールバーを使って

フォントのサイズを **24** に変更します。

一般にオブジェクトをダブルクリックすると、そのオブジェクトのプロパティダイアログが現れます。テキストオブジェクト（軸、タイトルまたはその他のテキストオブジェクト）の場合は、オブジェクトをダブルクリックするとオブジェクトがその場で編集されます。テキストオブジェクトのプロパティダイアログを開くときは、他のすべてのオブジェクトの場合のようにダブルクリックするかわりに **CTRL-1** を入力してください。



回帰式を追加する

プロットに回帰直線の式を追加します。**S-PLUS** では、これを簡単に行うことができます。

15. まず回帰直線か任意のデータ点をクリックして、線形回帰プロットを選択します。プロットの下に緑色の四角が現れ、そのデータオブジェクトが選択されたことを示します。
16. メインメニューから **挿入 ▶ 曲線近似式** を選択します。プロット上に回帰直線の式が現れます。テキストをダブルクリックし、**HOME** を押し、**y=** と入力して、式の前に **y=** を加えます。ツールバーを使ってフォントサイズを **16** に変更します。最後にテキストオブジェクトを選択し、オブジェクト（緑色の四角で囲まれた）の内側を左クリックしたあと希望の位置にドラッグして移動させます。

グラフに注釈をつける

グラフをフォーマットする最後のステップは、そのグラフにいくつかの注釈を加えることです。回帰式から線に向かって矢印を描き、どの車が最も燃料効率が高いかを示します。

17. **グラフ** ツールバーの **図形描画** ボタン  をクリックして、**図形描画パレット** を開きます。
18. まず、最も燃料効率の良い車にラベルを付けます。**図形描画** パレットの **Label Point** ボタン  を選択します。カーソルが大きいプラス記号に変わります（図 6.10 を参照）。

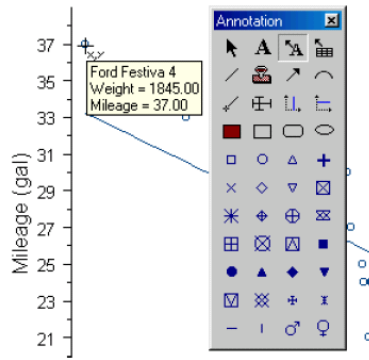


図 6.10 : 図形描画パレットの Label Point ツールを使用する

カーソルを y 軸の一番上の点まで移動させてクリックします。Ford Festiva 4 というラベルが、その車の重量および走行距離と一緒に現れます。

19. 図形描画パレットの Select Tool ボタン  をクリックし、ラベルをクリックして選択し、ツールバーを使ってラベルのフォントサイズを 16 に変更します。必要に応じてラベルの位置を変更します。
20. 最後のステップは、曲線近似式から線まで矢印を描くことです。図形描画パレットの Arrow Tool ボタン  をクリックすると、カーソルが矢印の付いたプラス符号に変わります (図 6.11 を参照)。

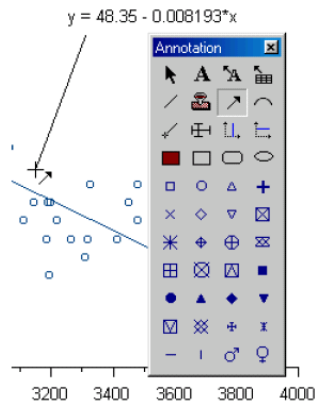


図 6.11 : 図形描画パレットの Arrow Tool を使用する

カーソルを曲線近似式のマイナス符号のすぐ下に移動させます。次にマウスをその位置（矢印の始まり）から回帰直線のすぐ上までドラッグします。マウスボタンを放すと矢印が描かれます。

これでフォーマットは完了しました。画面のグラフは、図 6.9 のようなものになったはずですが、この章の後半で、このグラフを引き続きフォーマットします。234 ページの「グラフをフォーマットする（続き）」の節を参照してください。

グラフシートをフォーマットする

グラフシートは、ページ全体のプロット環境を定義するオブジェクトです。グラフシートの名前、プロットの向き（縦長または横長）、「使われる単位」（デフォルトはインチ）、グラフシートのサイズ（デフォルトは、11×8.5 インチ）、ページ余白およびデフォルト色を設定することができます。さらに上級ユーザなら、1 つのグラフシートに複数のグラフをどのようにプロットするかを指定することができます。

グラフシートをフォーマットする

1. グラフシートに焦点を合わせている状態で、メインメニューから **書式** ▶ **グラフシート** を選択します。**Graph Sheet** ダイアログが現れます。
2. グラフシートに希望するフォーマットの変更を加え、**OK** をクリックします。

グラフシートのデフォルトを保存する

グラフシートの設定を保存して、新しいグラフシートを作成するときその設定をデフォルトとして利用することができます。これは、複数のグラフシートを同様のプロパティで作成するときに便利です。

現在アクティブなグラフシートの設定をデフォルトとして保存する

1. グラフシートを開くか、希望のフォーマットがされたグラフシートを作成します。
2. グラフシートのグラフ以外の空白領域をクリックします。
3. メインメニューから、**オプション** ▶ **ウィンドウサイズ・プロパティ** の今の状態をデフォルト値にするを選択します。

現在のグラフシートの設定（ウィンドウサイズと位置を含む）が、デフォルト設定値として保存されます。またグラフシートの端を右クリックして、**Save Graph Sheet as Default** を選択することもできます。

グラフをフォーマットする

グラフには、グラフ領域とプロット領域の外側の境界を定義するプロパティがあります。グラフ/プロットをサイズ変更/移動したい場合は、対話式かプロパティダイアログを使用してグラフをフォーマットすることができます。

グラフ領域をフォーマットする

グラフ領域を選択して、そのサイズまたはフォーマットを変更することができます。

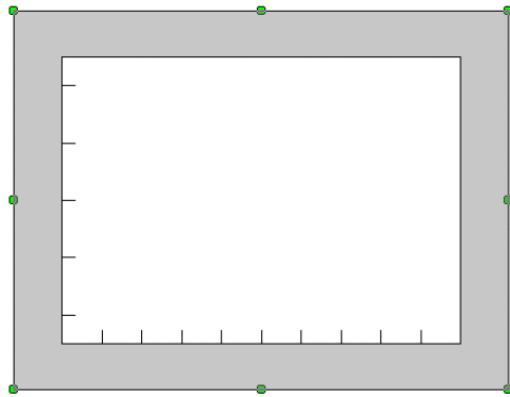


図 6.12 : グラフ領域（灰色）が選択されている

グラフ領域を選択する

- グラフ境界の内側で、かつ軸の外側の部分をクリックします。選択されると、グラフ領域のすべての辺とすべての4つの角に緑色の四角が現れます。

グラフ領域のサイズを変更する

- グラフ領域を選択し、角にあるサイズ変更ノブをドラッグして希望のサイズにします。

または

- グラフ領域内の他のオブジェクト以外の部分をダブルクリックして、**グラフプロパティ**のダイアログを表示させ、**Position/Size** タブをクリックします。**Graph Size** グループで、**Width** と **Height** を指定し **OK** をクリックします。

または

- グラフ領域を選択して右クリックし、ショートカットメニューから **Position/Size** を選択します。**Graph Size** グループで、**Width** と **Height** を指定し **OK** をクリックします。

グラフ領域を移動させる

- グラフ領域を選択し、そのグラフ領域を新しい位置にドラッグします。

または

- グラフ領域の内側をダブルクリックして、**グラフプロパティ**のダイアログを表示させます。**Position/Size** タブをクリックします。**Graph Position** グループでは、**Horizontal** と **Vertical** の位置を指定し **OK** をクリックします。

グラフ領域をフォーマットする

- グラフ領域の内側をダブルクリックして、**グラフプロパティ**のダイアログを表示させます。**Fill/Border** タブをクリックします。グラフ領域に希望するフォーマット変更を加えて、**OK** をクリックします。

プロット領域をフォーマットする

プロット領域を選択してそのサイズまたはフォーマットを変更することができます。2次元グラフでは、プロット領域は軸で境界が決められます。

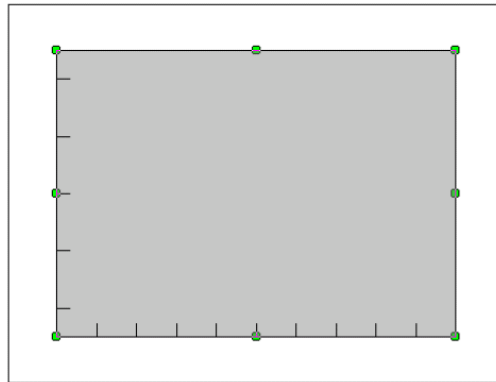


図 6.13 : プロット領域（灰色）が選択されている

プロット領域を選択する

- 軸で境界が決められた領域内の任意の場所をクリックします。選択されると、プロット領域のすべての辺と 4 つのすべての角に緑色の四角が現れます。

プロット領域のサイズを変更する

- プロット領域を選択し、角にあるサイズ変更ノブをドラッグして希望のサイズにします。

または

- グラフの内側をダブルクリックして**グラフプロパティのダイアログ**を表示させます。**Position/Size** タブをクリックします。**Plot Display Size** グループでは **Width** と **Height** を指定し、**OK** をクリックします。

または

- プロット領域を選択して右クリックし、ショートカットメニューから **Position/Size** を選択します。**Plot Display Size** グループでは **Width** および **Height** を指定し、**OK** をクリックします。

プロット領域を移動させる

- プロット領域を選択し、それを新しい位置までドラッグします。

または

- グラフの内側をダブルクリックして**グラフプロパティのダイアログ**を表示させます。**Position/Size** タブをクリックします。**Plot Origin Position** グループでは **X** と **Y** の値を指定し、**OK** をクリックします。

プロット領域をフォーマットする

- グラフの内側をダブルクリックして**グラフプロパティのダイアログ**を表示させます。**Fill/Border** タブをクリックします。プロット領域の境界線と塗りつぶしに希望するフォーマット変更を加え、**OK** をクリックします。

2次元軸をフォーマットする

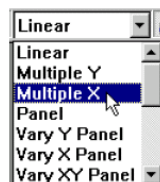
軸を様々な方法でカスタマイズすることができます。線形軸、対数軸、確率軸をいくつかの組み合わせで選択することができます。別の軸をプロットに追加したり、軸を削除および移動したり、軸プロパティダイアログを使用してこれらを完全にカスタマイズすることができます。**S-PLUS** は軸と軸ラベルを区別しており、それぞれに別々のプロパティダイアログがあることに注意してください。

軸タイプを選択する


デフォルトの2次元軸タイプを指定したり、個々の軸のフォーマットの一部として軸タイプを指定したりすることができます。





デフォルトの2次元軸タイプを指定する


- 標準ツールバーのドロップダウンメニューから希望する2次元軸タイプを選択します。



軸を対話的に 尺度変更する

指定矩形内を拡大ボタン  を使用して、プロットの指定部分を拡大することができます。2次元グラフの焦点を合わせたい領域のまわりに四角形をドラッグします。グラフのその領域だけを示すように、X軸とY軸が尺度変更されます。またこのツールは、メインメニューから書式 ▶ 指定矩形内を拡大を選択してアクセスすることもできます。

プロットの他の部分を表示させるときは、グラフツールパレットのパンボタン     を使用してください。

プロット全体を復元するとき、グラフツールボタンの元のスケールに戻す 、またはメインメニューから書式 ▶ 元のスケールに戻すを選択してください。

軸の選択

軸を選択する

- 軸の目盛り部分の内側をクリックします。軸が選択されると、軸の中心に正方形の緑色のノブが現れます。三角形の緑色のノブが表示されている場合は、軸ではなく軸ラベルが選択されています。

軸をフォーマットする

- 軸をダブルクリックするか、軸を選択してメインメニューから書式 ▶ 選択された Axis を選択します。両方の軸に同じ変更を加える場合は、CTRL-クリックで両方を選択し、書式 ▶ 選択したオブジェクトを選択してください。
- 軸ダイアログから、Display/Scale、Range、Grids/Ticks または Axis Breaks の希望のページを選択します。変更を加え OK をクリックします。

または

- 軸を選択、右クリックし、ショートカットメニューから軸プロパティメニューの1つを選択することによって、プロパティダイアログのページを開きます。

軸のプロパティダイアログで、軸の表示特性を変更することができます。範囲と目盛りの間隔を変更し、大きい目盛りと小さい目盛りの細かい表示を指定することもできます。上級ユーザなら、軸に切れ目を入れること（軸を一部省略すること）ができます。

軸を移動してプロット領域から外す

- 軸を選択します。軸の中心に四角のノブが現われます。軸が希望の位置になるまでノブをドラッグします。

他の軸を追加する

2次元グラフに軸を簡単に追加することができます。最も簡単な方法は、標準ツールバーから **Multiple X** または **Multiple Y** の2次元軸タイプを選択することにより、2つ目のX軸またはY軸を追加することができます。

また、次のようにドラッグ アンド ドロップまたはメニューオプションを使用して軸を追加することもできます。

- **グラフツールパレット**から付加軸の1つをドラッグし、それをグラフ領域の内側にドロップします。付加軸が選択したグラフに追加されます。軸が既にその位置にある場合、新しい軸は自動的に元の軸から少し外されます。



図 6.14 : グラフツールパレットの付加軸ボタン

または

- グラフを選択し、メインメニューから**挿入 ▶ 軸**を選択します。

同じ方法で枠線付きの軸を追加することができます。枠線は本当の軸ではなく反対側の軸を対称に写したもので、目盛りは常に同じになります。

2次元軸 ラベルを フォーマット する

軸ラベルは軸と同じようにフォーマットされます。フォーマットのオプションには、ラベルのタイプ (**Decimal**、**Scientific**、**Percent**、**Currency**、**Months**、**Date** など)、目盛りラベルの正確な位置、そしてラベルのフォントと色があります。軸タイトルは、軸の指定には含まれないことに注意してください。これは別々にフォーマットすることができるテキストオブジェクトです。

(228 ページを参照)

2次元軸ラベルを選択する

- ラベルの任意の場所をクリックしてそのラベルを選択します。三角形の選択ノブが現れます。四角い緑色のノブは、軸ラベルではなく軸が選択されていることを示しています。

2次元軸ラベルをフォーマットする

- ラベルをダブルクリックするか、ラベルを選択し、メインメニューから書式 ▶ 選択された X Axis Label を選択します。Axis Label ダイアログから、Label 1、Label 2/Minor Labels/Position、Font のうちの希望ページを選択します。必要な変更を加え、OK をクリックします。

2次元軸ラベルを移動する

- 軸ラベルを選択します。三角形の選択ノブをドラッグして、ラベルを軸の内側または外側にドラッグします。

または

- ラベルをダブルクリックするか、またはラベルを選択してメインメニューから書式 ▶ 選択された X Axis Label を選択します。Position タブを選択し Horizontal と Vertical フィールドに値を指定し、OK をクリックします。

複数行テキストを追加しフォーマットする

テキスト（コメント）、メインタイトルまたはサブタイトル、軸タイトルそして日付および時刻スタンプの形で、グラフに複数行テキストをいくつでも追加することができます。これらの様々なテキストタイプのプロパティダイアログのフォーマットオプションは、基本的に同じです。

複数行テキストを追加する

- メインメニューから挿入 ▶ テキストを選択します。テキスト編集ボックスが開きます。（挿入メニューから、他のタイプのテキスト、たとえばタイトルを利用することもできます。）

または

- **図形描画**パレットを開き、**コメントツール** **A** ボタンをクリックします。カーソルが**コメントツール**に変わります。**グラフシート**上でクリックし、カーソルをドラッグし、マウスボタンを離します。デフォルトのテキストがグラフに追加されます。選択したテキストをクリックすると、編集ボックスが開きます。

表示させたい文字をタイプ入力してください。**ENTER** を押すと新しい行が作成されます。編集を終了してその結果を保存するときは、編集ボックスの外側をクリックしてください。代わりに、**F10** または **CTRL-ENTER** を押すこともできます。保存せずに終了するときは、**ESC** キーを押してください。

既存のテキストをその場で編集する

- テキストを右クリックし、メニューから **Edit In-place** を選択します。
または
- テキストをクリックしてそのテキストを選択し、もう1度クリックします。

編集ボックス内の文字に変更を加えます。**ENTER** を押すと新しい行が作成されます。編集を終了してその結果を保存するときは、編集ボックスの外側をクリックします。代わりに、**F10** または **CTRL-ENTER** を押すこともできます。保存せずに終了するときは、**ESC** キーを押してください。

テキストを移動する

- テキストを選択します。テキストボックスのまわりに緑色の四角い選択ノブが現われます。選択された部分の「内部」をクリックし、そのボックスを新しい位置にドラッグします。

また、プロパティダイアログの **Position** ページの **X** と **Y** の位置を変更することによってテキストを移動させることもできます。

テキストのサイズ変更をする

- テキストを選択します。テキストボックスのまわりに緑色の選択ノブが現われます。「正方形」の緑色のノブの1つをドラッグして、ボックスのサイズを大きくしたり小さくしたりします。テキストの比率(高さ と 幅の比率) は変化しません。

あるいは、ツールバーボタンを使用してフォントサイズを変更します。

プロパティダイアログを使用してテキストをフォーマットする

- テキストを選択して **CTRL-1** を押すか、メインメニューから **書式 ▶ 選択された Comment** を選択します。**Comment** ダイアログが現れます。**Comment** ダイアログでフォーマットを変更し、**OK** をクリックします。

テキストをその場でフォーマットする

- テキスト編集ボックスを開き、テキストを選択します。**グラフツール** バーのオプション (**フォントボタン**、**フォントサイズボタン**、**太字ボタン**、**下線ボタン**、**斜体ボタン**、**上付きボタン**および**下付きボタン**を使用) を選択して、テキストのフォーマットを変更します。フォントとポイントサイズを変更し、選択したテキストをボールドまたはイタリックにしたり、下線を引いたりすることができます。

または

- テキスト自体を右クリックしてショートカットメニューを表示させます。テキストをフォーマットするときは **Superscript** または **Subscript** を選択すると、上付き文字、下付き文字の入力が可能です。フォントタイプを編集するダイアログを開くときは **Font** を選択します。記号またはギリシャ文字を追加または編集するダイアログを開くときは **Symbol** を選択します。

テキストを削除する

- テキストを選択します。**DELETE** キーを押します。または、**編集** メニューから **クリア** を選択するか、右クリックして **Cut** を選択することもできます。

タイトルと凡例を追加する

タイトルは自動的に位置決めされるため、タイトルの挿入は通常のテキスト挿入と異なります。タイトルの編集とフォーマットについては、224 ページを参照してください。

メインタイトルまたはサブタイトルを追加する

- メインメニューから、**挿入 ▶ タイトル** を選択してから、**メインタイトル** または **サブタイトル** を選択します。テキストを入力するための編集ボックスが開きます。必要なテキストをタイプ入力します。**ENTER** を押すと新しい行が作成されます。編集を終わって結果を保存するときは、編集ボックスの外側をクリックします。代わりに、**F10** または **CTRL-ENTER** を押すこともできます。保存せずに終了するときは、**ESC** キーを押してください。

2次元軸タイトルを追加する

グラフに軸タイトルを入れることができます。軸タイトルは、自動的に位置決めされるので便利です。軸タイトルの編集とフォーマットについては、224ページを参照してください。

2次元軸タイトルを追加する

- タイトルを追加したい軸を選択します。メインメニューから**挿入 ▶ タイトル ▶ 軸**を選択します。テキストを入力する編集ボックスが開きます。必要なテキストをタイプ入力します。新しい行を作成するときは **ENTER** を押してください。編集を終了して結果を保存するときは、編集ボックスの外側をクリックします。代わりに、**F10** または **CTRL-ENTER** を押すこともできます。保存せずに終了するときは、**ESC** キーを押してください。

3次元軸タイトルを追加する

3次元軸タイトルは、対話式に移動したりサイズを決めたりすることができないという点で、2次元軸タイトルと異なります。3次元軸タイトルのテキストは **3D Axes** ダイアログで指定され、複数の行にすることはできません。ただし、3次元グラフに複数行テキストをコメントやタイトルの形で追加することができます。テキストの指定についての詳細は、225ページを参照してください。

3次元軸タイトルを追加する

- 軸をダブルクリックするか、軸を選択しメインメニューから**書式 ▶ 選択された 3D Axes** を選択します。**3D Axes** ダイアログから **X Text**、**Y Text** または **Z Text** タブを選択します。**Text**、**Font**、**Size** および **Color** フィールドに必要な変更を加え、**OK** をクリックします。


または

- メインメニューから**挿入 ▶ タイトル ▶ 軸**を選択します。編集用の **3D Axes** ダイアログが現われます。

日付と時刻スタンプを追加する

日付と時刻スタンプで、指定したテキストと共に日付と時刻を表示させることができます。既存のテキストの編集とフォーマットについては、226ページを参照してください。

日付スタンプを追加する

- **図形描画** パレットの日付、時間ボタン  をクリックします。カーソルが**日付スタンプツール**に変わります。**グラフシート**上をクリックし、テキストボックスが希望のサイズになるまでドラッグします。マウスボタンを離すと、デフォルトの日付スタンプがグラフに表示されます。文字がテキストボックスの境界線外を越えるほど長くなると、テキストボックスが自動的に大きくなります。


または

- メインメニューから**挿入 ▶ 図形描画 ▶ 日付、時間**を選択します。テキストを入力する編集ボックスが開きます。必要なテキストをタイプ入力します。新しい行を作成するときは**ENTER**を押してください。編集を終了し結果を保存するときは、編集ボックスの外側をクリックします。代わりに、**F10**または**CTRL-ENTER**を押すこともできます。保存せずに終了するときは、**ESC** キーを押してください。

凡例を追加する

凡例は、グラフ上で様々なプロットを説明するテキストと図形を組み合わせたものです。マルチパネルグラフでも、凡例は1つのみです。

凡例を追加する

- グラフツールバーの**凡例の自動作成**ボタン  をクリックするか、メインメニューから**挿入 ▶ 凡例**を選択します。グラフシートに複数のグラフがある場合は、希望するグラフを選択した後で**挿入**メニューから**凡例**を選択してください。

凡例を削除するときは、**凡例の自動作成**ボタン  をもう1度クリックします。


凡例ボックスをフォーマットする

- 凡例を囲む境界線をダブルクリックするか、凡例を選択してメインメニューから**書式 ▶ 選択された Legend**を選択します。**Legend** ダイアログで、凡例ボックスの凡例の位置とフォーマットを指定することができます。あるいは凡例を右クリックし、ショートカットメニューから**Legend** ダイアログのページを選択することができます。

点にラベルを追加する

グラフ上に2次元散布図がある場合は、選択した点のラベル（行名などに決定されている）を自動的に表示することができます。


点にラベル付けする

- グラフシート上に複数の散布図がある場合は、使用したい散布図を選択します。グラフツールパレットの**行名の表示**ボタン  をクリックします。散布図のデータ点をクリックします。ラベルが現れます。このラベルは、他のコメントと同じように移動したり編集したりすることができます。
- 違う点のラベルを表示させるには、別のデータ点をクリックします。最初のラベルが消え、新しく選択したデータ点に新しいラベルが現れます。

- 前の点のラベルを残したまま、別の点にラベルを追加するときは、別のデータ点を **SHIFT** クリックします。別のラベルが現れます。

データ・ウィンドウ内の点を識別する

グラフ上に 2 次元散布図がある場合は散布図内の点をクリックすることにより、データ・ウィンドウ内の該当する点のデータ行を選択することができます。

- グラフ上に複数の散布図がある場合は、表示させたい散布図を選択します。散布図で x 列と y 列に使用されているデータのウィンドウを開きます。メインメニューから **ウィンドウ ▶ 縦に並べる** を選択し、データとグラフを並べて表示させます。グラフツールパレットの **矩形内のデータ選択** ボタン  をクリックします。散布図のデータ点をクリックします。データ・ウィンドウ内の対応する点のデータ行が選択されます。
- 違う行を選択するときは、散布図内の別の点をクリックします。
- さらに別の行を選択に加えるときは、別のデータ点を **SHIFT** クリックします。
- 点のグループを選択するときは左マウスボタンを押し、選択する点のまわりに四角形をドラッグします。

近似式を追加する

グラフ上に近似プロットがある場合は、線の式を自動的に表示することができます。

近似式を挿入する

- 式を表示させたい線を選択します。メインメニューから **挿入 ▶ 曲線近似式** を選択します。線の式が、グラフ上に表示されます。

近似式の表示桁数を指定する

- 式を表示させたい線を右クリックし、ショートカットメニューから **Results** を選択します。**Curve Fitting Plot** ダイアログの **Results** ページが表示されます。**Equation** グループで、**Precision** フィールドで近似式に用いる表示桁数を指定します（有効範囲は 1~15）。**OK** をクリックします。メインメニューから、**挿入 ▶ 曲線近似式** を選択します。線の式が、指定した表示桁数でグラフに表示されます。

注意

近似式をグラフに追加した後は、式の表示桁数を変更することはできません。

既存の近似式を編集する

- 式をダブルクリックするか、式を選択してメインメニューから**書式 ▶ 選択された Comment**を選択します。式のフォーマットについての詳細は、224 ページを参照してください。

近似式オプションは、グラフ上に少なくとも 1 つの近似数があるときだけ使用することができます。複数の近似線がある場合は、それぞれを選択して線を示す式を自動的に表示させることができます（種類によっては、ノンパラメトリックなので式が表示されません）。

線、形、記号を追加する

グラフにテキスト、線、形、記号などの項目を追加することもできます。このような描画オブジェクトは、**図形描画**パレットを使って追加することができます。

図形描画パレットからオブジェクトを追加する

- **図形描画**パレットから描画オブジェクトアイコンをグラフ上にドラッグ アンド ドロップします。

または



- 描画オブジェクトボタンをクリックして、マウスを描画ツールに変えます。描画するオブジェクトが、マウスポインタの右下横に小さい記号として現われます。このツールをプロット上に置き、クリックおよびドラッグして指定したサイズの描画オブジェクトを挿入します。オブジェクトは別のオブジェクトを描画するか、シート上のどこか他の場所をクリックするまで選択されたままになります。

四角の選択ノブの 1 つを選んでドラッグして、オブジェクトのサイズを変更したり、オブジェクトの中心を選択しドラッグして、オブジェクトの移動が可能です。描画オブジェクトの他のプロパティを編集するときは、そのオブジェクトをダブルクリックしてプロパティダイアログを開くか、右クリックしてショートカットメニューからダイアログの適切なページを選択します。

グラフオブジェクトでの作業

オブジェクトを重ねる


グラフ上でオブジェクトが重なっていると、オブジェクトによっては他のオブジェクトに、完全にまたは部分的に覆われることがあります。オブジェクトを前面に出したり背面に入れたりすることによって、重なっているオブジェクトの順序を変更することができます。

- 前面に出すか背面に入れたいオブジェクトを選択します。グラフツールバーの**前面へ**ボタン  または**背面へ**ボタン  をクリックします。またはメインメニューから、**書式 ▶ 前面へ**または**背面へ**を選択します。オブジェクトを1つ分（1回に1レベル）だけ前または後にするときは、**書式**メニューから**1レベル前面へ**または**1レベル背面へ**を選択してください。

オブジェクトを削除する

オブジェクトをグラフから削除する

- 削除したいオブジェクトを選択します。DELETE キーを押すか、メインメニューから**編集 ▶ クリア**を選択します。

オブジェクトを削除した直後に、削除を取り消すことができます。メインメニューから**編集 ▶ 元に戻す**を選択するか、**標準**ツールバーの**元に戻す**ボタン  をクリックします。

オブジェクトは**グラフシート**から削除することができますが、**Cut** コマンドを使用すると永久的には抹消される訳ではありません。オブジェクトはクリップボード上に置かれ、そのオブジェクトを現在の**グラフシート**上の別の位置、または別の文書に貼り付けることができます。

オブジェクト・エクスプローラとグラフオブジェクト

グラフの個々のオブジェクトを選択する他の方法は、**オブジェクト・エクスプローラ**によるものです。グラフのオブジェクトはすべて、**オブジェクト・エクスプローラ**の**グラフシート**のルートノードの下に表示されます。そのノードを展開する（プラス記号をクリックする）と、現在表示されているすべてのグラフのノードが現れます。**オブジェクト・エクスプローラ**の右パネルに、個々のグラフオブジェクトが表示されるまで、**グラフシート**のツリーの下の方にノードを展開することができます。グラフオブジェクトをダブルクリックすると、そのグラフオブジェクトのフォーマットダイアログが表示されます。

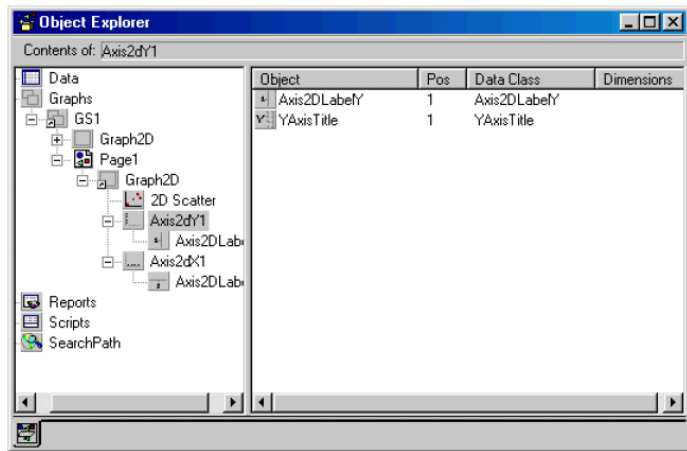


図 6.15 : オブジェクト・エクスプローラで表示されたグラフのオブジェクト

重なったり隠れたりしている複数のオブジェクトを含む複雑なグラフをフォーマットするとき、**オブジェクト・エクスプローラ**からグラフオブジェクトを選択する機能が役立ちます。**オブジェクト・エクスプローラ**からしかオブジェクトを選択できないケースは、ほとんどありません。

図表示

これまでプロットの作成方法を説明してきました。S-PLUS は汎用プログラムで、S-PLUS のプロットごとに多数のオプションを指定することができます。たとえば線形回帰プロットではデータがプロットされ、そのデータから線形回帰直線が描かれます。たとえばその線に信頼区間を追加したり、線の色を変更したり、データ点の色と記号を変更することなど様々なことができます。この節ではよく用いられるプロットについて説明し、プロットの外観を変更するプロット構成オプションのいくつかを紹介します。詳細は、オンラインヘルプを参照してください。

プロット プロパティ・ ダイアログ

プロットのプロットプロパティを変更する際の最初のステップは、他のグラフオブジェクトを選択するときと同じようにプロットを選択することです。プロット領域内のデータ点をクリックして、プロットオブジェクトを選択します。緑色の四角いノブがプロットの下の方に現われます。プロット領域を囲む 8 つの緑色の四角いノブが現われた場合は、プロット自体ではなくプロット領域が選択されています。

プロットを選択した後、次のいずれかの方法でプロットプロパティ・ダイアログを開きます。

- プロットをダブルクリックします。
- メインメニューから書式 ▶ 選択された Plot を選択します。
- プロットを右クリックしてそのショートカットメニューを表示させ、ダイアログページを選択します。

プロットプロパティ・ダイアログが現われたら、ダイアログの希望するプロパティを編集し、OK をクリックします。

グラフを フォーマット する (続き)

この章の始めの方で (213 ページを参照)、グラフのフォーマット方法を段階的に説明しました。次に引き続き回帰直線に信頼区間を追加し、更にプロットしたデータの外観を変更します。得られたプロットを、図 6.16 に示します。

Weight vs. Mileage in Passenger Cars

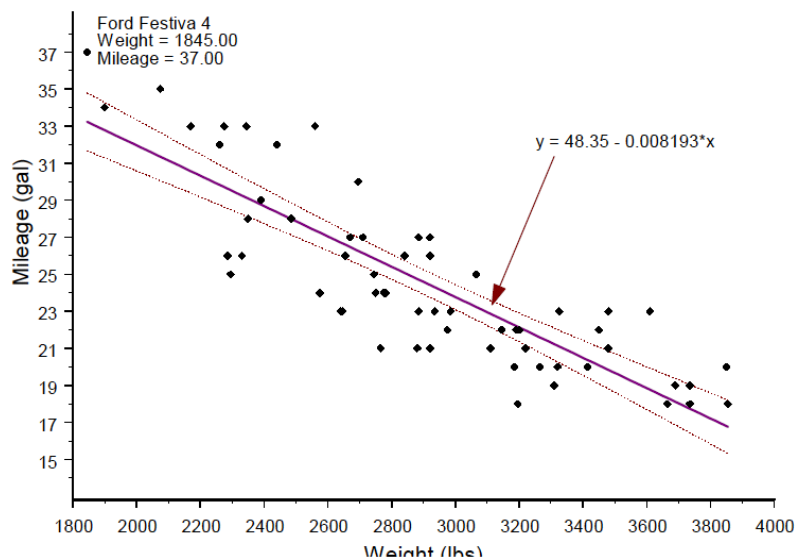


図 6.16 : 95%の信頼区間を追加しプロット記号を変更した後の図 6.9

この例のとおりにするためには、図 6.8 に示したプロットを作成しなければなりません。213 ページのステップに従って図を生成してください。少なくとも、基本のプロットを生成するためにはステップ 1~3 を実行しなければなりません。

線、記号、色

1. 回帰直線または任意のデータ点をクリックしてプロットを選択します。プロットが選択されていることを示すために、プロット領域の右下角に緑色の四角いノブが現われるはずです。メインメニューから、**書式** ▶ **選択された Curve Fitting Plot** を選択し、プロットプロパティのダイアログを開きます（または、この節の始めの方で説明した他の方法のいずれかを使用します）。
2. **Line/Symbol** ページに移り、**Line** グループの **Color** に **Magenta** を選択し、**Line** グループの **Weight** に **2** を選択します。**Apply** をクリックしたときに回帰直線がどのように変化するか注意してください。
3. 同じ **Line/Symbol** ページで、**Symbol** グループの **Style** として **Diamond, Solid** を選択し、**Symbol** グループの **Color** として **Black** を選択します。**Apply** をクリックします。

信頼区間

次に、プロットに信頼区間を追加します。

4. ダイアログの **By Conf Bound** タブをクリックします。回帰直線に **95%** の信頼区間を追加するときは、**Confidence 0.95** をハイライトし、**Line Attributes** グループの **Style** を選択して点線（最後のオプション）にし、**Line Attributes** グループの **Color** が **Red** になるように選択してください。**OK** をクリックし、プロットに **95%** の信頼区間が現われたことを確認してください。

これで、プロットは図 6.16 のようになります。

モデル オプション


プロットプロパティのダイアログを用いて、データの特性を探索することができます。たとえば、これまでは単純な線形の回帰直線 $y = \alpha + \beta x + \varepsilon$ をあてはめてきましたが、回帰モデル $y = \alpha + \beta_1 x + \beta_2 x^2 + \varepsilon$ の方が適切ではないかと疑問に思われているかも知れません。

5. このモデルをあてはめるためにはプロットプロパティ・ダイアログを開き、**Curve Fitting** ページの **Poly.Order** に **2** を入力します。グラフが再描画されるときに曲線と信頼区間がどのように変化するかに注意してください。

いくつか他の値を試してみてください（このデータセットでは、7 よりも大きい数だと問題が生じます）。またデータをあてはめるために、デフォルトの線形モデルではなく指数モデルやべき乗モデルを選択することもできます。

グラフの種類を 変更する

グラフの種類を変更しても設定済みの詳細なフォーマットは失われません（202 ページを参照）。

6. グラフの種類を変更するときは、プロットが選択されていること（緑色の四角いノブが右下角にあること）を確かめてください。**2D プロットパレット** を開き、**局所重み付き最小二乗法** ボタン  をクリックします。グラフは、このデータによって描かれた **Loess** 線で再描画されます。
7. 曲線近似式と矢印（**Loess** 線にあてはまっていない）を削除してください。オブジェクトをそれぞれ選択し、メインメニューからの **編集 ▶ 切り取り** を選択してそれらを削除します（**Loess** はノンパラメトリック回帰なので式の表示ができません）。

グラフスタイルを使用し色をカスタマイズする

グラフスタイルを使用して、デフォルトのプロットの外観をカスタマイズすることができます。**カラースタイル**と**白黒スタイル**の2つのグラフスタイルを使用することができます。**カラースタイル**を使用すると、様々な色を使用してプロットを作成することができます。**白黒スタイル**は異なるスタイルの線と記号を使用してプロットを識別しますが、色は変化しません。新しい**グラフシート**に使用するグラフスタイルを指定するときは、メインメニューから**オプション ▶ グラフオプション**を選択し、希望する**グラフスタイル**を選択してください。**グラフシート**をあるスタイルから別のスタイルに変更するときは、メインメニューから**書式 ▶ スタイルの適用**を選択してください。

オプション ▶ グラフスタイルより得られるダイアログを使用すると、それぞれのグラフスタイルをカスタマイズすることができます。それぞれのダイアログの最初のページで、**User Colors** と背景色を含む配色を指定することができます。**User Colors** は、オブジェクトの色を編集するときにドロップダウン・カラーリストに現れる16の選択可能な色です。これらの色は、新しく作成された**グラフシート**に必ず設定され、デフォルトの**グラフシート**での色指定を無効にします。メインメニューから**書式 ▶ グラフシート**を選択して作成した後で、特定の**グラフシート**の背景色と色パレットを編集することができます。

グラフスタイルのダイアログの最初のページで、グラフに複数のプロットを描く場合に色、線のスタイル、パターン、そして記号のスタイルを変化させるかを指定することもできます。たとえば、同じ軸の目盛りを使用する散布図で2つの線を作成することがしばしばあります。このような場合に異なる線の色、線のスタイルそして記号のスタイルを使用してプロットを作成するように指定することができます。

グラフスタイルのダイアログの残りのページで、第1のプロット、第2のプロットなどをプロットするために使用される色、線のスタイル、パターンそして記号のスタイルを指定することができます。色として**Plot Default**を選択すると、**S-PLUS** は作成しているグラフのデフォルトのプロットオブジェクトの色設定を使用します。プロットのデフォルトを修正するときは希望するグラフのプロットを作成し、希望する外観となるように編集します。次にプロットを右クリックして出てくるショートカットメニューで、**Save [Plot Type] as Default** を選択します。

可能なカラースキームは 8 つです。配色は、**オプション ▶ カラースキーム**より得られるダイアログを使ってカスタマイズすることができます。この場合各カラースキームごとに **Background Color**、**User Colors** および **Image Colors** を修正することができます。たとえば **Standard** の配色の **User Colors** を編集すると、新しく作成される **グラフシート** はすべて修正した **User Colors** で表示されます。**Image Colors** は、イメージ鳥瞰図またはカラー等高線図に使用することができる特別なカラーパレットです。**Image Colors** は、指定したそれぞれの色を補間する陰影の数を指定することができます。

プロジェクトによって異なる背景色とユーザ設定色を使用したい場合は、必要に合わせて 8 つのカラースキームのいずれかまたはすべてを編集することができます。カラースキームを設定した後でも、それらの間で容易に変更することができます。たとえば、**Color Graph Style** を使用する場合は、**オプション ▶ グラフスタイル ▶ 色** を選択します。**User Colors** を通常 **Trellis** グラフに使用されている配色である **Trellis** に設定します。以降、新しいすべての **グラフシート** は、**Trellis** 配色で指定された **Background Color** と **User Colors** を使用して作成されます。デフォルトで **Standard** 色を使用するように戻すときは、**オプション ▶ グラフスタイル ▶ 色** に戻り、**User Colors** を **Standard** に設定しなおしてください。

グラフシートへのデータの埋込みと抽出

グラフにさらに修正を加えたいと思っている仲間と、**グラフシート**を共有したい場合があるかもしれません。これは、**グラフシート**にグラフの作成に使用されるデータを埋め込むだけで可能になります。**グラフシート**にデータを埋め込んでしまえば、あとは**グラフシート**を配布するだけです。(統計のダイアログによって作成されたグラフには、自動的にデータが埋め込まれます。)

グラフシートにデータを埋め込むときは、実際にグラフに表示されている変数だけが埋め込まれます。したがって、たとえば **20** の変数を含むデータセットから **4** つの変数を使用してプロットを作成しデータを埋め込む場合は、**4** つの変数だけが埋め込まれます。

任意の**グラフシート**からデータを抽出することができます。これは、所定のグラフを作成するために使用されるデータだけで構成される新規データセットを作成するのに便利な方法です。

グラフシートにデータを埋め込む

1. **グラフシート**にグラフを作成します。
2. メインメニューから**グラフ ▶ データのグラフへの埋め込み**を選択します。プロットを再現するために必要なデータが、**グラフシート**に埋め込まれます。

データを埋め込んだ後、**グラフシート**は最初にグラフを作成するために使用したデータセットにもうリンクされません。すなわち、最初のデータセットへの変更が**グラフシート**に反映されなくなります。またデータを埋め込むと、**グラフツール**の**矩形内のデータ選択**ボタンを使用することができません。

グラフシートからデータを抽出する

1. データを表示する**グラフシート**を開きます。プロットプロパティ・ダイアログでは、データを埋め込んだ**グラフシート**の**データセット名**が **=Internal** となっていますが、どの**グラフシート**からもデータを抽出することができます。
2. メインメニューから**グラフ ▶ データのグラフからの抽出**を選択します。**Extract Data** ダイアログが現れます。
3. 抽出したデータセットの名前を入力し、**OK** をクリックします。

オブジェクトのリンクと埋込み

S-PLUS は、他のアプリケーションで作成されたデータまたはオブジェクトをグラフシートで使用できるようにリンク機能と埋込み機能をサポートしています。この節では別のアプリケーションからのデータと S-PLUS のプロットをリンクし、そのプロットをソースデータに接続する方法を説明します。

次のようなときに、S-PLUS のプロットをデータにリンクさせてください。

- データを変換する可能性があるとき。
- S-PLUS のプロットに最新のデータを必要とするとき。
- 元の文書がいつも使用可能で、更新が必要なとき。

また、S-PLUS のデータまたはグラフシートを Word や Excel などの別のアプリケーションに埋め込むことができます。埋め込んだオブジェクトは、別のアプリケーションから S-PLUS を使用して編集することができます。この節では、別のアプリケーションにグラフシートを埋め込む方法を説明します。

次のようなときに、データまたはグラフを埋め込んでください。

- 埋め込んだ情報を変更する予定がないとき。
- 埋め込んだ情報を複数の文書に使用する必要がないとき。
- 元の文書がリンクされている場合に、その文書が更新に使用されないとき。

注意

リンク機能または埋込み機能を使用するときは、元のアプリケーションが OLE (object linking and embedding) をサポートしていなければなりません。たとえば Excel 5.0 は OLE をサポートしているため、Excel 5.0 のデータを S-PLUS のグラフシートに埋め込み、リンクすることができます。

リンクを編集する

S-PLUS グラフシートのそれぞれのリンクを制御することができます。デフォルトでは、データは自動更新によってプロットにリンクされます。これは、リンクダイアログで手動更新に変更することができます。

リンクを編集する

1. メインメニューから**編集 ▶ S-PLUS ウィザード**を選択します。
2. リンクダイアログで、編集するリンクを選択します。
3. **自動**または**手動**リンクを選択します。**手動**の更新では、**リンクダイアログ**から**今すぐ更新**ボタンを選択したときだけリンクが更新されます。

リンクの再接続または変更

ソースファイルを名前変更または移動する場合は、リンクを再度設定する必要があります。**リンクダイアログ**ではソースファイルを名前変更するか、ソースファイルの新しい位置を指定する必要があります。

リンクの再設定または変更

1. メインメニューから**編集 ▶ S-PLUS ウィザード**を選択します。
2. リンクしたソースファイルの名前を変更するか、異なるファイル名を指定します。**OK**をクリックします。

他のアプリケーションに S-PLUS グラフを埋め込む

S-PLUS は OLE をサポートしているため、PowerPoint や Word などの他のアプリケーションに S-PLUS の**グラフシート**を埋め込むことができます。

S-PLUS のグラフシートを埋め込む



1. Word などのクライアントアプリケーションをロードします。クライアントアプリケーションの**挿入メニュー**から**オブジェクト**を選択します。
2. **新規作成タブ**を選択し、**オブジェクトの種類**から **S-PLUS Graph Sheet** を選択します。**OK** をクリックします。これで、新しい**グラフシート**を作成しアクティブにすることができます。

3. グラフシートを編集し終わったら、そのグラフシートの外側をクリックして非アクティブ化します。埋め込まれたグラフシートが文書中で表示されます。


埋め込まれているグラフシートをその場で編集する

1. クライアントアプリケーションの文書で、埋め込んだ S-PLUS グラフシートをダブルクリックします。あるいはグラフシートを選択して、編集メニューから **Graph Sheet** オブジェクトを選択し、次に **Edit** を選択します。
2. 埋め込まれたグラフシートはクライアントアプリケーションに残りますが、メニューとツールバーは、S-PLUS のものになります。S-PLUS のメニューとツールバーを使用して、グラフシートを編集してください。
3. 終了したら埋め込まれたオブジェクトの外側をクリックすると、クライアントアプリケーションのメニューとツールバーに戻ります。

グラフを印刷する

S-PLUS のグラフシートを印刷するときは、Windows 標準の印刷ボタン  またはファイルメニューのグラフシート印刷オプションを使用します。グラフシートを1度に全ページでも1ページでも印刷することができます。印刷ボタン  は必ず現在のページだけを印刷しますが、グラフシート印刷を選択したときに表示される **Print** ダイアログには、全ページまたは一部のページを印刷するオプションがあります。

グラフシートを印刷する

- 標準ツールバーの印刷ボタン  をクリックします。印刷の選択の確認をするダイアログが現われます。

または

1. メインメニューからファイル ▶ グラフシート印刷を選択します。**Print** ダイアログが現れます。
2. **Print** ダイアログで必要なオプションを選択します。オプション説明についてはオンラインヘルプを参照してください。
3. **OK** をクリックすると印刷が始まります。

グラフをファイルにエクスポートする

S-PLUS のグラフを、他のアプリケーションで使用するための様々なフォーマットのファイルにエクスポートすることができます。そのようなファイルは、S-PLUS とリンクされなくなります。グラフシート (*.SGR ファイル) の別のアプリケーションとの共有については、240 ページの「オブジェクトのリンクと埋込み」を参照してください。

グラフをファイルにエクスポートする

1. エクスポートしたいグラフが、現在のグラフシートに表示されていることを確かめてください。
2. メインメニューから **ファイル** ▶ **グラフのエクスポート** を選択します。図 6.17 に示したような **Export Graph** ダイアログが表示されます。

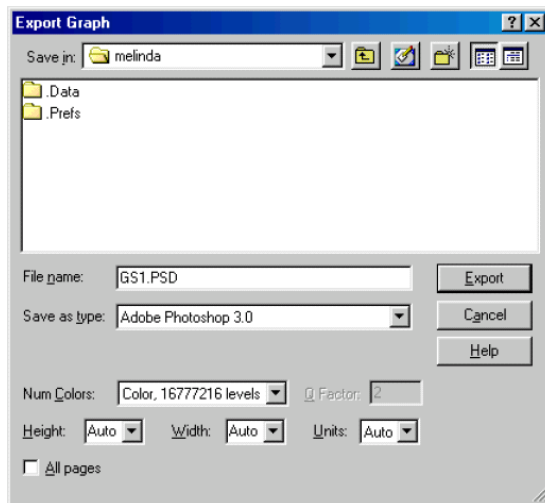


図 6.17 : Export Graph ダイアログ

3. 必要に応じて希望のターゲットフォルダに進みます。
4. **ファイルの種類** ドロップダウンリストから希望のファイルタイプを選択します。
5. **ファイル名** ボックスでエクスポートされたグラフのファイル名を入力します。
6. **Export** をクリックします。

マルチページ グラフシートを エクスポート する

グラフシートに複数のページがあって、そのすべてのページを同時にエクスポートしたい場合は、以下のように行います。

1. マルチページグラフシートがアクティブな文書であることを確認します。
2. メインメニューから、**ファイル ▶ グラフのエクスポート**を選択して**Export Graph** ダイアログを開きます。
3. 必要に応じて希望のターゲットフォルダに移動します。
4. **ファイルの種類**ドロップダウンリストから希望のファイルタイプを選択します。
5. **ファイル名**ボックスにエクスポートされたグラフのベースになるファイル名を入力します。
6. **Export Graph** ダイアログの左下隅にある**All pages** チェックボックスを選択します。このオプションを選択すると、ファイル名ボックスが使用できなくなり、ファイル名の最後のドットの前に記号#数字が挿入されます。
7. **Export** をクリックします。

すべての**グラフシート**ページが別のファイルにエクスポートされ、それぞれのファイル名の記号#数字が、**グラフシート**の対応するページ番号を示します。

第7章 S-PLUS GRAPHLETS™

はじめに	248
要件	250
例	250
Graphlet データファイルの作成	252
対話式のファイル作成	252
プログラムによるファイル作成	252
ページタイトルとタグ	252
アクティブ領域	253
アクション文字列	256
例の詳細	259
ウェブページへの Graphlet の埋め込み	263
アプレットパラメータ	264
例の詳細 (続き)	268
Graphlet の使い方	269

はじめに

S-PLUS Graphlet™は、S-PLUS グラフの表示および S-PLUS グラフとの対話を可能にするファイルに保存されている Java アプレットです。Graphlet は、ウェブページに埋め込み、ウェブブラウザで見ることができます。たとえば、図 7.1 は、Netscape のウィンドウに Graphlet の例を表示させたところです。

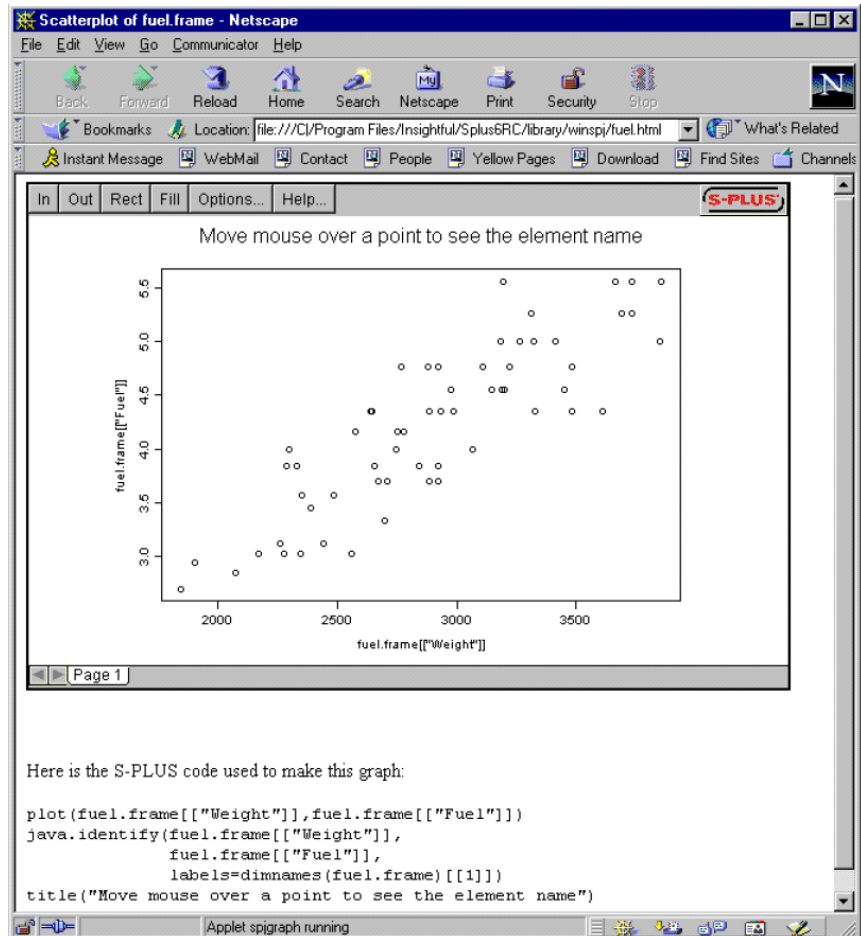


図 7.1 : Graphlet の例

Graphlet を使って S-PLUS グラフを表示すると、JPEG や GIF 画像などの静的なグラフ画像では得ることができないいくつかのインタラクティブな機能を用いることができます。

1. Graphlet は、S-PLUS `java.graph` グラフ表示と同じように、複数のタブ付きページのあるグラフウィンドウを表示します。多くの場合、関連した一連のグラフを 1 つのグラフウィンドウの複数ページに集めておくと便利です。グラフの作成者は、ページタブのタイトルに分かりやすい名前を設定することができます。
2. 表示させたグラフをパンしたりズームしたりして、細部をより詳しく調べることができます。グラフは、ビットマップとしてではなく、グラフィックコマンドで作成されたベクタグラフとして表示されるため、ズームで拡大した画像も元のサイズの時と同様にはっきり表示されます。
3. マウスがグラフ上を移動するのに合わせて、マウスの「グラフ座標」を表示させることができます。このグラフ座標は、単にウィンドウ内の xy ピクセル座標ではなく、グラフの軸に対するマウスの位置を xy 座標で表したものです。したがって、この表示を使って、座標の対象となっている散布図やその他のグラフ上の点の位置を調べることができます。1 つのページに複数のグラフがある場合は、マウスのあるグラフ上の座標が表示されます。
4. グラフの作成者は、ページ内に長方形の「アクティブ領域」をいくつでも定義することができます。あるアクティブ領域上にマウスを移動すると、その領域がアウトライン表示され、そのアクティブ領域に関連付けられているラベルが表示されます。ラベルは、グラフの要素に関する追加情報を表示する文字列です。たとえば、散布図内のすべての点のまわりにアクティブ領域を定義し、それぞれの点に名前情報を与えることができます。
5. それぞれのアクティブ領域は、「アクション文字列」と関連付けることができます。このアクション文字列は、あるアクティブ領域内でマウスがクリックされたときに起きる動作を指定します。オプションには、同一 Graphlet 内での他ページへの切り替え、特定のウェブページにジャンプさせること、あるいはメニューをポップアップさせて複数のアクションから 1 つを選択することなどの機能があります。これにより、グラフの特定領域に関するより詳細な情報を“ドリルダウン”して表示させることができます。

Graphlet と S-PLUS は、同時に動作していないことに注意してください。S-PLUS によって、一連のグラフィックコマンドがファイルに保存されます。その後で、Graphlet が、そのファイルを読み取り、グラフとして表示します。グラフに表示されるのは、S-PLUS セッションが終了してからずっと後かもしれません。S-PLUS を持っていないユーザがウェブを介して見ているかもしれません。

このような S-PLUS グラフを作成して使用するには、次の 3 つの主要なステップがあります。

1. Graphlet データファイルを SPJ フォーマットで作成する。
2. Graphlet をウェブページに埋め込む。
3. Graphlet を表示し操作する。

この章では、これらのステップを詳しく説明します。「グラフ作成者」は、一般に、最初の 2 つのステップに関係し、「エンドユーザ」が関係するのは 3 つ目のステップだけです。S-PLUS Graphlet の操作方法について知りたいエンドユーザは、「Graphlet の使い方」(269 ページ) にスキップしてください。

要件

Graphlet ファイルは、`java.graph` グラフ・デバイスに準拠しており、したがって、Graphlet ファイルの作成者は `java.graph` にアクセスできなければなりません。S-PLUS Graphlet は、Java バージョン 1.1 以降の Java アプレットをサポートしているウェブブラウザで使用するよう設計されています。

Graphlet は、従来のコマンド形式の S-PLUS グラフを表示します。グラフィックコマンドおよびオプションの詳細については、『S-PLUS Programmer's Guide』の次の 2 つの章を参照してください。

- 第 2 章「Traditional Graphics」
- 第 3 章「Traditional Trellis Graphics」

例

`$SHOME/library/example5/winspj` フォルダには、いくつかの Graphlet ファイルの例があります。この章では `fuel.html` と `map.html` の 2 つについて説明します。ブラウザによっては、セキュリティ保護のため、コンテンツが表示されない場合がありますが、表示された警告に従って、表示させてください。

Insightful Corporation のウェブサイト www.insightful.com/graphlets にも、Graphlet を使った見本が多数あります。

注意

java.graph を使って Graphlet を作成するには、最初に winjava または winspj ライブラリを追加する必要があります。

Graphlet データファイルの作成

S-PLUS グラフィックスを Graphlet ファイルに対話的またはプログラム実行によって保存することができます。

Winjava ライブラリのロード

Graphlet を作成するのに必要なライブラリをロードします。ファイルメニューで、ライブラリのロード、Library Name で winjava を選択してください。

対話式の ファイル作成

Graphlet データファイルを対話式に保存するときは、java.graph ウィンドウでグラフを作成します。java グラフウィンドウの **File** メニューで **SaveAs** メニュー項目を選択します。ファイルタイプとファイル名を選択するダイアログが現れます。ファイルタイプ **S-PLUS Graphlet (SPJ)** を選択し **OK** を押して、カレントグラフのすべてのページを特定のファイルに書き込みます。

プログラムによる ファイル作成

Graphlet データファイルをプログラムで保存するときは、java.graph 関数を呼び出し、ファイル名とファイルタイプ"SPJ"を次の例のように指定します。

```
> java.graph(file = "myfile.spj", format = CSPJ)
```

これにより、特定のファイルに Graphlet データを保存する S-PLUS グラフィックデバイスが保存されます。このコマンドではウィンドウが開かないことに注意してください。次に、希望のグラフを作成するのに必要なグラフィックコマンドをすべて実行します。この作業が終わったら、dev.off 関数でファイルを閉じます。グラフィックデバイスを終了させると、その内容が、名前を付けたファイルに書き込まれます。

ビットマップフォーマットの中には、JPEG のように 1 つのファイルに 1 つの画像しか保存できないものもありますが、SPJ フォーマットでは同一ファイルに複数のグラフページを保存することができます。java.graph を使って SPJ ファイルを作成すると、描画したページはすべて 1 つのファイルに保存されます。

ページタイトル とタグ

複数ページの Graphlet 画面の各ページには、対応する「タイトル文字列」と「タグ文字列」があります。タイトル文字列は、タブ上に表示され、空文字列""で小さい空タブを示すことがあります。タイトルが、文字列"#"の場合は、表示されるタイトルは、たとえば"Page 3"のような自動生成される文字列です。タグ文字列は表示されませんが、「Graphlet 内の別のページにジャンプする」(257 ページ)で説明している page アクション文字列のターゲットページを特定するのに使用されます。

java.graph デバイスでグラフを作成している時点で、「カレントページ」が

表示されます。カレントページのタイトルおよびタグは、次のような S-PLUS 関数で設定されます。

```
> args(java.set.page.title)
function(name)

> args(java.set.page.tag)
function(tag)
```

アクティブ領域

java.identify 関数は、グラフ内の長方形「アクティブ領域」とそれに対応するテキストラベルとアクションを指定します。ユーザが指定された領域内でマウスを動かしたとき、Graphlet は、これらのラベルを表示します。以下は、Weight と Fuel の関係を示す散布図内の各点の上にマウスを移動させるときに自動車の型名が表示される簡単な例です。

```
# Open a java.graph device.
> java.graph(file = "fuel.spj", format = "SPJ")

# Plot the data and identify the points.
> plot(fuel.frame[["Weight"]], fuel.frame[["Fuel"]])
> java.identify(fuel.frame[["Weight"]],
+ fuel.frame[["Fuel"]],
+ labels = dimnames(fuel.frame)[[1]])

# Close the device.
> dev.off()
```

図 7.2 は、S-PLUS 上に表示されているグラフです。fuel.spj ファイルをウェブページに埋め込んだ後、エンドユーザは、そのグラフをウェブブラウザで見て、またグラフ上の各点にマウスを移動することによって、それぞれの自動車の型名を表示させることができます。一例として、**\$SHOME/library/winspj** のファイルを参照してください。SPJ ファイルを埋め込むために必要なステップは、「ウェブページへの Graphlet の埋め込み」(263 ページ) で説明します。

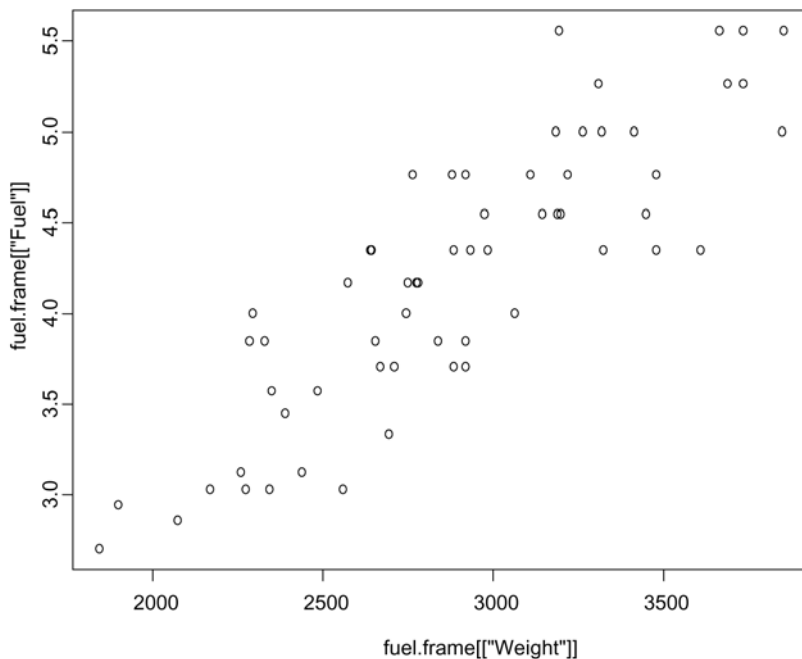


図 7.2: fuel.frame データセットの Weight と Fuel の関係を示す散布図

java.identify 関数では、多くのオプション引数を使用することができます。それぞれの引数について、以下で説明します。

```
> args(java.identify)
function(x1, y1, x2, y2, labels, actions = character(0),
size = 0.01, size.relative = T, adj = 0.5)
```

x1、y1、x2、y2

これらの引数は、現在のプロット内にある一連の長方形を指定します。それぞれの長方形が、1つのアクティブ領域を定義しています。引数には、数値ベクトル、長さも含めて使用できます。引数の様々な組み合わせを、次のように指定することができます。

1. $x1$ 、 $y1$ 、 $x2$ 、 $y2$ が、すべて指定される場合は、1組の任意の長方形領域が、 $(x1,y1)$ 、 $(x1,y2)$ 、 $(x2,y1)$ 、および $(x2,y2)$ の各点で定義されます。この領域は、プロット領域の内側でも外側でもかまいません。
2. $x1$ と $y1$ だけが与えられた場合は、領域が各点 $(x1,y1)$ のまわりに作成されます。この方法で、上記 **fuel.spj** の例のアクティブ領域を指定することができます。
3. $x1$ と $x2$ が与えられた場合は、領域が x 軸に沿って $x1$ から $x2$ まで作成され、この間の y 軸の値はすべてプロット領域となります。
4. $y1$ と $y2$ が与えられた場合は、領域が y 軸に沿って $y1$ から $y2$ まで作成され、この間の x 軸の値はすべてプロット領域となります。
5. $x1$ だけが与えられた場合は、領域が x 軸上のこれらの値のまわりに作成され、この間の y 軸の値はすべてプロット領域となります。
6. $y1$ だけが与えられた場合は、領域が y 軸上のこれらの値のまわりに作成され、この間の x 軸の値はすべてプロット領域となります。

labels

labels 引数は、プロット内のアクティブ領域と対応する文字列のベクトルです。各ラベルは、ユーザがマウスをそのラベルに対応した領域の上に移動させたときに表示されます。ラベルは、復帰改行文字"`\n`"を含み、複数行のテキストを表示することができます。上記 **fuel.spj** の例において、`labels = dimnames(fuel.frame)[[1]]`と指定し、これにより、`fuel.frame` データセットの行名を、図 7.2 の各点と関連付けます。

actions

labels 引数と同様に、**actions** 引数も、プロット内のアクティブ領域と対応する文字列のベクトルです。この引数は、ユーザが各アクティブ領域内でマウスを左クリックしたときに起こるアクションを定義します。`length(actions) < length(labels)`の場合、**actions** ベクトルは、適切な長さになるまで（アクションを指定しない）空白文字列でくり返されます。アクション指定文字列のフォーマットについては、「アクション文字列」（256 ページ）で説明します。

size

点だけが指定されている場合は、size 引数が、各点のまわりのアクティブ領域の大きさを決定します。size 引数は、別々の x 値と y 値として解釈される 2 つの数字のベクトルで表すことができます。size.relative=T の場合、size 値は、プロットの x 軸または y 軸上の範囲に対する割合です。その他の場合、size は、ピクセルで表した絶対値として解釈されます。

size.relative

これは、size 引数をどのように解釈するかを決定する論理値です。

adj

点だけが指定されているケースでは、adj 引数が、点に対してアクティブ領域が配置される場所を決定します。size と同様に、adj 引数も、別々の x 値と y 値を指定する 2 つの数字のベクトルで表すことができます。デフォルト値では、点はアクティブ領域の中央に位置決めされます。adj=0.0 の場合は、関連付けられた左下角の点でアクティブ領域が位置決めされます。

アクション 文字列

java.identify 関数の actions ベクトルの各要素は、対応するアクティブ領域内でマウスが左クリックされたときに起こるアクションを指定します。特に、それぞれ別のアクションになる選択肢のメニューをポップアップさせるときには、アクションがかなり複雑になる可能性があります。そのような複雑なアクションを 1 つの文字列で対応し、かつ将来の機能強化を可能にするために、アクション文字列は、XML フォーマットで指定されます。このフォーマットを間違わずに記述するのは難しいため、S-PLUS には、後で説明するように、文字列を作成するための一組の関数が含まれています。アクション文字列が、そのようなフォーマットでない場合は、アクティブ領域上でクリックしても何も起きません。

別の ウェブページに ジャンプする

別のウェブページにジャンプするための XML フォーマットのアクション文字列の例は、次の通りです。


```
<link href="http://www.insightful.com" target="_top" />
```

href プロパティは、1つのウェブページを指定する URL を示します。この例では、ウェブページは、**www.insightful.com** です。target プロパティは、このウェブページを表示する HTML のフレームを指定します。target が示されていない場合、デフォルトで "_top" となり、この場合、ウェブブラウザに表示されている現在のページの代わりに指定したページが表示されます。もう1つの有用な target 値は、"_blank" です。これは、URL をウェブブラウザの新規ウィンドウに表示します。ウェブブラウザの中には、target プロパティに対応していないものもありますので注意してください。

このアクションに対応する S-PLUS 関数は、次のような `java.action.link` です。

```
> args(java.action.link)
function(url, target = "_top")
```

文字列として URL が与えられると、`java.action.link` は、アクション文字列を上記のフォーマットで返します。URL のベクトルを与えると、対応するアクション文字列のベクトルが返されます。target が指定されている場合、それがアクション文字列に使用されます。指定されていない場合は、デフォルトで "_top" になります。たとえば、上記のアクション文字列は、次の S-PLUS コードで生成することができます。

```
> java.action.link("http://www.insightful.com")
[1] "<link href=¥"http://www.insightful.com¥"
target=¥"_top¥" />"
```

Graphlet 内の別のページにジャンプする

Graphlet 内の別のページにジャンプするための XML フォーマットのアクション文字列の例は、次の通りです。

```
<page tag="p3" />
```

tag プロパティは、選択するページタグを指定します。指定されたタグのあるページが Graphlet 内にはない場合は、アクションが選択されても何も起きません。ページタグは、「ページタイトルとタグ」(252 ページ) で説明したように、`java.set.page.tag` 関数で定義されます。

このアクションに対応する S-PLUS 関数は、次に示す `java.action.page` です。

```
> args(java.action.page)
function(tag)
```

ページタグが与えられると、`java.action.page` は、アクション文字列を上記のフォーマットで返します。この関数は、ページタグのベクトルを与えると対応するアクション文字列のベクトルが返されます。たとえば、上記のアクション文字列は、次の S-PLUS コードで生成することができます。

```
> java.action.page("p3")
[1] "<page tag=¥"p3¥"/>"
```

アクションの メニューを ポップアップ する

アクション選択枝のメニューをポップアップするための XML フォーマットのアクション文字列の例は、次の通りです。

```
<menu title="some actions">
  <menuitem label="go to page1">
    <page tag="p1"/>
  </menuitem>
  <menuitem label="go to URL">
    <link href="http://www.insightful.com" target="_top"/>
  </menuitem>
</menu>
```

対応する S-PLUS 関数は、次のような `java.action.menu` と `java.action.menuitem` です。

```
> args(java.action.menu)
function(items = character(0), title = "")

> args(java.action.menuitem)
function(action, label = "item")
```

`java.action.menu` 関数では、それぞれが `java.action.menuitem` で作成された `menuitem` アクションのベクトルを指定することができます。この関数は、1つのメニューアクションコマンドを定義する文字列を返します。メニューのタイトルは、`title` 引数を使って指定されます。`title=""` の場合、メニューにはタイトルがありません。プラットフォームの中には、ポップアップメニュー上のタイトルをサポートしていないものがあります。その場合は、`title` 引数は無視されます。

`java.action.menuitem` 関数は、アクション文字列のベクトルを指定可能で、対応する `menuitem` オブジェクトのベクトルを返します。メニューに、各メニューアイテムが、指定されたラベルと共に現れます。たとえば、上記のアクション文字列は、次の S-PLUS コードで生成することができます。

```
> java.action.menu(
+   java.action.menuitem(
+     action = c(java.action.page("p1"),
+               java.action.link("http://www.insightful.com")),
```

```
+     label = c("go to page1", "go to URL"),
+     title = "some actions")

[1] "<menu title=¥"some actions¥">¥n<menuitem label=¥
"go to page1¥"><page tag=¥"pl¥"/></menuitem>¥
n<menuitem label=¥"go to URL¥"><link href=¥
"http://www.insightful.com¥"target=¥"_top¥
"/></menuitem>¥n</menu>"
```

XML 文字列 ユーティリティ 関数

XML フォーマットでは、二重引用符や不等号 "<" を含む特定の文字が、特定の目的のために予約されています。そのような予約文字を文字列値に使用するときは、S-PLUS 関数 `java.xml.string` を使って予約文字を特殊な文字シーケンスに変換する必要があります。この関数は、文字列のベクトルを指定することができ、XML 引用符シーケンスを使って文字列に変換します。次に例を示します。

```
> java.xml.string("bad characters: ¥<hello & Goodbye>¥")

[1] "bad characters: &quot;&lt;/hello &amp;
Goodbye&gt;&quot;"
```

`java.xml.string` 関数は、XML 文字列を明示的に構築するときに便利です。`java.action.link` のような関数は、そのすべての文字列引数に関して `java.xml.string` を呼び出すので、通常はユーザがこの関数を呼び出す必要はありません。

例の詳細

この節では、サンプルファイル `map.spj` と似た Graphlet のファイルを作成するために必要なステップを説明します。`map.spj` の Graphlet を見るには、`/$SHOME/library/winspj` ディレクトリにある `map.html` ファイルを開きます。この Graphlet を操作するための手引きは、「Graphlet の使い方」(269 ページ) を参照してください。この節で作成する Graphlet は、`map.html` にあるものよりやや単純ですが、S-PLUS で SPJ ファイルを作成するための要点を理解することができます。使用するコードには、`java.set.page.title`、`java.set.page.tag`、`java.identify`、`java.action.page` など、この章で既に説明した S-PLUS 関数も一部含まれています。

`map.spj` Graphlet は、米国内 47 都市の人種による人口分布に関する複数ページにわたる人口調査データの複雑なグラフを表示しています。データはヴァージニア大学の Geospatial & Statistical Data Center

(<http://fisher.lib.virginia.edu/ccdb>) から取得したもので、通常の S-PLUS サンプルデータの一部ではありません。米国の地図上で各都市を探してラベルを付けるには、組み込まれているデータベクトル `city.name`、`city.x`、および `city.y` を使用します。

map.spj Graphlet には、次のような 3 つの主要な機能があります。

1. グラフの最初のページには、米国の地図が表示されます。この地図上には 47 の都市がラベル付けされており、それぞれが 1 つのアクティブ領域に対応しています。各都市を示す丸は、その都市の人口にしたがって大きさが決められており、大きい都市ほど丸が大きくなっています。
2. 地図上のアクティブ領域をクリックすると、その領域と関連した都市の人種データの棒グラフを含むページに切り替わります。また、特定の都市のページタブをクリックすることにより、上記の棒グラフを表示させることができます。
3. 棒グラフを含む各ページの下の部分には、クリックすることで最初のページに切り換えることができるアクティブ領域があります。この機能により、ユーザは、**Graphlet** 内を簡単に移動し、複数の都市のグラフを次々と選択し表示させることができます。

データが、**S-PLUS** のデータセット **census** に保存されていると仮定します（実際にはありません）。このデータセットには、以下に示したような列があるとします。

- **AreaName** : 組み込まれているベクトル **city.name** に表示されるときに都市名
- **Population** : 1986 年における都市の人口
- **Pct.white** : 1980 年における白人の人口の割合
- **Pct.black** : 1980 年における黒人の人口の割合
- **Pct.amerind** : 1980 年におけるアメリカン・インディアンの人口の割合
- **Pct.asiapac** : 1980 年におけるアジアまたは太平洋諸島の人口の割合
- **Pct.hisp** : 1980 年におけるヒスパニックの人口の割合
- **x** : 組み込まれているベクトル **city.x** に相当し、**usa** 関数によって設定された座標系によるその都市の負の（度数で表した）経度です。
- **y** : 組み込まれているベクトル **city.y** に相当し、**usa** 関数によって設定された座標系による（度数で表した）緯度です。

以下の各ステップでは、これらのデータの Graphlet を作成するのに使用した S-PLUS コードについて説明します。まず、ファイル **map.spj** を `java.graph` グラフィックデバイスとします。

```
java.graph(file = "map.spj", format = "SPJ")
```

次に、米国の地図をプロットし、`census` 内にある都市にラベル付けします。`symbols` 関数が地図上に丸を描きます。丸の大きさは、各都市の人口によって決定されます。`text` 関数が、丸の隣に都市名を書き込みます。

```
usa()
symbols(census$x, census$y,
        circles = sqrt(census$Population), inches = 0.05,
        col = 2, add = T)
text(census$x, census$y, paste("", census$AreaName),
     adj = 0, cex = 0.5, col = 3)
title("Racial distribution in US cities")
```

地図を含む Graphlet のページにラベル付けします。

```
java.set.page.title("USA map")
java.set.page.tag("p0")
```

次に、地図上の点のまわりにアクティブ領域を定義します。この段階で、各都市がそれぞれページタグ "p1"、"p2"、"p3"、... の 1 つと関連付けられます。`java.identify` に対する `labels` 引数は、マウスが各都市のアクティブ領域上に移動したときに Graphlet の右上角に都市の名前を表示します。

```
java.identify(census$x, census$y, labels = census$AreaName,
             actions = java.action.page(
                 paste("p", 1:length(census$x), sep=""))) )
```

都市名をアルファベット順にソートし、各都市の人種データを表す棒グラフを生成します。以下の `for` ループが繰り返されるたびに、棒グラフが 1 つ作成され、その棒グラフが専用の Graphlet ページに表示されます。

```
cities <- census$AreaName
for(thecity in sort(cities))
{
  datum <- census[census$AreaName == thecity, ]
  grps <- c("Pct.white", "Pct.black", "Pct.amerind",
           "Pct.asiapac", "Pct.hisp")
  grp.eng <- c("White", "Black", "American Indian",
             "Asian/Pacific", "Hispanic")
  x <- unlist(datum[grps])
  # The barchart() command creates a new Graphlet page.
  print(barchart(grp.eng ~ x, xlim = c(0,100),
```

```

        xlab = "Percent", ylab = "")
title(paste("Racial distribution in", thecity))

# Label this page of the Graphlet.
java.set.page.title(datum$AreaName)
java.set.page.tag(paste("p",
    match(thecity, cities, nomatch = 0),
    sep = "", collapse = ""))
# Define an active region below the x axis in the bar plot
# that returns you to the first page in the Graphlet.
title(xlab = "[Click here to return to USA Map]",
    adj = 0.0)
java.identify(y1 = par("usr")[[3]],
    y2 = par("usr")[[3]] - 100, labels = "return to map",
    actions = java.action.page("p0"))
}

```

最後に、グラフィックデバイスを閉じます。これにより、グラフィックコマンドからの情報がすべて **map.spj** に書き込まれます。

```
dev.off()
```

ウェブページへの Graphlet の埋め込み

S-PLUS Graphlet をウェブページに埋め込むためには、次に示すいくつかのファイルが必要です。

1. Graphlet の Java コードを含むファイル **spjgraph.jar**。このファイルは、配布された S-PLUS に含まれており、`$SHOME/library/winspj` ディレクトリにあります。このファイルのコピーは、無料で配布されています。
2. Graphlet 上で表示される S-PLUS グラフィックコマンドを含む SPJ ファイル。このファイルは、「Graphlet データファイルの作成」(252 ページ) で説明したように、S-PLUS のバージョン 6 以降で生成することができます。
3. ウェブページを定義する HTML ファイル。このファイルは、手動で作成することもでき、また多くの入手可能なウェブページエディタの 1 つを使用して作成することもできます。

ウェブページ内に S-PLUS Graphlet をアプレットとして埋め込むには、そのウェブページの HTML ファイルに APPLET オブジェクトが含まれていなければなりません。たとえば、「アクティブ領域」(253 ページ) で定義した **fuel.spj** ファイル用のきわめて簡単なウェブページは、次の通りです。

```
<HTML>
<BODY>
A Web page with an embedded S-PLUS graph:<BR>
<APPLET code="spjgraph.class" ARCHIVE="spjgraph.jar"
  WIDTH="967" HEIGHT="601">
<PARAM NAME=spjgraph.filename VALUE="fuel.spj">
</APPLET>
</BODY>
</HTML>
```

このページが、`/u/web/fuel.html` のファイルであると仮定します。Graphlet の Java クラスにアクセスするには、ファイル **spjgraph.jar** も同じ `/u/web` ディレクトリになければなりません。あるいは **spjgraph.jar** が `/u/web` のサブディレクトリにある場合には、ファイルの場所を相対パスで指定してください。たとえば次の APPLET コマンドは、ディレクトリ `/u/web/graph` 内で **spjgraph.jar** を探します。

```
<APPLET code="spjgraph.class" ARCHIVE="graph/spjgraph.jar"
  WIDTH="967" HEIGHT="601">
```

アプレットパラメータ `spjgraph.filename` は、Graphlet がロードする SPJ グラフファイルの名前を指定します。上の例では、`/u/web/fuel.spj` のファイルがロードされています。`spjgraph.filename` が指定されていない場合は、デフォルトで **graph.spj** がロードされます。アプレットの幅と高さは、APPLET コマンドで指定され、ウェブブラウザで変更することはできません。

このとき、**spjgraph.jar** 内の Graphlet コードが、Graphlet が最新の Java プリント（回転フォントを含む）を処理できる JVM（Java バーチャルマシン）で動作しているかどうかを判断し、Graphlet **Print** ボタンから印刷するときこれをを使用することに注意してください。特に、プロットラベルはプリントが改善されています。これまでのプログラムは、文字の縁を「ぎざぎざ」にプリントしていました。

JVM が従来のも（Internet Explorer のデフォルトの Microsoft JVM など）である場合、プリントはこれまでと同じになります。新しいプリント機能を使用するときは、Java プラグインが必要になります。

比較のため、次のようにアプレットプロパティを指定して、従来のプリント動作を行わせることができます（新しい JVM でも）。

```
<PARAM NAME=3Dspjgraph.graphics.2d VALUE=3DFALSE>
```

新しいプリントコードは新しいプリントダイアログを使用します。プリンタは、別の **Printer...** ボタンで指定されます。

Graphlet の **Print** ボタンを使用して Graphlet の内容をプリントするときは、変更されたコードだけがプリント品質に影響します。Graphlet を含むブラウザーページ全体を印刷するときはプリント品質は改善されません。これは、ブラウザーがプリント Java アプレットを処理する方法によるもの（制御範囲を超えている）と考えられます。

例えば、次のページをプリントしてみましょう。

<http://www.insightful.com/products/graphlets/gallery/TestJavaApplet.asp>

このページには十字線のある枠を描く単純なアプレットが含まれています。プリントされたページでは、文字が「ぎざぎざ」の斜線で縁取りされます。

アプレット パラメータ

1 組のオプションの PARAM 値で、グラフファイルと Graphlet の上部に表示されるボタンの両方を制御します（図 7.3 を参照してください）。上記で説明したように、パラメータ `spjgraph.filename` が、ロードする SPJ グラフファイルの名前を指定します。残りのパラメータは、Yes、ON、Y、T のどれも ON と解釈し、それ以外はすべて OFF と解釈します。たとえば、APPLET コマンドに次のような要素を追加すると、初期状態でマウスの座標が表示され、エンドユーザは、グラフのサイズを変更できなくなります。


```
<PARAM NAME=spjgraph.mouse.position VALUE="ON">  
<PARAM NAME=spjgraph.resize.buttons VALUE="OFF">
```

利用可能なそれぞれの PARAM オプションについて、下の表 7.1 で簡単に説明します。S-PLUS Graphlet 上に表示されるボタンについての詳細は、「Graphlet の使い方」(269 ページ) を参照してください。

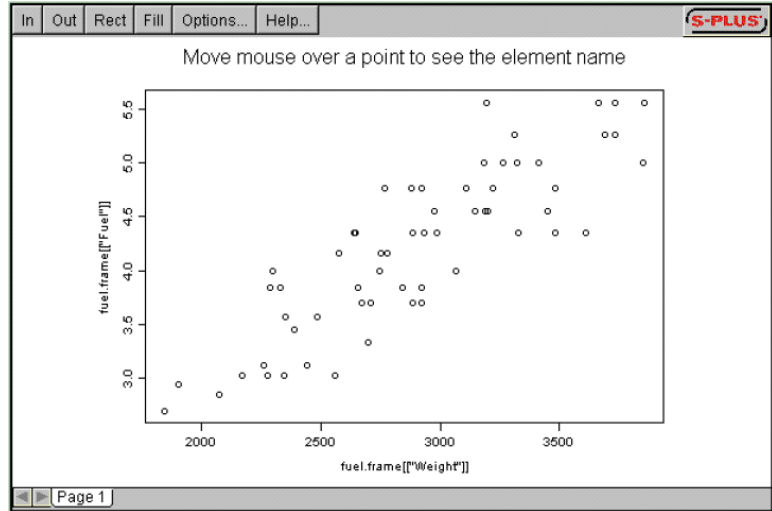


図 7.3 : ファイル **fuel.html** 上の Graphlet の例
Graphlet の上部にある各ボタンは、それぞれオプションの
アプレットパラメータに対応している

表 7.1 : SPJ グラフファイルの名前と Graphlet の上部に表示されるボタンの両方を制御するアプレットパラメータ

アプレットパラメータ	説明
spjgraph.filename	Graphlet がロードする SPJ グラフファイルの名前を指定します。デフォルトは graph.spj です。
spjgraph.mouse.position	マウスがグラフ上にあるとき、マウスの座標を表示するかどうかを指定します。デフォルトは OFF です。
spjgraph.mouse.position.checkbox	オプションボタンのダイアログ上にマウスの座標を表示する (Display mouse position) チェックボックスを表示するかどうかを指定します。ON に指定すると、Graphlet 上にマウス座標を表示させるかどうかを、エンドユーザが変更することができます。デフォルトは ON です。
spjgraph.active.regions	アクティブ領域を使用可能にするかどうかを指定します。ON にした場合は、領域がアウトライン表示され、マウスがアクティブ領域上に移動すると関連付けられたラベルが表示されます。OFF にした場合は、領域もラベルも表示されず、マウスをアクティブ領域上に移動しても関連するアクションは実行されません。デフォルトは ON です。
spjgraph.active.regions.checkbox	オプションボタンのダイアログ上に Active regions を有効にする (Enable active regions) チェックボックスを表示するかどうかを指定します。ON に指定すると、Graphlet 上のアクティブ領域を使用可能にするかどうかを、エンドユーザが制御することができます。デフォルトは ON です。

表 7.1 : SPJ グラフファイルの名前と Graphlet の上部に表示されるボタンの両方を制御するアプレットパラメータ (続き)

アプレットパラメータ	説明
spjgraph.rect.button	Graphlet の上部に 選択領域 (Rect) ボタンを表示するかどうかを指定します。デフォルトは ON です。
spjgraph.resize.buttons	Graphlet の上部に、いずれかのサイズ変更ボタン (ズームイン 、 ズームアウト 、 選択領域 、および 元に戻す で In 、 Out 、 Rec 、 Fill と表示されます) を表示するかどうかを指定します。指定しないと、エンドユーザは、グラフのサイズを変更できません。デフォルトは ON です。
spjgraph.options.button	オプション (Options) ボタンを表示するかどうかを指定します。デフォルトは ON です。
spjgraph.help.button	ヘルプ (Help) ボタンを表示するかどうかを指定します。デフォルトは ON です。
spjgraph.tabs	グラフの下にページタブを表示するかどうかを指定します。デフォルトは ON です。グラフのページが 1 つしかない場合、またはユーザがアクティブ領域の選択でしかページの切り替えを行えないようにする場合は、このオプションを外しておくとう便利です。
spjgraph.tabs.checkbox	オプション (Options) ボタンのダイアログに タブを表示する (Display page tabs) チェックボックスを表示するかどうかを指定します。ON に指定すると、エンドユーザが、Graphlet のグラフの下にページタブを表示させるかどうかを制御することができます。デフォルトは ON です。

例の詳細 (続き)

ここでは、引き続き「Graphlet データファイルの作成」(252 ページ)の例について、Graphlet のデータファイル **map.spj** を生成するのに必要なステップについて説明します。以下の HTML コードは、**map.spj** ファイルをウェブページに埋め込み、APPLET コマンド内の特定のオプションを設定するものです。

```
<HTML>
<HEAD>
<TITLE>Racial Distribution in US cities</TITLE>
</HEAD>
<BODY>
<APPLET code="spjgraph.class" archive="spjgraph.jar"
        WIDTH=600 HEIGHT=400 >
    <PARAM NAME=spjgraph.filename VALUE=map.spj>
    <PARAM NAME=spjgraph.mouse.position VALUE=ON>
    <PARAM NAME=spjgraph.options.button VALUE=OFF>
</APPLET>
</BODY>
</HTML>
```

この HTML コードを、**map.html** と名前をつけたファイルに保存し、**map.spj** および **spjgraph.jar** ファイルのあるディレクトリに置きます。その後、ウェブブラウザを使って Graphlet を表示させたり操作することができます。**\$SHOME/library/winspj** ディレクトリにあるサンプルファイル **map.html** は、上記で作成した HTML とは異なることに注意してください。見本ファイル **map.html** には、すべての PARAM 値のデフォルト値が維持されていますが、上記で作成したファイルでは、**spjgraph.mouse.position** と **spjgraph.options.button** の値が変更されています。

Graphlet の使い方

ウェブページに埋め込まれた S-PLUS Graphlet を見ているとします。ウィンドウには、上部にいくつかのラベル付きボタンがあり、その下にグラフを表示するための大きな領域があり、さらにウィンドウの下端に沿ってタブの列があります。タブの 1 つを左クリックすると、タブに対応するページのグラフが表示されます。たとえば、図 7.4 は、`map.html` ファイルの Graphlet の例を示します。この Graphlet は、米国内 47 都市の人種による人口分布を表しています。この Graphlet を見るときは、`$$SHOME/library/winspj` ディレクトリにある `map.html` ファイルを開きます。

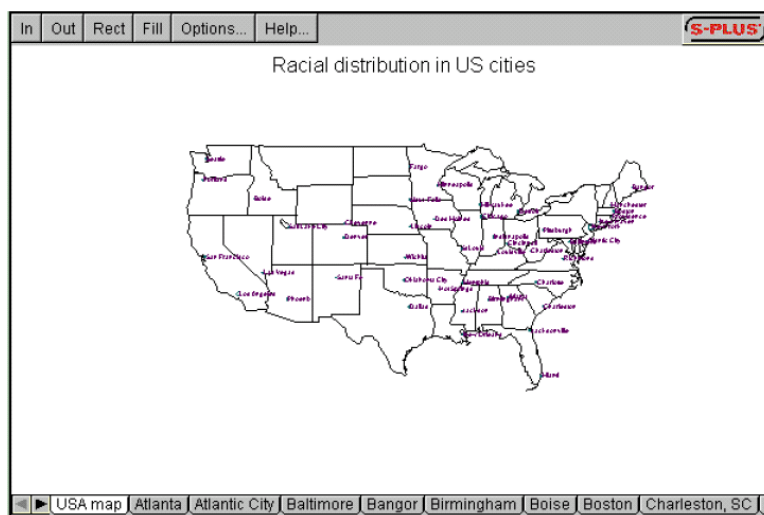


図 7.4 : `map.html` ファイルの Graphlet の例

Graphlet には、画像のサイズを変更するためのボタンがいくつかあります。**ズームイン (In)** ボタンは、画像を拡大し、**ズームアウト (Out)** ボタンは縮小します。**元に戻す (Fill)** ボタンは、ウィンドウいっぱいになるようにグラフのサイズを変更します。グラフが、ウィンドウよりも大きいサイズに変更された場合は、グラフのまわりにスクロールバーが現れます。

選択領域 (Rect) ボタンを使用すると、特定の領域を拡大することができます。グラフ・ウィンドウ内でマウスの左ボタンをクリックしてドラッグすると、四角形の境界ボックスが定義されます。そのような四角形を定義した後で、**選択領域 (Rect)** ボタンを押します。これにより、指定した長方形の領域がウィンドウいっぱいに表示されます。

オプション (Options) ボタンを押すと、図 7.5 に示すダイアログが表示されます。このダイアログを使用して、Graphlet のアクションを制御するいくつかのオプションを変更することができます。マウスの座標を表示する (Display mouse positions) チェックボックスで、マウスのグラフ上の移動にしたがってグラフ座標を表示するかどうかを指定します。このボックスは、最初はチェックされておらず、したがって、マウスの座標は表示されません。表示桁数テキストフィールド (Mouse position digits) で、マウス座標の精度を表示するとき使用する桁数を指定します。Active regions を有効にする (Enable active regions) チェックボックスは、アクティブ領域を使用可能にするかどうかを指定します。このボックスは、最初にチェックされているため、アクティブ領域が使用可能になっています。タブを表示する (Display page tabs) チェックボックスで、グラフの下にページタブを表示するかどうかを指定します。デフォルトでは、このボックスはチェックされており、ページタブが表示されます。グラフのページが 1 つしかない場合は、このオプションを外しておくとも便利です。

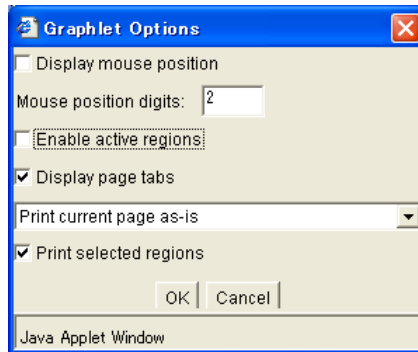


図 7.5 : S-PLUS Graphlet の Options ボタンを押したときに表示されるダイアログ

S-PLUS Graphlet 上のヘルプ (Help) ボタンを押すと、Graphlet の使い方に関する簡単な説明を含むウィンドウが現れます。右上角にある S-PLUS を押すと、S-PLUS Graphlet について説明しているウェブページが表示されます。

注意

上記で説明したすべてのボタンが Graphlet に表示されるとは限りません。「ウェブページへの Graphlet の埋め込み」(263 ページ) で説明したように、表示するボタンを指定するアプレットパラメータを、ウェブページ上で設定することができます。これで、グラフの作成者は、たとえばグラフのサイズ変更ボタンをエンドユーザが使用できないようにすることもできます。

map.spj Graphlet を作成するために使用するデータを“ドリルダウン”するには、米国の地図内の点上にマウスを移動させます。特定の都市の上にマウスを移動させると、アクティブ領域がハイライトされ、その都市の名前が Graphlet の右上角に表示されることに注意してください。画面に表示される都市のラベルが小さすぎる場合は、**選択領域 (Rect)** ボタンを使って地図内の特定の領域を拡大してください。

地図内のアクティブ領域の 1 つをクリックすると、その領域に対応する都市の人種データの棒グラフを示す別の Graphlet ページにジャンプします。たとえば、米国の地図内のサンフランシスコのアクティブ領域をクリックします。すると、Graphlet は、図 7.6 に示すページにジャンプします。あるいは、Graphlet の左下角にある矢印を使ってページタブの間を移動することもできます。また、“San Francisco”というタイトルのタブをクリックしても、図 7.6 に示す棒グラフを表示させることができます。

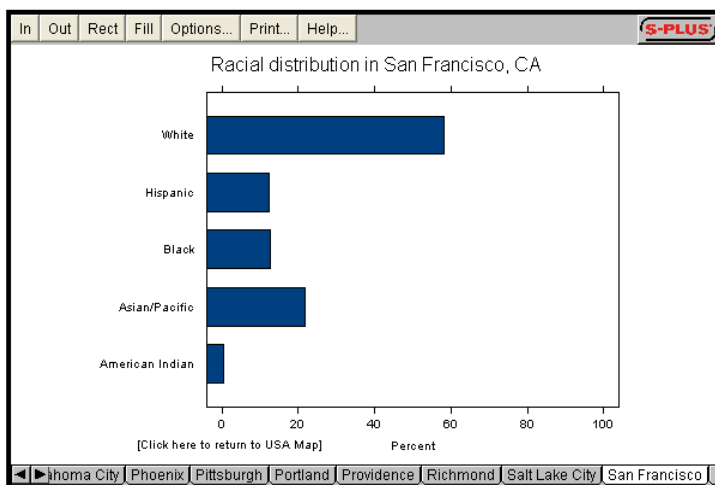


図 7.6 : map.spj Graphlet のサンフランシスコの人種データの棒グラフ

米国の地図に戻るときは、棒グラフの下の部分にマウスを持って行きます。すると“Click here to return to USA map.”という文字を囲むアクティブ領域がハイライトされます。このアクティブ領域をクリックすることにより、グラフの最初のページに戻り、そこで、各都市の人種データの棒グラフを選択することができます。

第 8 章

統計

はじめに	276
概要	276
基本手順	278
ダイアログ	278
ダイアログのフィールド	279
統計ダイアログからプロットする	280
解析による結果を保存する	281
要約統計	282
要約統計	282
クロス表	285
相関	287
テーブル	289
標本比較	291
1 標本検定	291
2 標本検定	301
k 標本検定	311
仮説検定	320
検出力と標本数	335
平均	335
二項比率	337
実験計画	340
要因計画	340
直交要因計画	341
要約統計量プロット	342
箱ひげ図	343
交互作用のチェック	344
回帰	346
線形回帰	347

ロバスト MM 回帰	356
ロバスト LTS 回帰	357
ステップワイズ線形回帰	359
一般化加法モデル	361
局所回帰	362
非線形回帰	364
一般化線形モデル	368
対数線形 (ポアソン) 回帰	370
ロジスティック回帰	371
プロビット回帰	373
分散分析	376
固定効果 ANOVA	376
ランダム効果 ANOVA	377
多重比較	379
混合効果モデル	382
線形	382
非線形	383
一般化最小二乗モデル	386
線形	386
非線形	387
生存時間解析モデル	389
ノンパラメトリック	389
コックス比例ハザードモデル	390
パラメトリック	392
打ち切りのあるデータの解析	394
樹形モデル	397
樹形モデル	397
樹形モデルのチェック	398
モデルの比較	401
クラスター分析	404
非類似度の算出	404
κ -means 型クラスタリング	405
pam アルゴリズムによるパーティショニング	406
ファジィ・パーティショニング	408
擬集型階層クラスタリング	410
区分型階層クラスタリング	411
系統分類クラスタリング	413

多変量解析	415
判別分析	415
因子分析	416
主成分分析	417
多変量分散分析	419
QC チャート (品質管理チャート)	421
グループのあるデータ	421
グループのないデータ	422
P 管理図	424
標本からのサンプリング	426
ブートストラップ推定	426
ジャックナイフ推定	428
平滑化	430
核関数による平滑化	431
局所回帰 (Loess)	431
スプライン平滑化	431
スーパースムーザー	432
時系列解析	434
自己相関	434
ARIMA	436
ラグプロット	438
スペクトル	439
乱数と分布	441
ランダムサンプリング	441
分布関数	442
乱数	447
参考資料	451

はじめに

S-PLUS の特長の 1 つは、統計解析機能にグラフィックス機能を統合していることです。このマニュアルの他の章では、S-PLUS のグラフィックスを紹介しています。この章では S-PLUS で行われる統計処理について説明します。

統計解析を実行する前にこの章全体を読む必要はありません。統計処理を実行する基本的な方法を理解すれば、関心のある節を直接参照することができます。

この章では最初に統計ダイアログの使い方について一般的な情報を説明し、残りの節でそれぞれのダイアログについて説明し、例を取りあげます。

概要

図 8.1 は、S-PLUS インタフェースの多くの要素を示しています。

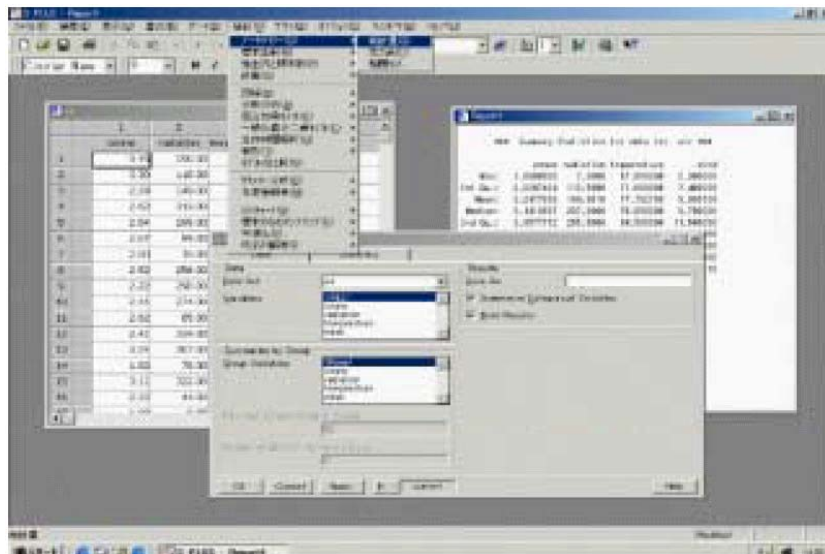


図 8.1 : 統計に関連するメニューとウィンドウ

- **統計メニュー**：統計メニューから、S-PLUS で使用することができる統計処理のほとんどを行うことができます。処理方法は、目的に応じたサブメニューによって論理的に分類されています。例として図 8.1 に、統計量のメニューツリーを示します。これは、**統計 ▶ データサマリー ▶ 統計量**を選択します。
- **統計ダイアログ**：図 8.1 に示した **Summary Statistics** と名前のついたオープンダイアログは、計算するデータなどを指定するために使用されます。
- **データメニュー**（図示せず）：統計メニューの左側にある**データ**メニューは、データを生成し処理する機能を実行するために用いられます。
- **データ・ウィンドウ**：図 8.1 の左側に開いているウィンドウは、**データ・ウィンドウ**です。**データ・ウィンドウ**を使用して、データセットを作成したり編集したりすることができます。**データ・ウィンドウ**の詳しい説明は、第 2 章「データによる作業」を参照してください。
- **レポート・ウィンドウ**：**レポート・ウィンドウ**は、統計解析の結果を表示するウィンドウです。図 8.1 の例では、**レポート・ウィンドウ**は要約統計の結果を示しています。
- **グラフシート**（図示せず）：**グラフシート**は、**統計**メニューから作成したグラフを表示します。
- **メッセージ・ウィンドウ**（図示せず）：**メッセージ・ウィンドウ**は、統計処理によってエラー、警告または情報メッセージが生じた場合だけ現れます。**メッセージ・ウィンドウ**を閉じると、次に情報メッセージが生成されたときに再び現れます。統計解析中に大きな障害が生じないように、**メッセージ・ウィンドウ**の内容を常に確認するようにしてください。

基本手順

データを分析する基本手順は、解析のタイプに関係なく同じです。

1. 作業をしたいデータを選択します。
2. **統計**メニューから実行したい統計処理（要約統計、線形回帰、ANOVA など）を選択します。その処理に対応するダイアログが開きます。
3. 選択した処理に対し、データセット、変量および様々なオプションを選択します。（それぞれのダイアログにはわずかな違いがあります。）
OK または **Apply** ボタンをクリックして解析を開始します。**OK** をクリックすると、統計解析が実行されたときにダイアログが閉じます。**Apply** をクリックすると、ダイアログが開いたままになります。
4. メッセージを確認してください。メッセージが生成されると、**メッセージ・ウィンドウ**に表示されます。**メッセージ・ウィンドウ**が開かない場合、メッセージは生成されていません。
5. 結果を確認してください。処理が問題なく完了すると、解析の結果が**レポート・ウィンドウ**に表示されます。また、統計処理によっては、グラフを生成するものがあります。

必要に応じてダイアログの変量、パラメータまたはオプションを変更し、**Apply** をクリックすると、新しい結果を生成することができます。**S-PLUS** では、簡単にオプションを試したり解析手法の違いを試したりすることができます。

ダイアログ

S-PLUS の統計機能のほとんどは、**データ**メニューと**統計**メニューから利用することができます。

データメニューにはデータ・テーブルを作成し、分布関数を計算し、乱数を生成するダイアログが含まれています。**データ**メニューにはまた、データのソート／積み重ね／変換などを処理するダイアログが含まれています。詳細は、第2章の「データによる作業」を参照してください。

統計メニューにはデータサマリーを作成し、統計モデルをあてはめるダイアログが含まれています。ダイアログの多くはモデルのあてはめ、作図および予測を含む完全な解析を可能にするタブ付きページで構成されています。それぞれのダイアログには、関数の引数にダイアログ入力で指定した値を用い実行させる、対応する関数があります。通常はタブ付きダイアログの1ページのいくつかのフィールドに入力するだけで、関数の実行が可能です。

ダイアログのフィールド

多くのダイアログには、**Data Set** フィールドがあります。このフィールドには**ファイル ▶ 開く**で一番最後に開いたデータセットか、または**Select Data**ダイアログで選択したデータセットの名前が自動的に入力されています。別のデータセットを指定するときは、**Data Set** フィールドにその名前を直接入力するか、ドロップダウンリストから選択をします。ドロップダウンリストに表示されるデータセットは、**オブジェクト・エクスプローラ**によってフィルタリングされたものに限られています。

統計モデルをあてはめるほとんどのダイアログには、データセットの一部分だけを指定するために使用できる**Subset Rows** フィールドが含まれています。解析にデータのサブセットを使用するときは、使用する行を特定する**S-PLUS** 式を**Subset Rows** フィールドに入力してください。この式は論理値のベクトルとして評価され、真の値は、解析にどの行を含めるかを示し、偽の値は、どの行を外すかを示します。あるいは次のように、式で行を特定するベクトルを指定することができます。

- 式 `Species=="bear"` は、列 `Species` が `bear` を含む行だけを抽出します。
- 式 `Age>=13 & Age<20` は、変量 `Age` の十代の値に対応する行だけを抽出します。
- 式 `1:20` は、データの最初の 20 行を抽出します。

データセット内のすべての行を使用するときは、**Subset Rows** フィールドを空白のままにしてください。

いくつかのダイアログには、**Variables** フィールドがあります。このフィールドには、選択されているデータセットの列名が自動的に入力されています。

ダイアログによっては、データセットの列を選択した場合には自動的に入力されている**Formula** フィールドを必要とするものがあります。最初に選択される列が反応変量（目的変数）であり、残りの列が予測変量（説明変数）です。自動的に表示されたモデル式が希望のものではない場合は、別のモデル式を**Formula** フィールドに入力したり、**Create Formula** ボタンをクリックして、ダイアログを使って式を作成することができます。**Generalized Additive Models** ダイアログなどのいくつかのダイアログは、特別なモデル式を必要とします。このケースでは、使用可能な特別の項目が、**モデルビルダー**にリストされます。

ほとんどのダイアログには、解析結果を保存するオブジェクト名を指定する**Save As** フィールドがあります。このオブジェクトを**オブジェクト・エクスプローラ**から処理することによって、モデリングをあてはめた後でも追加のサマリーやプロットを得ることができます。また、ダイアログの多くには、1 つまたは複数の**Save In** フィールドがあります。**Save In** フィールドは、新しい列を保存するデータセットの名前に対応しています。たとえば、あてはめ

統計 ダイアログ からプロット する

値／残差／予測／標準誤差の値が保存されるためのものがあります。

ほとんどの統計ダイアログは、解析に役立つデフォルトのグラフを生成します。その多くは通常、個別の **Plot** タブにいくつかのプロットオプションとして選択可能です。

デフォルトでは、統計ダイアログから作成されたプロットは編集できません。好みにより、統計ダイアログのデフォルトの挙動を、通常のグラフィックスから編集可能なグラフィックスに変更することができます。

1. メインメニューから **オプション ▶ グラフオプション** を選択します。
図 8.2 に示すように、**Graphs** ダイアログが開きます。
2. **Statistics Dialogs Graphics** グループの **Create Editable Graphics** オプションをチェックします。
3. **OK** をクリックします。

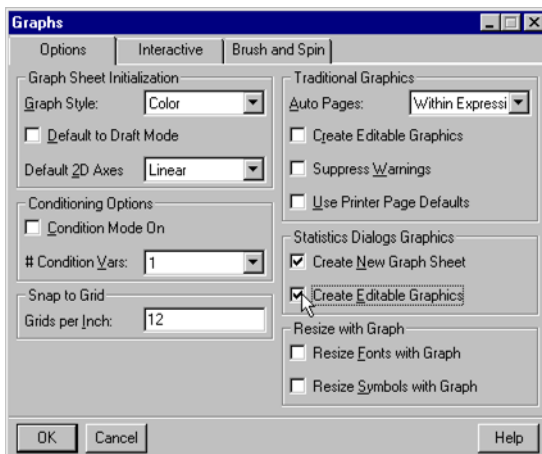


図 8.2 : Graph ダイアログの Options ページ

動作を速くするため、マルチパネルの Trellis 表示のような複雑なプロットは、通常のグラフィックスで作成されることをお勧めします。編集可能でないグラフを編集可能なグラフに変換するときは、グラフの点を右クリックし、コンテキストメニューから **Convert to Objects** を選択します。

解析による 結果を 保存する

統計モデル結果のオブジェクトは、ダイアログの **Save As** フィールドでオブジェクト名を指定することにより作成することができます。実行が完了すると、**オブジェクト・エクスプローラ**にオブジェクトが現われます。オブジェクトをダブルクリックすると結果が**レポート・ウィンドウ**に表示され、オブジェクトのサマリーが表示されます。線形回帰による結果といったモデルオブジェクトの場合は、右クリックのコンテキストメニューを使用することができます。関連メニューを表示させるときは、**オブジェクト・エクスプローラ**内のモデルオブジェクトを右クリックしてください。ほとんどのメニューの選択肢は、ダイアログからのタブ付きページに対応しています。これによりダイアログ全体を再度呼び出すことなくモデルのプロットと予測を実行することができます。線形モデルオブジェクトの例を、図 8.3 に示します。

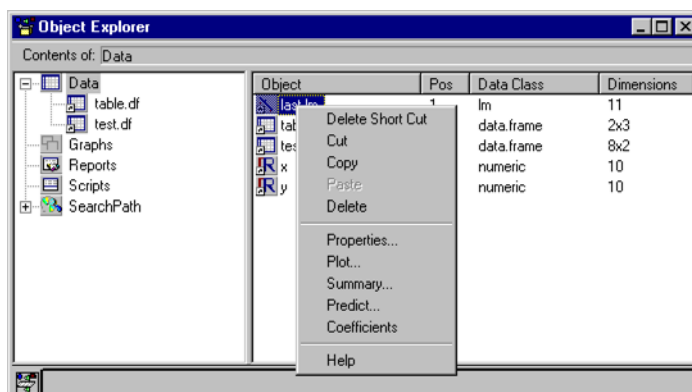


図 8.3 : 線形モデルオブジェクトに関して示された
右クリックのコンテキストメニュー

要約統計

データ解析の最初のステップの 1 つは、サマリーを作成することです。これは **Summary Statistics**、**Crosstabulations**、および **Correlations and Covariances** ダイアログによって数值的に行うことができます。

- **Summary statistics (統計量)** : 平均／中央値／分散／総和／四分位値などの要約統計を計算します。
- **Crosstabulations (クロス表)** : 変量間の因子の組み合わせごとに発生件数の表を作成し、その表の統計量を生成します。
- **Correlations and Covariances (相関)** : 変量間の相関または共分散を計算します。

これらの 3 つの処理は、**統計 ▶ データサマリー**メニューから選択することができます。さらに、**データの Tabulate** ダイアログは、データのテーブルサマリーを出力し、**Trellis** グラフに使用するのに便利なデータセットを作成します。

要約統計

Summary Statistics ダイアログは、連続的な変量に対しては基本的な 1 変量要約統計量を、カテゴリカル変量に対しては度数を提供します。サマリーは、1 つまたは複数のグループに分けた変量に基づいてグループごとに計算することができます。

統計量を計算する

メインメニューから、**統計 ▶ データサマリー ▶ 統計量**を選択します。図 8.4 に示すように、**Summary Statistics** ダイアログが開きます。

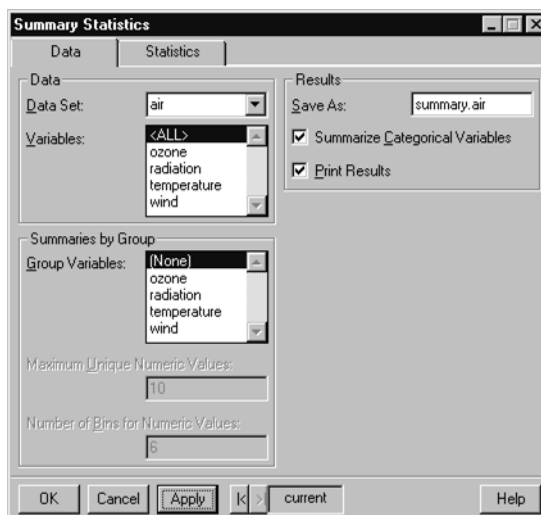


図 8.4 : Summary Statistics ダイアログ

例

データセット `air` を使用します。このデータセットは、ニューヨークで 111 日間にわたって、オゾン濃度／風速／気温／日射量を測定したものです。この例では、これらのデータの要約統計を計算します。


1. **Summary Statistics** ダイアログを開きます。
2. **Data Set** フィールドに `air` と入力します（ユーザーのデータではないのでリストボックスに `air` という名前は表示されません）。
3. **Variables** フィールドで要約統計したい変数を選択します。この例では、**<ALL>**（デフォルト）を選択します。
4. レポート・ウィンドウで結果を表示し、モデルオブジェクトに結果を保存したいときは、**Save As** フィールドに `summary.air` と入力し、保存する、この名前のオブジェクトを作成します。S-PLUS は、警告なしに、この名前の既存のデータに上書きします。結果がレポート・ウィンドウで表示されるように、**Print Results** チェックボックスが選択されていることを確かめてください。
5. 統計量を選択するときは、**Statistics** タブをクリックしてください。この例では、**Variance** と **Total Sum** チェックボックスを追加選択します。

6. **OK** をクリックします。レポート・ウィンドウに、次のような出力が作成されます。

```

*** Summary Statistics for data in: air ***

              ozone radiation temperature      wind
      Min:      1.00          7.00          57.00    2.30
      1st Qu.:   2.62         113.50          71.00    7.40
      Mean:     3.25         184.80          77.79    9.94
      Median:   3.14         207.00          79.00    9.70
      3rd Qu.:  3.96         255.50          84.50   11.50
      Max:     5.52         334.00          97.00   20.70
      Total N: 111.00        111.00          111.00  111.00
      NA's:     0.00          0.00          0.00    0.00
      Variance: 0.79        8308.74          90.82   12.67
      Std Dev.: 0.89         91.15          9.53    3.56
      Sum:     360.50       20513.00          8635.00 1103.20
    
```

7. 上記の出力がレポート・ウィンドウに表示されない場合は、メッセージ・ウィンドウにエラーメッセージがないか確認し、レポートされている問題を解決してください。
8. 結果が保存されたモデルオブジェクトを呼び出すときは、**オブジェクト・エクスプローラ**を使用します。**オブジェクト・エクスプローラ**を開くときは、**標準**ツールバーの**オブジェクト・エクスプローラ**ボタン  をクリックしてください。
9. **Data** フォルダをハイライトします。summary.air という名前 (上のステップ 4 で指定した名前) の新しいデータ行列をダブルクリックし、**データ・ウィンドウ**に要約統計をロードします。

これで完了です。要約統計の計算が簡単なことがお分かりでしょう。他の統計処理も上記と同じ基本的なステップを使用します。

クロス表

Crosstabulations ダイアログは、指定されたカテゴリカル（因子）変量のすべての組み合わせの度数表を作成します。さらに、セルのパーセント値を計算し、独立性に関するカイ二乗検定を行います。**Crosstabulations** ダイアログは、その結果を ASCII 形式の表で返します。

独立性に関するカイ二乗検定は、データがカテゴリカル共変量の様々な組み合わせ結果の出現回数から成るときに役立ちます。これは出現回数が共変量の周辺値によるものか、共変量間の交互作用による影響を受けているのかを判断するために使用されます。

さらに詳しい解析に使用する度数表を作成するときは、**データ**メニューから **Tabulate** ダイアログを使用してください。

クロス表を計算する

メインメニューから、**統計** ▶ **データサマリー** ▶ **クロス表** を選択します。図 8.5 に示すように、**Crosstabulations** ダイアログが開きます。

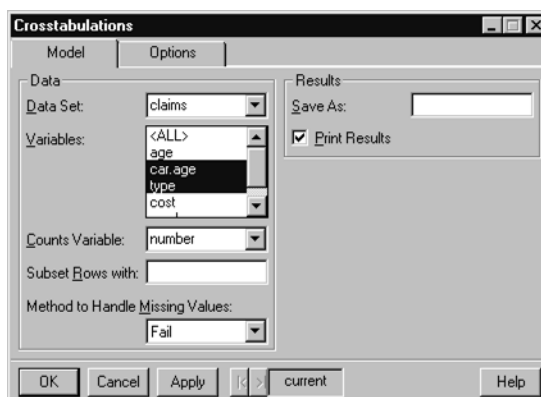


図 8.5 : Crosstabulations ダイアログ

例

成分 age/car.age/type/cost/number を含むデータセット claims を検討します。元データは、8,942 件の保険金請求から得たものです。データセット claims の 128 の行は、3 つの予測変量としての変数（列）age、car.age および type のすべての組み合わせを表します。さらに追加の変数 number は、各セルにおける請求件数を示します。変数 cost は請求の平均金額です。

分割表を使用して、車の使用年数と車種による請求件数の分布を調べることができます。独立性の検定によって、請求発生 の 尤度 が 使用年数 と 車種 の 組み合わせによって変化するかどうか、あるいは車の使用年数と車種が無関係かどうか分かります。

claims データの分割表を作成するには、

1. **Crosstabulations** ダイアログを開きます。
2. **Data Set** フィールドに claims と入力します。
3. **Variables** フィールドの car.age をクリックし、次に type を CTRL-クリックします。これで、両方の変数が解析用に選択されます。
4. **Counts Variable** フィールドで、変数のリストをスクロールし、number を選択します。
5. **OK** をクリックします。

レポート・ウィンドウに、以下のような表が現れます。

```

*** Crosstabulations ***
Call:
crosstabs(formula = number ~ car.age + type, data =
  claims, na.action = na.fail, drop.unused.levels = T)
8942 cases in table
+-----+
|N      |
|N/RowTotal|
|N/ColTotal|
|N/Total  |
+-----+
car.age|type
      |A      |B      |C      |D      |RowTotal|
-----+-----+-----+-----+-----+
0-3   | 391   |1538   |1517   | 688   |4134   |
      |0.0946|0.3720 |0.3670 |0.1664 |0.462  |
      |0.3081|0.3956 |0.5598 |0.6400 |      |
      |0.0437|0.1720 |0.1696 |0.0769 |      |
-----+-----+-----+-----+-----+
4-7   | 538   |1746   | 941   | 324   |3549   |

```

	0.1516	0.4920	0.2651	0.0913	0.397
	0.4240	0.4491	0.3472	0.3014	
	0.0602	0.1953	0.1052	0.0362	
8-9	187	400	191	44	822
	0.2275	0.4866	0.2324	0.0535	0.092
	0.1474	0.1029	0.0705	0.0409	
	0.0209	0.0447	0.0214	0.0049	
10+	153	204	61	19	437
	0.3501	0.4668	0.1396	0.0435	0.049
	0.1206	0.0525	0.0225	0.0177	
	0.0171	0.0228	0.0068	0.0021	
ColTotl	1269	3888	2710	1075	8942
	0.14	0.43	0.30	0.12	

Test for independence of all factors
Chi^2 = 588.2952 d.f.= 9 (p=0)
Yates' correction not used

表の各セルには車の使用年数と車種の組み合わせに対する請求件数と共に、セル内の観測値の列に対するパーセント値、行に対するパーセント値および合計に対するパーセント値が表示されます。独立性の検定の結果は、各セル内の観測値のパーセント値が、列合計のパーセント値と行合計のパーセント値の積と大きく異なっていることを示します。したがって車の使用年数と車種との間には交互作用があり、これは請求件数に影響しています。すなわち車の使用年数が請求件数に及ぼす影響は、車種によって異なります

Crosstabulations ダイアログは、行と列のパーセント値を調べ独立性の検定を行うのに最も有効です。簡単なデータの表を作成するときは、289 ページで説明するような **Tabulate** ダイアログを使用してください。

相関

Correlations and Covariances ダイアログは、相関と共分散の基本的な 2 変量のサマリーを生成します。

相関と共分散を計算する

メインメニューから、統計 ▶ データサマリー ▶ 相関を選択します。図 8.6 に示すように、**Correlations and Covariances** ダイアログが開きます。

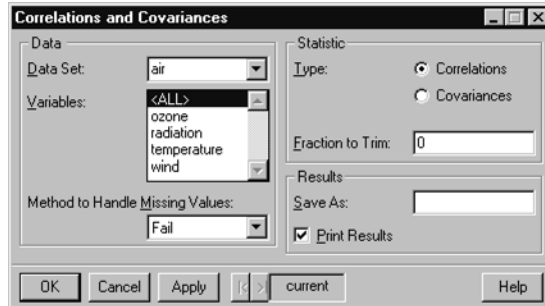


図 8.6 : **Correlations and Covariances** ダイアログ

例

282 ページの要約統計では、データセット `air` の 1 変量サマリーについて調べました。次に、データセットの 4 つすべての変量間の相関を生成します。基本ステップは、次の通りです。

1. **Correlations and Covariances** ダイアログを開きます。
2. **Data Set** フィールドに `air` と入力します。
3. **Variables** フィールドで **<ALL>** を選択します。
4. **OK** をクリックします。

レポート・ウィンドウに、以下の 4 つの変量間の相関が表示されます。

```
*** Correlations for data in: air ***
           ozone   radiation temperature   wind
ozone  1.000000   0.4220130   0.7531038  -0.5989278
radiation 0.4220130   1.0000000   0.2940876  -0.1273656
temperature 0.7531038   0.2940876   1.0000000  -0.4971459
wind -0.5989278  -0.1273656  -0.4971459   1.0000000
```

`ozone` と `temperature` の間の 0.75 という強い相関に注目してください。気温が高いほどオゾンの濃度も増大します。`ozone` と `wind` の間の -0.60 の負の相関は、風速が早くなるほどオゾンの濃度が減少することを示しています。最後に `wind` と `temperature` の間の -0.50 の相関は、風速が早くなるほど気温が低くなること（または、風が弱くなるほど気温が高くなること）を示しています。

テーブル

Tabulate ダイアログは、データセットからデータのテーブルサマリーを作成します。

テーブルを作成する

メインメニューから、**データ ▶ テーブル**を選択します。図 8.7 に示すように、**Tabulate** ダイアログが開きます。

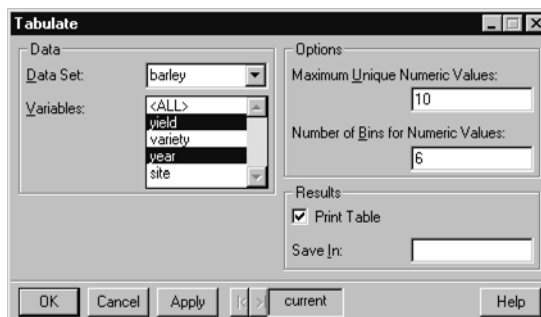


図 8.7 : Tabulate ダイアログ

例

barley データセットは、大麦の収穫高を研究した 1930 年代の農業産地試験による観測値を含んでいます。ミネソタ州の 6 つの生産地で、1931 年と 1932 年にそれぞれ 10 種類の大麦を育てました。データは生産地、品種および年のすべての組合せによる収穫高であるため、120 (6×10×2) の観測値となります。第 3 章「データの探索」では、このデータを Trellis グラフで調べました。この例では、様々な方法でデータのテーブルを作成します。

barley データのテーブルを作成するには、

1. **Tabulate** ダイアログを開きます。
2. **Data Set** フィールドに barley と入力します。
3. 変量 yield をクリックし、次に CTRL-クリックして year を選択します (選択の順序は重要です)。
4. **OK** をクリックします。

レポート・ウィンドウに以下のテーブルが現われます。

*** Table of yield,year in barley ***

	1932	1931
13.92000+ thru 22.64667	13	2
22.64667+ thru 31.37334	18	20
31.37334+ thru 40.10002	17	17
40.10002+ thru 48.82669	10	13
48.82669+ thru 57.55336	1	4
57.55336+ thru 66.28003	1	4

収穫高は、6つの“ビン（固まり）”に分割されています。収穫高が1931年には1932年よりもどのように高くなる傾向があるかに注目してください。

Tabulate ダイアログで複数の変数を選択することができます。たとえば、さらに表を大麦の10品種に分割するために、yield、year および variety を（この順序で）選択してみてください。

標本比較

1 標本検定

S-PLUS は、単一母集団に関する仮説の検定を行うための様々な検定をサポートしています。これらのほとんどの検定は、仮定の値に関するパラメータの検定となります。すなわち、帰無仮説は、 $H_0: \Theta = \Theta_0$ の形です。ここで Θ は対象となるパラメータであり、 Θ_0 はパラメータの仮説の値です。

- **1 標本 t 検定**：母平均 μ の検定。母平均が特定の値であるかどうかを検定します。小さいデータの場合は、母集団が正規分布でなければなりません。
- **1 標本ウィルコクソンの符号付き順位検定**：母平均 μ のノンパラメトリック検定。t 検定と同じように、母平均が特定の値であるかどうか検定しますが、分布の仮定を行いません。
- **1 標本コルモゴロフ=スミルノフ適合度検定**：データが仮説の分布とみなしてよいかどうかを判定する検定。連続した変量に好ましい適合度検定です。
- **1 標本カイ二乗適合度検定**：データが仮定した分布とみなしてよいかを確かめる検定。離散変量に好ましい適合度検定。

1 標本 t 検定

「1 標本 t 検定」は、変量の平均が特定値となるかどうかを検定するために使用されます。t 検定での主な仮定は、データがガウス（正規）分布によるものであることです。そうでない場合は、ウィルコクソンの符号付き順位検定のようなノンパラメトリック検定のほうが、位置（平均）の検定に適していることがあります。

1 標本 t 検定を実行する

メインメニューから統計 ▶ 標本比較 ▶ 1 標本 ▶ t 検定を選択します。図 8.8 に示すように、**One-sample t Test** ダイアログが開きます。

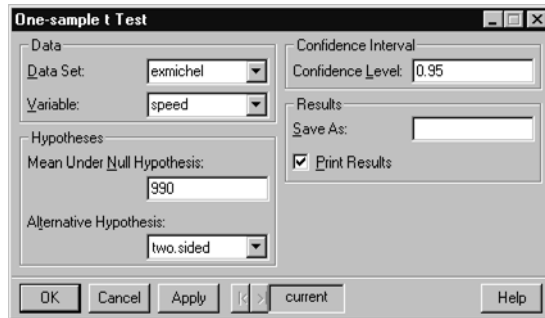


図 8.8 : One-sample t Test ダイアログ

例

1876 年に、フランスの物理学者コルニュが、光の速さ c について 299,990km/秒の値を報告しました。1879 年には、アメリカの物理学者 A.A. マイケルソンが、いくつかの実験を行ってコルニュの値を検証し精度を高めました。マイケルソンは、光の速さに関して次のような 20 の測定値を得ました。

850 740 900 1070 930 850 950 980 980 880
1000 980 930 650 760 810 1000 1000 960 960

マイケルソンの実測値を得るためには、上記の各値に 299,000km/秒を加えます。第 3 章「データの探索」で、マイケルソンデータを含む exmichel データセットを作成しました。

この 20 の観測値は、共通で未知の平均値 μ を持つ 20 の確率変量の観測値と見なすことができます。光の速さを測定する実験環境に偏りがないとすると、 μ が真の光の速さであると仮定するのが適切です。これらのデータを評価する際、少なくとも 4 つの質問に対する答えを検討します。

1. 光の速さ μ はどれだけか。
2. 最良と思われる値 $\mu_0 = 299,990\text{km/秒}$ に対して光の速さが変化したか。
3. 1 と 2 に対する答えに関連する不確実性は何か。
4. データの分布はどんな形か。

最初の 3 つの質問は、マイケルソンがデータを収集したときおそらく考慮されていたでしょう。データから有効な統計的推定を得るために使用できる方法を決定するには、最後の 2 つの質問に答えなければなりません。たとえば分布の形がほとんど外れ値のない正規分布を示す場合は、2 の質問に答えようとする際にスチューデント t 検定を使用することができます。データが外れ値を含むか正規分布からかけ離れている場合は、ロバスト法あるいはウィルコクソンの符号付き順位検定のようなノンパラメトリック法を使用しなければなりません。この節では、**S-PLUS** を使用してマイケルソンデータを詳しく解析します。同じ方法で他の 1 標本データを探索し、解析することができます。

探索的データ解析

マイケルソンデータを視覚的に探索するために、箱型図、ヒストグラム、確率分布、および正規 QQ プロットを作成します。第 3 章では、マイケルソンデータを解析するために図 8.9 に示すグラフと類似のグラフを作成しました。

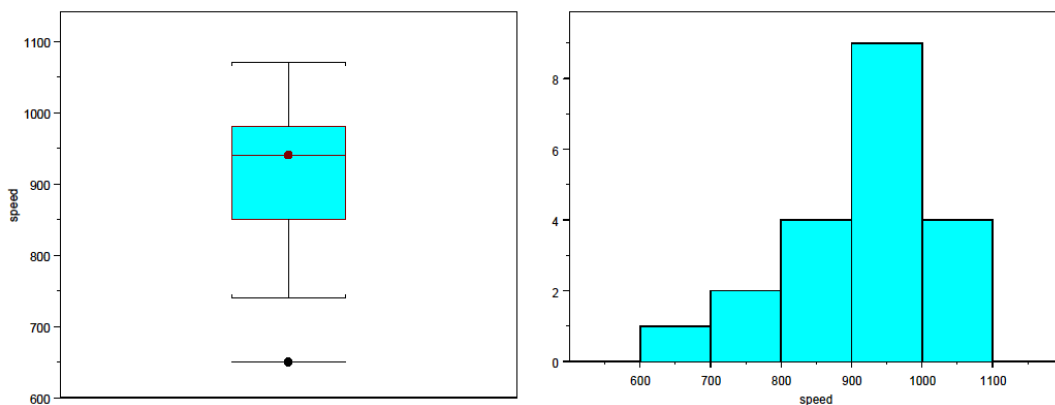


図 8.9 : マイケルソンデータの探索的データ解析のプロット

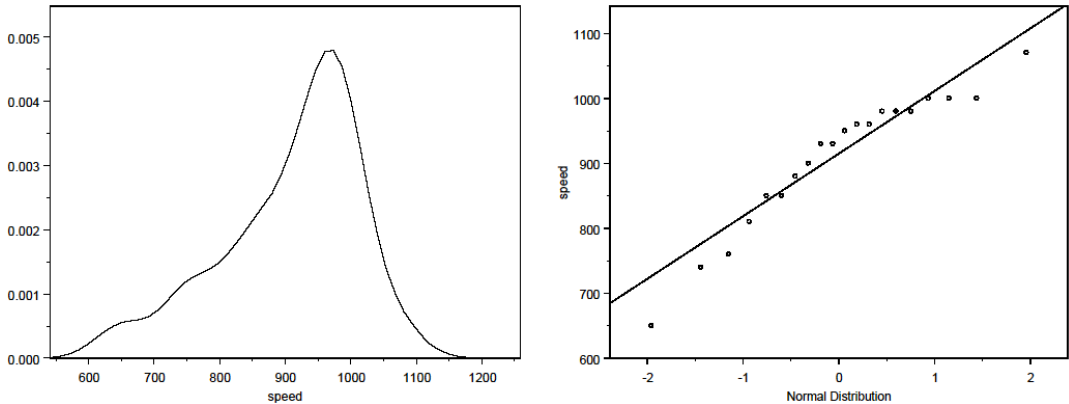


図 8.9 : マイケルソンデータの探索的データ解析のプロット (続き)

データが正規分布しているかどうかを確認するために、分布の形を評価したいと思います。これらのプロットは、左（すなわち小さい値の方）の裾が広い明らかに非対称の分布を示しています。分布は正規ではなく、“ほぼ”正規でもないようです。したがって、t 検定は標本が少ない場合に正規性を必要とするため、統計的推測にスチューデント t 検定を使用しない方がよいでしょう。

箱型図の横の実線は、データの「中央値」の位置で、箱の上端と下端はそれぞれデータの「第 3 四分位値」と「第 2 四分位値」の位置にあります。中央値と四分位値の正確な値を知りたいときは、**Summary Statistics** ダイアログを使用してください。

1. **Summary Statistics** ダイアログを開きます。
2. **Dat Set** として `exmichel` を入力します。
3. **Statistics** タブをクリックし、**Mean**、**Minimum**、**First Quartile**、**Median**、**Third Quartile** および **Maximum** 以外のすべてのオプションを選択解除します。
4. **OK** をクリックします。出力がレポート・ウィンドウに現われます。

```
*** Summary Statistics for data in: exmichel ***
```

```

                speed
Min:           650.000
1st Qu.:       850.000
Mean:          909.000
Median:        940.000
3rd Qu.:       980.000
Max:           1070.000

```

このサマリーは、上から最小値／第 1 四分位値／平均値／中央値／第 3 四分位値／最大値を示しています。このサマリーから、「四分位範囲」 $IQR=3Q-1Q$ を計算することができます。四分位範囲は外れ値を識別するのに有効な基準を提供します。第 3 四分位値よりも上と第 1 四分位値よりも下の、 IQR の 1.5 倍以上の観測値はすべて疑わしい外れ値です。

統計的推測

マイケルソンデータはおそらく正規分布でないため、統計的推測にはスチューデント t 検定ではなくウィルコクソンの符号付き順位検定を使用した方がよいでしょう。ここでは説明のために、両方の検定を行います。

母平均値の位置パラメータ μ のスチューデント t の信頼区間を計算するために、**One-sample t Test** ダイアログを使用します。またこのダイアログは、パラメータ $\mu_0=299,990$ のスチューデント t 有意検定の p 値を計算します。

1. **One-sample t Test** ダイアログを開きます。
2. **Data Set** フィールドに `exmichel` と入力します。
3. **Variable** として `speed` を選択します。
4. 帰無仮説値 $\mu_0=990$ (プラス 299,000) の両側検定を検定し、95%の信頼区間を作成したいと思います。**Mean Under Null Hypothesis** として 990 を入力します。
5. **OK** をクリックします。

レポート・ウィンドウに、1 標本 t 検定の結果が現われます。

One-sample t-Test

```
data: speed in exmichel
t = -3.4524, df = 19, p-value = 0.0027
alternative hypothesis: true mean is not equal to 990
95 percent confidence interval:
 859.8931 958.1069
sample estimates:
mean of x
 909
```

マイケルソンデータの算出平均は 909、p 値は 0.0027 です。この p 値はたいへん有意です。光の速さ 299,909km/秒のマイケルソンの平均値が、299,990km/秒のコルニュの値と有意に差があることは明らかです。

S-PLUS は、p 値の他に t 値/自由度/標本平均/信頼区間などの有効な情報を返します。

1 標本ウィルコクソンの符号付き順位検定

「ウィルコクソンの符号付き順位検定」は、変量の中央値が特定の値をもつかどうかを検定するために使用されます。これは 1 標本 t 検定と違って、観測値がガウス（正規）分布によることを想定していません。

1 標本ウィルコクソンの符号付き順位検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ 1 標本 ▶ ウィルコクソンの符号付き順位検定を選択します。図 8.10 に示すように、**One-sample Wilcoxon Test** ダイアログが開きます。

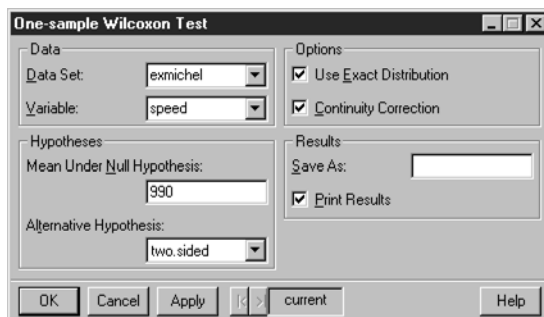


図 8.10 : One-sample Wilcoxon Test ダイアログ

例

291 ページの 1 標本 t 検定では、マイケルソンデータに t 検定を行いました。この検定により、光の速さ 299,909km/秒のマイケルソンの平均値が、299,990km/秒のコルニュの値と有意に差があるという結論に達しました。しかし、データが正規分布ではないために t 検定の結果に疑いがあることが分かりました。そこで次にウィルコクソンの符号付き順位検定を行って、2 つの光の速さの値が互いに有意に差があるか確認します。

1. **One-Sample Wilcoxon Test** ダイアログを開きます。
2. **Data Set** フィールドに `exmichel` と入力します。
3. **Variable** として `speed` を選択します。
4. **Mean Under Null Hypothesis** として `990` を入力します。
5. **OK** をクリックします。

レポート・ウィンドウに、次のように表示されます。

```
Wilcoxon signed-rank test
data: speed in exmichel
signed-rank normal statistic with correction Z = -3.0715,
p-value = 0.0021
alternative hypothesis: true mu is not equal to 990
```

変量 `Speed` に同じ値があるという警告メッセージが表示されることがあります。このメッセージは無視することができます。同じ帰無仮説を持ち、両側検定のために算出されたウィルコクソン検定の `p` 値 `0.0021` と `t` 検定の `p` 値 `0.0027` とはとても近い値になっています。結果としてウィルコクソンの符号付き順位検定により、光の速さのマイケルソンの平均値 `299,909km/秒` がコルニュの値 `299,990km/秒` と有意に差があることが確認されます。

コルモゴロフ＝ スミルノフ 適合度検定

「コルモゴロフ＝スミルノフ適合度検定」は、1組の観測値の経験分布が特定の理論分布から取り出したランダムサンプルと一致しているかどうかを検定するために使用されます。コルモゴロフ＝スミルノフ適合度検定は、連続変量にはカイ二乗適合度検定よりも一般的に有効です。離散変量には、一般にカイ二乗検定の方が適しています。

理論分布のパラメータ値が推定できない場合は、パラメータ値を正規（ガウス）分布または指数分布に対する検定の一部として観測値から自動的に推定することができます。他の分布で、パラメータを推定しなければならない場合には、カイ二乗検定を使用しなければなりません。その場合パラメータは検定とは別にデータから推定され、ダイアログに入力されます。

1 標本コルモゴロフ＝スミルノフ適合度検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ 1 標本 ▶ コルモゴロフ＝スミルノフ適合度検定を選択します。図 8.11 に示すように、**One-sample Kolmogorov-Smirnov Goodness-of-Fit Test** ダイアログが開きます。

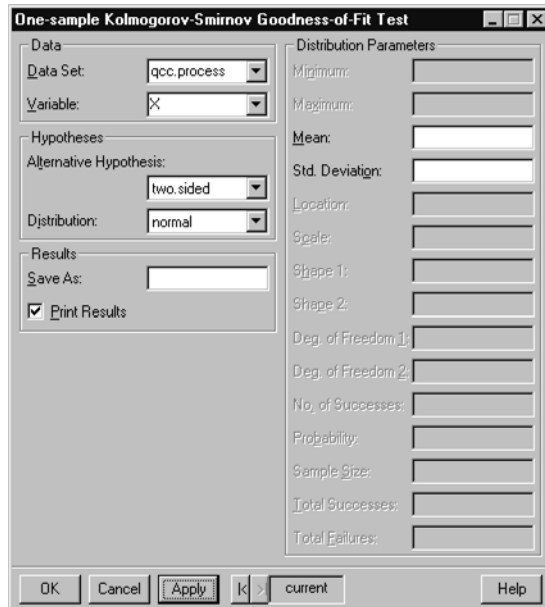



図 8.11 : One-sample Kolmogorov-Smirnov Goodness-of-Fit Test
ダイアログ

例

シミュレーションにより作成した 200 個の測定値からなる `qcc.process` と呼ばれるデータセットを作成します。合計 20 日間にわたり、1 日 10 個ずつ実測値を取得しました。**Random Numbers** ダイアログを使って、ガウス分布からデータセットを生成します。このダイアログの詳細は、441 ページの「乱数と分布」を参照してください。

1. 標準ツールバーの**新規データセット**  ボタンをクリックして空のデータセットを開きます。

2. メインメニューから、**データ ▶ 乱数**を選択します。**Data Set** フィールドに空のデータセットの名前があることを確認してください。デフォルトで、**Distribution** は **normal** です。**Mean** に 10 を入力し、**Std. Deviation** を 1 にしておきます。
3. **Target Column** に **x** を入力し、**Sample Size** に 200 を入力します。再現性のため、乱数ジェネレータのシードに設定します。**Set Seed with** フィールドに 21 と入力します。**OK** をクリックします。このステップにより、ガウス分布からランダムにサンプリングされた 200 個の要素を含む **x** と名付けられた列が作成されます。
4. メインメニューから、**データ ▶ 埋める**を選択します。**Data Set** フィールドにデータセットの名前があることを確認してください。**Columns** に **Day**、**Length** として 20、**Replications** の数として 10 を入力します。**Content** リストから **Grouped Sequence** を選択し、**OK** をクリックします。このステップにより、それぞれ 10 回繰り返された 1 から 20 の整数を含む **Day** という名前の列が作成されます。列 **Day** は、シミュレートされた実測値を取得した日を表します。
5. **データ・ウィンドウ**の左上角をダブルクリックして、データセットの名前を変更します。現われたダイアログで、**Name** フィールドに **qcc.process** を入力し、**OK** をクリックします。

コルモゴロフ＝スミルノフ適合度検定を使って、**qcc.process** がガウス分布であることを確認することができます。

1. P.298 の方法で **One-Sample Kolmogorov-Smirnov Goodness-of-Fit Test** ダイアログを開きます。デフォルトで、**Distribution** は **normal** です。
2. **Data Set** として **qcc.process** を選択します。
3. **Variable** として **x** を選択します。
4. **OK** をクリックします。

適合度検定のサマリーが**レポート・ウィンドウ**に表示されます。**p** 値 0.5 は、データが正規分布しているという仮説を棄却できないことを示しています。サマリーには、分布の平均と標準偏差の推定も含まれています。

p 値がきわめて小さい (**p** 値 ≤ 0.1) 場合には、この検定に使用される Dallal-Wilkinson 近似が最も正確であることを示す警告が**メッセージ・ウィンドウ**に表示されます。実際に算出された **p** 値は 0.776 であり、サマリーでは 0.5 に修正され、帰無仮説が棄却されないことを示していますが、**p** 値の推定としてはあまり正確ではありません。

カイ二乗 適合度検定

「カイ二乗適合度検定」は、ピアソンのカイ二乗統計を使用して、1組の観測値の経験分布が特定の理論分布から取り出されたランダムサンプルと一致するかどうかを検定します。

カイ二乗検定は、連続変量、離散変量、およびそれらの組み合わせなどの任意のタイプの変量に利用することができます。想定した分布が離散的で、かつ標本サイズが大きい ($n > 50$) 場合は、有効な検定はカイ二乗だけです。さらにカイ二乗検定は、分布のパラメータを推定した状況に簡単に適合します。ただし、連続変量の場合は、データのグループ化によって情報が失われます。

想定した分布が連続的で、帰無仮説の棄却が期待される場合、コルモゴロフ＝スミルノフ適合度検定は、その帰無仮説を棄却する可能性がカイ二乗検定よりも高くなります。コルモゴロフ＝スミルノフ検定は、カイ二乗検定よりも効果が高く、連続分布には便利です。

ピアソンのカイ二乗検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ 1 標本 ▶ カイ二乗適合度検定を選択します。図 8.12 に示すように、One-sample Chi-Square Goodness-of-Fit Test ダイアログが開きます。

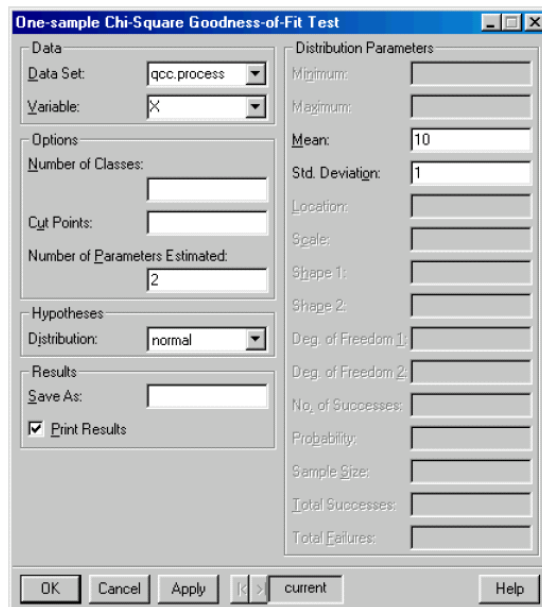


図 8.12 : One-sample Chi-Square Goodness-of-Fit Test ダイアログ

例

前の節で、200 個のシミュレーションによる計測値から成る `qcc.process` と呼ばれるデータセットを作成しました。合計 20 日間にわたり、1 日 10 個ずつ実測値を取得しました。カイ二乗適合度検定を使って、`qcc.process` がガウス分布をしていることを確認することができます。

1. `qcc.process` がまだ作成されていない場合は、298 ページに示した手順で `qcc.process` データを作成します。
2. P.300 の方法で **One-sample Chi-Square Goodness-of-Fit Test** ダイアログを開きます。デフォルトで、**Distribution** は **normal** です。
3. **Data Set** として `qcc.process` を選択します。
4. **Variable** として `x` を選択します。
5. カイ二乗検定では、分布の平均と標準偏差のパラメータ推定値を指定しなければなりません。**Mean** として 10、**Std.Deviation** として 1 を入力してください。データに適したパラメータ推定値が分からない場合は、**Summary Statistics** ダイアログを使ってそれを算出することができます。
6. データの平均と標準偏差を推定しているため、適合度検定を実行するときは、これらのパラメータ推定値を調整する必要があります。**Number of Parameters Estimated** として 2（推定が必要なパラメータ数）を入力してください。
7. **OK** をクリックします。

適合度検定のサマリーが、レポート・ウィンドウに表示されます。

2 標本検定

S-PLUS は、2 つの母数パラメータを比較するための様々な統計的検定をサポートしています。ここでは、 $H_0: \theta_1 = \theta_2$ という帰無仮説を検定します。ここで、 θ_1 と θ_2 とは 2 つの母数パラメータです。

- **2 標本 t 検定**：2 つの母平均 μ_1 と μ_2 を比較する検定。小さいデータでは、両方の母集団が正規分布していなければなりません。分散が等しくない 2 標本 t 検定や対のある 2 標本 t 検定などの 2 標本 t 検定の変形もサポートしています。
- **2 標本ウィルコクソン検定**：2 つの母平均 μ_1 および μ_2 を比較するためのノンパラメトリック検定。前に説明した t 検定と同じように 2 つの母平均 μ_1 と μ_2 を比較しますが、母集団に関する分布を仮定しません。符号付き順位と順位和の 2 つの形式の検定をサポートしています。
- **コルモゴロフ＝スミルノフ適合度検定**：2 つの標本が同じ分布によるものかどうかを決定する検定。

2 標本 t 検定

「2 標本 t 検定」は、2 つの標本が同じ平均値を持つ分布によるものかを検定するために使用されます。この検定は、対になった標本も独立した標本も扱うことができます。標本は、ガウス（正規）分布によるものと想定されます。そうでない場合は、ウィルコクソン順位和検定などのノンパラメトリック検定の方が位置（平均）の検定に適していることがあります。

2 標本 t 検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ 2 標本 ▶ t 検定を選択します。図 8.13 に示すように、Two-sample t Test ダイアログが開きます。

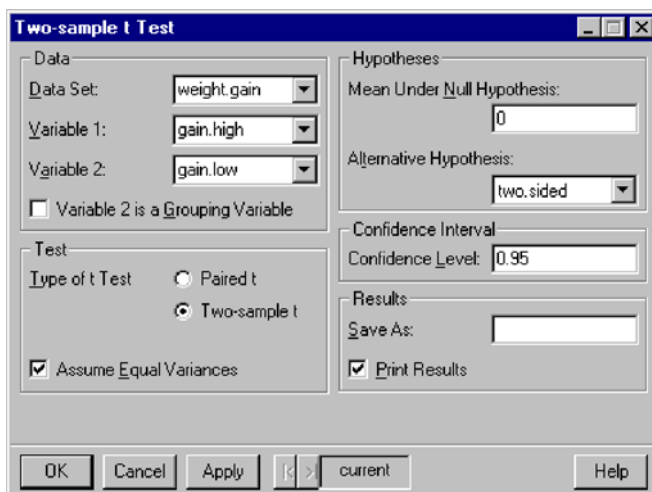


図 8.13 : Two-sample t Test ダイアログ

例

高タンパク質の食物と低タンパク質の食物の相対的な長所を研究している栄養学者を想定します。これらの 2 つの食物では、平均体重増加量に差がみられるでしょうか。そのような 2 つの食物を与えた 2 つのロットの雌ネズミの体重増加量（グラム）を示す、表 8.1 のデータを検討します。12 匹のネズミから成る第 1 のロットには高タンパク質の食物を与え、7 匹のネズミから成る第 2 のロットには低タンパク質の食物を与えました。このデータは、『Snedecor and Cochran』（1980）の 6.9 節にあります。


表 8.1 : 体重増加データ

高タンパク質	低タンパク質
134	70
146	118
104	101
119	85
124	107
161	132
107	94
83	
113	
129	
97	
123	

高タンパク質と低タンパク質の標本は、それぞれ平均値位置パラメータが μ_H および μ_L で、標準偏差尺度パラメータが σ_H および σ_L であると推定されます。主に μ に差があるかどうかに関心がありますが、2つの食物の標準偏差を計算することによって、ばらつきが異なるかどうかにも関心があるかもしれません。この例では、S-PLUS を使ってそのような問題を解決する方法を示します。

データをセットアップする

データは、2組の観測値から成るため、S-PLUS では長さの異なる2つの変数を含むデータフレームとして適切に記述されます。

1. 標準ツールバーの**新規データセット**ボタン  をクリックして、空のデータセットを開きます。

2. 表 8.1 にリストされた 19 個の観測値を入力します。最初に v1 をダブルクリックして gain.high と入力し、次に v2 をダブルクリックして gain.low と入力し、列名を変更します。ENTER を押すか、またはデータ・ウィンドウ内のどこかをクリックして変更を確定します。
3. データ・ウィンドウの左上角をダブルクリックして、データセットの名前を変更します。現われたダイアログの **Name** フィールドに weight.gain と入力し、**OK** をクリックします。

探索的データ解析

まず、分布の形を評価して両方の変量が正規分布しているかどうかを確認したいと思います。これを行うため、それぞれの変量について、箱型図、ヒストグラム、密度関数プロット、および正規 QQ プロットを作成します。データ・ウィンドウの weight.gain の列を選択し、2D プロットパレットを開き、適切なプロットボタンをクリックして、3 つのプロットを作成することができます。

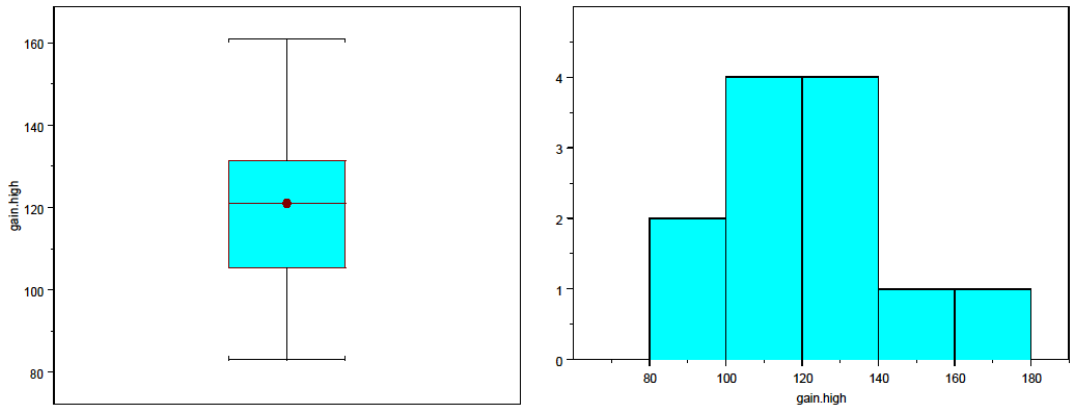


図 8.14 : 高タンパク質グループのデータの探索的データ解析プロット

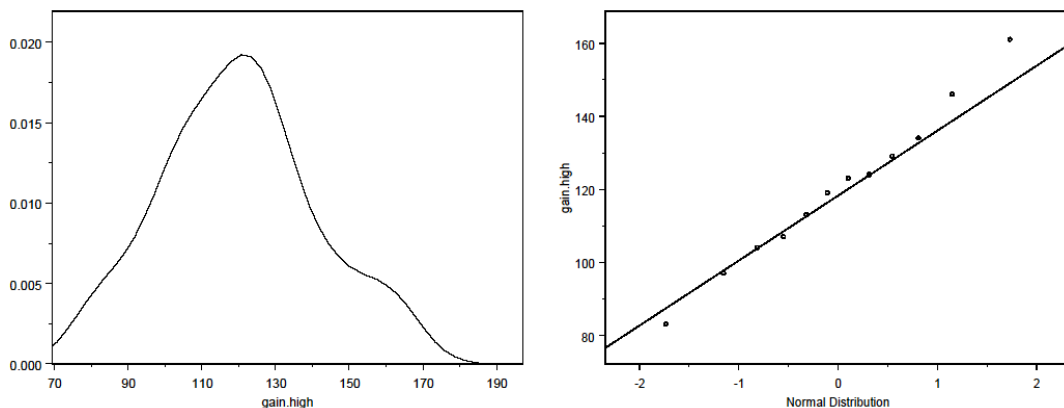


図 8.14 : 高タンパク質グループのデータの探索的データ解析プロット (続き)

高タンパク質グループのプロットを、図 8.14 に示します。この図は、データがほぼ正規分布であり、外れ値がないことを示しています。図に示していない低タンパク質グループも同じ結論が裏付けられています。

統計的推測

2つのグループのネズミの平均体重増加量は同じでしょうか。特に高タンパク質のグループは、高い平均体重増加量を示しているのでしょうか。探索的データ解析から、スチューデント t 検定が仮説に有効な検定であると言えます。1 標本検定の場合と同じように、2 つの平均値位置パラメータ μ_1 と μ_2 の差 $\mu_1 - \mu_2$ の信頼区間と仮説検定 p 値を得ることができます。これを行うため、**Two-sample t Test** ダイアログと **Two-sample Wilcoxon Test** ダイアログを使用します。

それぞれの 2 標本検定は、検定する仮説、信頼水準、2 つの標本平均の「差」と呼ばれる仮説 μ_0 によって指定されます。しかし、2 つの標本が異なる分布をもつ可能性があるため、2 つの標本が等しい分散を持つかどうかを指定することもできます。オプション **Assume Equal Variances** のチェックの有無を決定するために、分散と箱型図の略式の検討を加えることもできますし、等分散を確認するために正式に F 検定を行うこともできます。2 つの箱型図の箱の高さがほとんど同じ場合は、2 つの標本の分散もほぼ等しいでしょう。この `weight.gain` の例では、箱型図は、等分散の仮定がおそらく有効であることを示しています。この仮定を確かめるために、分散を正確に計算します。

1. P.282の方法で **Summary Statistics** ダイアログを開きます。
2. **Data Set** として `weight.gain` を入力します。
3. **Statistics** タブをクリックし、**Variance** チェックボックスを選択します。
4. **OK** をクリックします。

レポート・ウィンドウに、以下の出力が表示されます。

```
*** Summary Statistics for data in: weight.gain ***
      gain.high gain.low
Min:   83.00000  70.00000
1st Qu.: 106.25000  89.50000
Mean:  120.00000 101.00000
Median: 121.00000 101.00000
3rd Qu.: 130.25000 112.50000
Max:   161.00000 132.00000
Total N:  12.00000  12.00000
NA's :    0.00000   5.00000
Variance: 457.45455 425.33333
Std Dev.:  21.38819  20.62361
```

この2つの標本の実際の分散は、それぞれ 457.4 と 425.3 です。これらの値は、分散が等しいという主張を裏付けます。

$\mu_H - \mu_L = 0$ の両側対立と $\mu_H - \mu_L > 0$ の片側対立の2つの対立仮説に関心があります。これらの仮説を検定するために1回目にデフォルトの両側対立、2回目に片側対立仮説 `greater` の標準的な2標本t検定を2回実行します。

1. **Two-sample t Test** ダイアログを開きます。
2. **Data Set** フィールドに `weight.gain` を入力します。
3. **Variable 1** として `gain.high`、**Variable 2** として `gain.low` を選択します。デフォルトでは、**Variable 2 is a Grouping Variable** チェックボックスは選択されておらず、**Assume Equal Variances** チェックボックスは選択されています。
4. **Apply** をクリックします。データに NA があるため、メッセージが表示されますが、無視することができます。

レポート・ウィンドウに、以下のような結果が表示されます。

```
Standard Two-Sample t-Test

data: x: gain.high in weight.gain , and y: gain.low
      in weight.gain
t = 1.8914, df = 17, p-value = 0.0757
alternative hypothesis: true difference in means is
      not equal to 0
95 percent confidence interval:
      -2.193679 40.193679
sample estimates:
mean of x mean of y
      120      101
```

p 値は 0.0757 であり、したがって帰無仮説が有意水準 0.05 では棄却されませんが、有意水準 0.10 では棄却されます。信頼区間は (-2.2,40.2) です。すなわち、水準 0.05 で 2 つの食物間の体重増加量に有意に差がないことが分かります。

$\mu_H - \mu_L > 0$ の片側対立を検定するために、**Two-sample t Test** ダイアログの **Alternative Hypothesis** フィールドを **greater** に変更します。**OK** をクリックして検定を実行し、次のような出力を確認してください。

```
Standard Two-Sample t-Test

data: x: gain.high in weight.gain , and y: gain.low
      in weight.gain
t = 1.8914, df = 17, p-value = 0.0379
alternative hypothesis: true difference in means is
      greater than 0
95 percent confidence interval:
      1.525171      NA
sample estimates:
mean of x mean of y
      120      101
```

このケースでは、 p 値は両側対立の p 値のちょうど半分です。 p 値の間の関係は一般的に保たれます。また使用する対立仮説が `greater` で、信頼境界が下側だけになることが分かります。これは、“より大きい” 対立仮説に対応する自然な片側信頼区間です。

2 標本 ウィルコクソン 検定

「ウィルコクソン順位和検定」は、2 組の観測値が同じ分布によるかを検定するために使用されます。対立仮説は、観測値が形は同じでも位置が異なる分布であるというものです。この検定は、2 標本 t 検定と違って、観測値が正規（ガウス）分布によるものとは想定していません。ウィルコクソン順位和検定は、マン=ホイットニー検定に相当します。

データが対になっている場合は、ウィルコクソン順位検定のタイプとして“順位和”を指定してください。

2 標本ウィルコクソン順位検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ 2 標本 ▶ ウィルコクソン順位検定を選択してください。図 8.15 に示すように **Two-sample Wilcoxon Test** ダイアログが開きます。

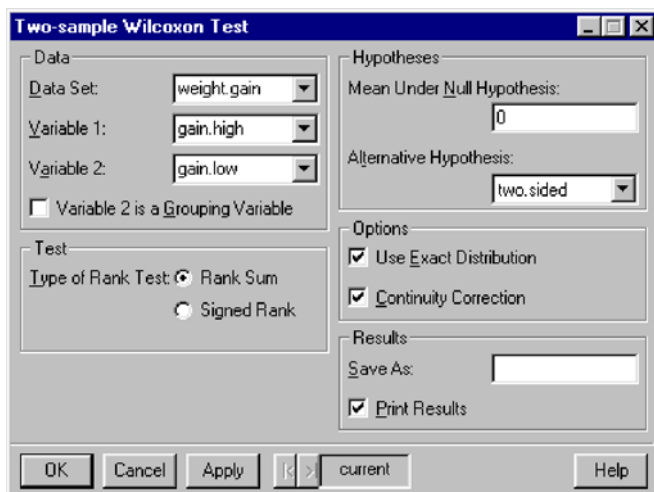


図 8.15 : Two-sample Wilcoxon Test ダイアログ

例

302 ページの「2 標本 t 検定」において、高タンパク質の食物による平均体重増加量が、低タンパク質の食物の平均体重増加量と異なるかを確認するために検定を行いました。2 標本 t 検定は、水準 0.05 では有意でなく、水準 0.10 で有意でした。正規性が保たれるため、これらのデータには 2 標本 t 検定の方が適切だと思われます。しかし説明のために、2 標本ウィルコクソン検定を

行って、2つの食物の平均体重増加量が異なるかどうかを調べます。帰無仮説が、食物の差が0であるという場合は、両側検定を行います。すなわち、平均体重増加が各食物に関して同じであるかどうかを検定します。

1. `weight.gain` データセットをまだ作成していない場合は、303 ページに示した手順で `weight.gain` データセットを作成します。
2. P.308 の方法で **Two-sample Wilcoxon Test** ダイアログを開きます。
3. **Data Set** として `weight.gain` を指定します。
4. **Variable 1** として `gain.high`、**Variable 2** として `gain.low` を選択します。デフォルトでは、**Variable 2 is a Grouping Variable** チェックボックスが選択されておらず、**Type of Rank Test** が **Rank Sum** に設定されています。**OK** をクリックします。

レポート・ウィンドウに、以下のように出力されます。

```
Wilcoxon rank-sum test
data: x: gain.high in weight.gain , and y: gain.low in
weight.gain
rank-sum normal statistic with correction Z = 1.6911,
p-value = 0.0908
alternative hypothesis: true mu is not equal to 0
```

値 107 が 2 度現われるため、メッセージ・ウィンドウに警告が表示されます。この警告は、今のところ無視することができます。p 値 0.0908 は正規近似に基づいており、これは値 107 が 2 度現れているためです。t 検定 0.0757 に近い値です。これにより、平均体重増加量には、高タンパク質と低タンパク質の食物において水準 0.05 で有意差がないという結論が裏付けられます。

コルモゴロフ＝ スミルノフ 適合度検定

「2 標本コルモゴロフ＝スミルノフ適合度検定」は、2 組の観測値が同一分布であるかどうかを検定するために使用されます。この検定は 2 つの標本がランダムで、互いに独立しており、データが少なくとも順序尺度では実測されたと仮定しています。さらにこの検定は、基礎分布が連続的な場合のみ正確な結果を提供します。

2 標本コルモゴロフ＝スミルノフ適合度検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ 2 標本 ▶ コルモゴロフ＝スミルノフ適合度検定を選択します。図 8.16 に示すように、**Two-sample Kolmogorov - Smirnov Goodness-of-Fit Test** ダイアログが開きます。



図 8.16 : Two-sample Kolmogorov-Smirnov Goodness-of-Fit Test
ダイアログ

例

データセット `kyphosis` は、脊椎矯正手術を受けた 81 人の子供のデータを表す 81 行で構成されています。結果を表す `Kyphosis` は 2 値変数であり、他の 3 つの列 `Age`、`Number` および `Start` は数値です。脊柱後弯症は、脊椎手術を受けた一部の子供に見られる術後の変形です。子供の年齢、手術した椎骨の数または起点椎骨が、子供が変形を持つ確率に影響するかどうか調べたいと思います。探索ツールとして、`Age`、`Number` および `Start` の分布が、脊柱後弯症の子供と脊柱後弯症でない子供で同じかどうか検定します。

1. P.310 の方法で **Two-sample Kolmogorov-Smirnov Goodness-of-Fit Test** ダイアログを開きます。
2. **Data Set** フィールドに `kyphosis` と入力します。
3. `Kyphosis` でグループ分けした各ケースで、3 つの共変数それぞれに別々の検定を実行します。**Variable 2** として `Kyphosis` を選択します。**Variable 2 is a Grouping Variable** チェックボックスを選択します。
4. **Variable 1** として `Age` を選択します。**Apply** をクリックします。
5. **Variable 1** として `Number` を選択します。**Apply** をクリックします。
6. **Variable 1** として `Start` を選択します。**OK** をクリックします。

レポート・ウィンドウに3つの適合度サマリーが表示されます。Age、Number、および Start の p 値はそれぞれ 0.076、0.028、および 0.0002 です。これは、脊柱後弯症の子供と脊柱後弯症でない子供は年齢による分布に大きな違いはなく、手術を受けた椎骨の数と起点椎骨がどの椎骨であるかの分布が大きく異なっていることを示しています。これは、371 ページの「ロジスティック回帰」で説明するこのようなデータに適したロジスティック回帰モデルの結果と一致しています。

k 標本検定

S-PLUS は、計画実験でグループの平均差を解析する様々な方法をサポートしています。

- **一元分散分析**：単純な 1 因子の分散解析。主効果間の交互作用は仮定されません。すなわち k 標本は独立しているとみなされ、データは正規分布でなければなりません。
- **クラスカル=ワリス順位和検定**：一元分散分析に代わるノンパラメトリック解析。分布は仮定されません。
- **フリードマン順位和検定**：反復のないブロック因子を含む 1 因子計画実験の平均のノンパラメトリック解析。

ANOVA ダイアログは、複数の因子を含む分散モデルを解析することができます。376 ページの「分散分析」を参照してください。

一元分散分析

One-Way Analysis of Variance ダイアログは、データにグループ化のための変量があり、そのグループ化変量によりデータにいくつかの独立な標本が定義されるときに、単純な分散分析 (ANOVA) テーブルを生成します。主効果間の交互作用は仮定されません。すなわち、標本は独立しているとみなされます。ANOVA テーブルは、すべてのグループの平均値が等しいかをどうかを検定する F 統計を結果に含みます。これらの統計量は、観測値が正規 (ガウス) 分布していると仮定します。

より複雑なモデルや複数の予測変量を持つ ANOVA には、**Analysis of Variance** ダイアログを使用してください。

一元 ANOVA を実行する

メインメニューから、統計 ▶ 標本比較 ▶ k 標本 ▶ 一元分散分析を選択します。図 8.17 に示すように、One-way Analysis of Variance ダイアログが開きます。

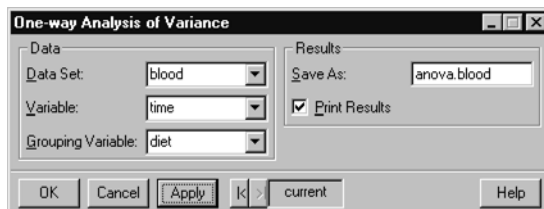


図 8.17 : One-way Analysis of Variance ダイアログ

例

最も簡単な実験は、実験「因子」のいくつかの「水準」ごとに 1 つの連続する「反応」変量を何回か実測することです。たとえば、表 8.2（「Box, Hunter, and Hunter」(1978) から）のデータを検討します。このデータは、4 つの各食物の血液凝固時間の数値から成ります。凝固時間は連続する反応変量であり、食物は A、B、C、D の 4 つの水準がある「質的」変量または「因子」です。水準 A、B、C、D に対応する食物は実験者が決定しました。

表 8.2 : 4 つの食物の血液凝固時間


Diet			
A	B	C	D
62	63	68	56
60	67	66	62
63	71	71	60
59	64	67	61
	65	68	63
	66	68	64
			63
			59

因子“食物”が血液凝固時間の平均値に影響を及ぼすか確認することに主に関心があります。“食物”のような実験因子を「処理 (treatment)」と呼ぶことがあります。


因子水準が平均凝固時間に影響をおよぼすかどうかを検定する正式な統計検定は、分散分析法 (ANOVA) を使用して行われます。この方法は、モデルの仮定が正式な ANOVA 結論を検証できるくらい正確であることを確認するために、探索的なグラフィックスで補う必要があります。S-PLUS は、データ探索と正式な ANOVA の両方を行うツールを提供します。

データをセットアップする

1 つの因子変量 (食物) と 1 つの反応変量 (時間) を持つデータを用意します。データは、2 つの列のあるデータセットとして S-PLUS に適切に記述されます。以下のステップは、表 8.2 からのデータを含む blood と名前の付けられたデータフレームを作成します。

1. 標準ツールバーの**新規データセット**ボタン  をクリックして、空のデータセットを開きます。
2. 表 8.2 にリストされた 4 列の観測値を入力します。V1 をダブルクリックして A を入力し、V2 をダブルクリックして B を入力し、その後は同じように続けて列名を変更します。ENTER を押すか、またはデータ・ウィンドウ内のどこかをクリックして変更を確定します。
3. メインメニューから、**データ ▶ 構築 ▶ スタック**を選択します。フィールドの **From** グループで、**Data Set** ボックスにデータセットの名前が表示されていることを確認します。**Stack Columns** リストで **<ALL>** を選択します。
4. **To** グループのフィールドで、**Data Set** として blood を入力し、**Stack Column** に time、**Group Column** に diet と入力します。OK をクリックします。このステップで、表 8.2 からのデータ点を含む blood と名付けられたデータセットが 2 列フォーマットで作成されます。
5. メインメニューから、**データ ▶ 構築 ▶ パック**を選択します。blood が **Data Set** フィールドに表示されたことを確認し、**Columns** リストで **<ALL>** を選択します。このステップでは、NA の観測値がすべて除去されます。

探索的データ解析

箱型図は、データを最初に調べるための迅速で簡単な方法です。blood データ・ウィンドウで diet と time の 2 つの列順にハイライトします。2D プロットパレットを開き、箱ボタン  をクリックして箱型図を生成します。

得られた箱型図は図 8.18 のようになります。このプロットは、食物 A と D の反応はよく似ている一方、箱の高さによって分かるばらつきを考慮しても食物 B と C の中央値の反応がかなり大きいことを示しています。したがって、食物は血液凝固時間に影響を及ぼすと考えられます。

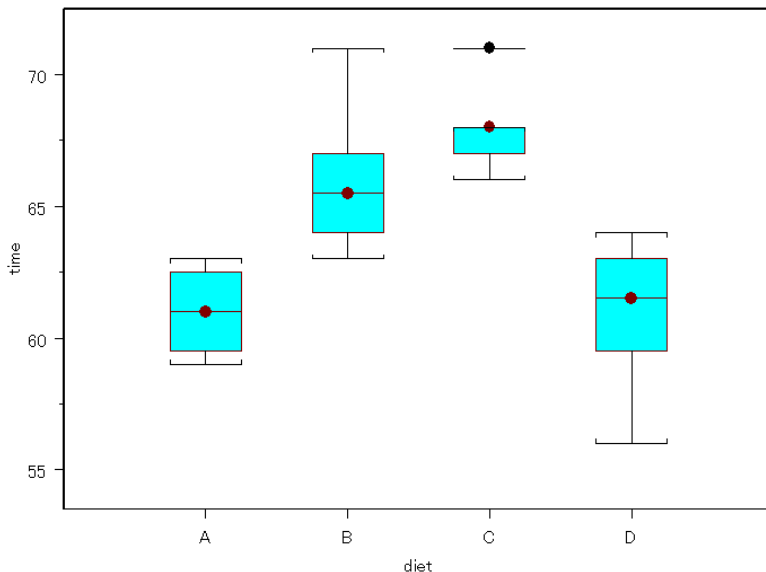


図 8.18 : blood データセットにおける 4 つの各食物の箱型図

一元配置モデルと分散分析

単一因子を使った実験の伝統的モデルは、次の通りです。

$$y_{ij} = \mu_i + \varepsilon_{ij} \quad \begin{array}{l} j = 1, \dots, J_i \\ i = 1, \dots, I \end{array}$$

ここでは μ_i は実験因子の第*i*水準の反応の平均値です。水準*I*の実験因子があり、 J_i の実測値 $y_{i1}, y_{i2}, \dots, y_{iJ_i}$ は、水準*i*の実験因子の反応変量を表します。処理という用語を使用すると*I*の処理があり、 μ_i は*i*番目の処理平均と呼ばれます。上記のモデルは、「一元配置」モデルと呼ばれることがあります。血液凝固の実験では*I*=4の食物があり、平均 $\mu_1, \mu_2, \mu_3, \mu_4$ は、食物A、B、C、Dにそれぞれ対応します。観測値の数は、 $J_A=4, J_B=6, J_C=6$ および $J_D=8$ です。

One-way Analysis of Variance ダイアログを使って分散分析を行うことができます。

1. P.312 の方法で **One-way Analysis of Variance** ダイアログを開きます。
2. **Data Set** フィールドに blood と入力します。
3. **Variable** として time、**Grouping Variable** として diet を選択します。
4. 後の節で多重比較を生成するために、**Save As** フィールドに anova.blood を入力して結果を保存します。
5. **OK** をクリックして ANOVA を実行します。

レポート・ウィンドウに次のような結果が表示されます。

```
*** One-Way ANOVA for data in time by diet ***
```

```
Call:
```

```
aov(formula = structure(.Data = time ~ diet, class =
"formula"), data = blood)
```

```
Terms:
```

```
                diet Residuals
Sum of Squares  228         112
Deg. of Freedom    3          20
```

```
Residual standard error: 2.366432
```

```
Estimated effects may be unbalanced
```

```
      Df Sum of Sq Mean Sq  F Value    Pr(F)
diet   3      228    76.0 13.57143 0.00004658471
Residuals 20      112     5.6
```

p 値は 0.000047 であり、これはきわめて有意です。これにより、食物が血液凝固時間に影響を及ぼすことが分かります。

クラスカル＝ワリス順位和検定

「クラスカル＝ワリス順位検定」は、一元分散分析の代わりにノンパラメトリック解析です。帰無仮説は、それぞれのグループで y の真の位置パラメータ（平均）が同じであるというものです。対立仮説は、グループのうちの少なくとも 1 つで y の真の位置パラメータ（平均）が異なるというものです。この検定は一元 ANOVA と違い、正規性を必要としません。

クラスカル＝ワリス順位和検定を実行する

メインメニューから統計 ▶ 標本比較 ▶ k 標本 ▶ クラスカル＝ワリス順位検定を選択します。図 8.19 に示すように、**Kruskal-Wallis Rank Sum Test** ダイアログが開きます。

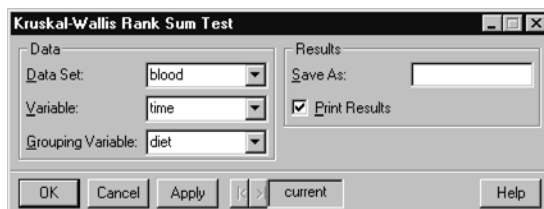


図 8.19 : Kruskal-Wallis Rank Sum Test ダイアログ

例

311 ページの一元分散分析で、食物が血液凝固時間に影響を及ぼすという結論を出しました。一元 ANOVA は、データが正規分布する必要があります。ノンパラメトリックのクラスカル＝ワリス順位和検定は分布を仮定せず、様々なデータに適用することができます。次に、blood データセットにクラスカル＝ワリス順位和検定を行います。

1. blood データセットがまだ作成されていない場合は、313 ページの手順で blood データセットを作成してください。
2. P.316 の方法で **Kruskal-Wallis Rank Sum Test** ダイアログを開きます。
3. **Data Set** フィールドに blood と入力します。
4. **Variable** として time、**Grouping Variable** として diet を選択します。
5. **OK** をクリックします。

レポート・ウィンドウに、次のような結果が表示されます。

```
Kruskal-Wallis rank sum test
data: time and diet from data set blood
Kruskal-Wallis chi-square = 17.0154, df = 3,
p-value = 0.0007
alternative hypothesis: two.sided
```

p 値は 0.0007 であり、これはきわめて有意です。一元 ANOVA の結果をクラスカル=ワリス順位和検定が確認しています。

フリードマン 順位検定

「フリードマン順位検定」は、反復のない完全なブロック計画から得られたデータに適しています。そのような種類の計画では、それぞれの処理において、それぞれの実験単位すなわち「ブロック」から 1 つの観測値だけが収集されます。 y の要素はグループ効果、ブロック効果、さらに独立で同一に分散した誤差の合計であると仮定されます。グループとブロック間の交互作用はないと仮定されます。

因子のグループとブロックによる二元配置の状況において、典型的な帰無仮説はブロック効果の総計である y の真の位置母数（平均）が各グループで同じというものです。対立仮説は、グループの少なくとも 1 つで真の位置母数が異なるというものです。

フリードマン順位検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ k 標本 ▶ フリードマン順位検定を選択します。図 8.20 に示すように、**Friedman Rank Sum Test** ダイアログが開きます。

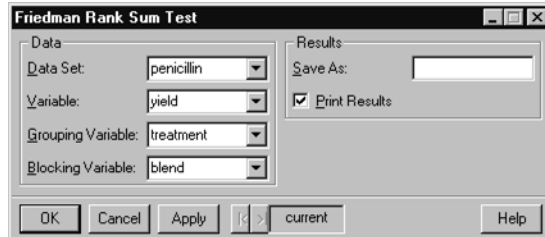


図 8.20 : Friedman Rank Sum Test ダイアログ

例

表 8.3 に示したデータセットは、最初に 1978 年に「Box, Hunter, and Hunter」によって使用されました。ペニシリン製造プロセスにおいて処理 A、B、C、D がペニシリンの収率に及ぼす影響を決定するために収集されたものです。反応変量は yield であり、処理変量は treatment です。それぞれの処理を適用するためにコーンステーパーリカーの別の混合を行わなければならなかったため、第 2 の因子 blend があります。

主な関心は、因子 treatment が yield に影響を及ぼすかどうかを判定することです。因子 blend は二次的な関心に過ぎず、処理の推測の感度を高めるために導入されたブロック化変量です。ブロック内の処理の順序は任意に選択しました。したがって、これは乱塊法実験です。

表 8.3 : ペニシリンの収率に対する 4 つの処理の影響

blend	treatment	yield
1	A	89
2	A	84
3	A	81
4	A	87
5	A	79

表 8.3 : ペニシリンの収率に対する 4 つの処理の影響 (続き)

blend	treatment	yield
1	B	88
2	B	77
3	B	87
4	B	92
5	B	81
1	C	97
2	C	92
3	C	87
4	C	89
5	C	80
1	D	94
2	D	79
3	D	85
4	D	84
5	D	88

データをセットアップする

Factorial Design ダイアログを使って、表 8.3 の情報を含む penicillin データセットを作成します。このダイアログについて詳細は、340 ページの「要因計画」を参照してください。

1. メインメニューから **統計 ▶ 計画 ▶ 要因計画** を選択します。
2. **Levels** フィールドに 5, 4 を入力します。このステップで、データセットの第 1 の列に 5 つの水準が指定され、第 2 の列に 4 つの水準が指定されます。
3. ここで、2 つの因子列の水準に名前を付ける必要があります。**Factor Names** フィールドに、次の式を入力します。

```
c("Blend 1", "Blend 2", "Blend 3", "Blend 4", "Blend 5"), c("A", "B", "C", "D")
```

4. **Save In** フィールドに penicillin を入力し、**OK** をクリックします。
5. penicillin データ・ウィンドウの第3列に、表 8.3 から yield の 20 の値を入力します。
6. 列の名前を、表の名前と一致するように、blend、treatment、および yield に変更します。

統計的推測

フリードマン順位検定を使用し、処理の影響がないという帰無仮説を検定します。

1. P.318 の方法で **Friedman Rank Sum Test** ダイアログを開きます。
2. **Data Set** フィールドに penicillin と入力します。
3. **Variable** として yield、**Grouping Variable** として treatment、**Blocking Variable** として blend を選択します。
4. **OK** をクリックします。

フリードマン検定のサマリーがレポート・ウィンドウに現われます。p 値は 0.322 で、これはあまり有意ではありません。この p 値は、漸近カイ二乗近似を使用して計算されます。

仮説検定

S-PLUS は、仮説検定を分析するための様々な方法をサポートしています。

- **二項検定**：データが指定の比率パラメータを持つ分布によるかを評価するために二項データに使用される正確検定。
- **比率検定**：二項標本の比率パラメータが特定された値を持つか、あるいは2つの二項標本が同じ比率パラメータを持つかを評価するカイ二乗検定。
- **フィッシャーの正確確率検定**：分割表の行と列間の独立性に関する検定。
- **マクネマーの検定**：対応する変量があるときの分割表内の独立性に関する検定。
- **マンテル=ヘンツェル検定**：3次元分割表の独立性のカイ二乗検定。
- **カイ二乗検定**：2次元分割表の独立性のカイ二乗検定。

二項データは、 n 回の試行のうち成功回数 k を表すデータです。観測値は、成功確率 p に依存せずに発生します。分割表は、複数のカテゴリカル（因子）変量のそれぞれの組み合わせの出現する回数を表わしたものです。

二項検定

「正確二項検定」は二項データを使用して、データが指定された比率パラメータ p を持つ分布による可能性があるかどうかを評価します。二項データは、 n 回の試行のうち成功回数 k を表すデータであり、観測値は成功確率 p と関係なく生じます。この例には、コイン投げデータがあります。

正確二項検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ 仮説検定 ▶ 二項検定を選択します。図 8.21 に示すように、**Exact Binomial Test** ダイアログが開きます。

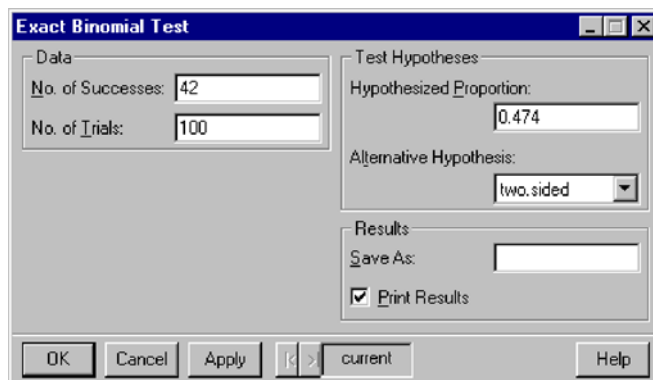


図 8.21 : Exact Binomial Test ダイアログ

例

ルーレットで赤に賭けるとき、勝つ確率が 0.5 よりもわずかに小さいと予想します。これは、アメリカのルーレットでは、円盤には、ボールが落ちることができるマス目が全部で 38 あり、そのうち赤いマス目が 18、黒いマス目が 18、さらに“0”と“00”とラベルのついたマス目が 2 つあるからです。したがって、公正なすなわち完全に均等な円盤の場合、赤の確率は、 $p_0 = 18 \div 38 = 0.474$ であることが想像されます。赤の確率が 0.474 よりも小さくなるように賭博場がルーレットの円盤を改造し、不正を働いていないことを期待しています。

たとえば赤に 100 回賭けて赤が 42 回出たとします。その結果が公正なルーレットの円盤による妥当なものかどうかを確認したいとします。

1. P.321 の方法で **Exact Binomial Test** ダイアログを開きます。
2. **No. of Successes** として 42 を入力します。 **No. of Trials** として 100 を入力します。
3. **Hypothesized Proportion** として 0.474 を入力します。
4. **OK** をクリックします。

検定のサマリーがレポート・ウィンドウに表示されます。p 値 0.3168 は、標本が比率パラメータ 0.474 による二項分布から取り出されたデータと近いことを示します。したがって、ルーレットの円盤は公正であると考えられます。

比率検定

「比率検定」は、ピアソンのカイ二乗統計を使用して、二項標本が指定された比率パラメータ p を持つかどうかを評価します。さらに、複数の標本が同じ比率パラメータを有するかどうかを評価することもできます。比率検定は二項分布に正規近似を使用するため、正確二項検定よりも有効ではありません。したがって、通常は正確二項検定の方が適しています。比率検定の長所は、比率パラメータの信頼区間を提供することと、複数の標本に使用できることです。

比率検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ 仮説検定 ▶ 比率検定を選択します。図 8.22 に示すように、**Proportions Test** ダイアログが開きます。

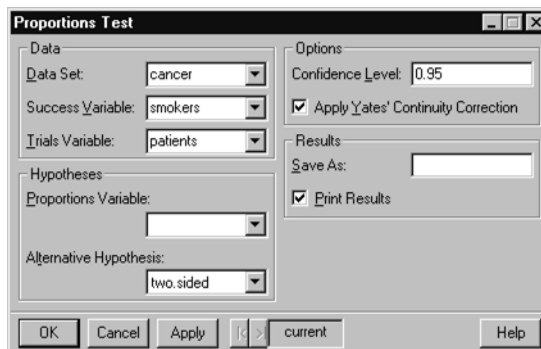


図 8.22 : Proportions Test ダイアログ

例


対象の複数の標本があり、それぞれの対象が何らかの特徴の有無によって特徴付けられることがあります。言いかえると、各試行の結果が成功か失敗から成る 3 組以上の試行がある場合です。たとえば表 8.4 に示したデータセットは、Fleiss (1981) が発表した肺ガン患者に対する異なる 4 つの研究結果をまとめたものです。それぞれの研究には patients 数があり、各研究のタバコを吸う患者の数が smokers でした。

表 8.4 : 肺ガン患者に対する異なる 4 つの研究

Smokers	Patients
83	86
90	93
129	136
70	82

データをセットアップする

表 8.4 の情報を含むデータセット cancer を作成します。

1. 標準ツールバーの**新規データセット**ボタン  をクリックして、データセットを開きます。
2. 表 8.4 に示した 2 つの列の観測値を入力します。v1 をダブルクリックして **smokers** を入力し、v2 をダブルクリックして **patients** を入力して列名を変更します。ENTER を押すか、または**データ・ウィンドウ**内のどこかをクリックして変更を確定します。
3. **データ・ウィンドウ**の左上角をダブルクリックしてデータセットの名前を変更します。現われたダイアログで、**Name** フィールドに **cancer** を入力し、**OK** をクリックします。

統計的推測

cancer データについては、喫煙者である患者の確率が 4 つの各研究で同じかどうかに関心があります。すなわち、各研究が等質な母集団の患者を含むかどうかを検定したいと思います。

1. P.323 の方法で **Proportions Test** ダイアログを開きます。
2. **Data Set** フィールドに **cancer** を入力します。
3. **Success Variable** として **smokers**、**Trial Variable** として **patients** を選択します。
4. **OK** をクリックします。

検定のサマリーが**レポート・ウィンドウ**に表示されます。p 値 0.0056 は、比率パラメータが等しいという帰無仮説を棄却することを示します。したがって、患者のうちの喫煙者に関して、すべてのグループが同じ確率を持つと結論づけることはできません。

フィッシャーの 正確確率検定

「フィッシャーの正確確率検定」は、分割表の行と列の変量の独立性を調べる検定です。データが 2 つのカテゴリカル変量から成るとき、各因子の組み合わせの出現回数を反映させた分割表を作成することができます。フィッシャーの正確確率検定は、ある因子の値が他の因子の値に対して独立かどうかを評価します。たとえば、政党加入状況が性別とは関係がないか調べるために使用されます。ある種の等質性、たとえば $k \times 2$ 表における比率の等質性は、独立性の仮説に相当します。したがってこの検定は、そのようなケースを対象とすることもあります。

これは正確検定であるため、クロス分類表の度数の合計数が 200 を超えてはいけません。そのようなケースには、カイ二乗検定による独立性の検定が適しています。

フィッシャーの正確確率検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ 仮説検定 ▶ フィッシャーの正確確率検定を選択します。図 8.23 に示すように、**Fisher's Exact Test** ダイアログが開きます。



図 8.23 : Fisher's Exact Test ダイアログ

例


表 8.5 に示したデータセットには、臨床試験の結果をまとめた分割表です。患者は、試験薬の投与を受ける治療群と試験薬の投与を受けない対照群とに分けられています。これらの患者を 28 日間モニターし、研究の最後にその患者たちの生存状態を記録しました。

表 8.5 : 臨床試験の結果をまとめた分割表

	Control	Treated
Died	17	7
Survived	29	38

データをセットアップする

表 8.5 の情報を含む `fisher.trial` データセットを作成します。

1. 標準ツールバーの**新規データセット**ボタン  をクリックして、空のデータセットを開きます。
2. 表 8.5 に示した 2 列の観測値を入力します。
3. まず `v1` をダブルクリックして `Control` を入力し、次に `v2` をダブルクリックして `Treated` を入力し、列名を変更します。
4. 各行の最初のエントリの隣の灰色のボックス内をダブルクリックして、行名を変更します。第 1 行名に `Died`、第 2 行名に `Survived` を入力します。行名を見やすくするように灰色の列のサイズを変更することができます。これを行うときは、列の右側の境界を希望する幅までドラッグ アンド ドロップしてください。
5. **データ・ウィンドウ**の左上角をダブルクリックして、データセットの名前を変更します。現われたダイアログの **Name** フィールドに `fisher.trial` と入力し、**OK** をクリックします。

統計的推測

治療が生存確率に影響を及ぼしたかどうかを調べたいと思います。

1. P.325 の方法で **Fisher's Exact Test** ダイアログを開きます。
2. **Data Set** フィールドに `fisher.trial` と入力します。
3. **Data Set is a Contingency Table** チェックボックスを選択します。
4. **OK** をクリックします。

検定のサマリーが**レポート・ウィンドウ**に表示されます。p 値 0.0314 は、独立性の帰無仮説を棄却することを示しています。したがって、治療が生存確率に影響を及ぼすと結論づけることができます。

マクネマーの検定

カテゴリカル変量が 2 つの実験では、変量のうちの 1 つで、異なる治療を受ける個体の複数のグループを指定します。そのような状況では、統計的推測の精度を高めるためにしばしば個体の突き合わせが行われます。ただし突き合わせを行うとき、観測値は通常独立していません。そのようなケースでは、カイ二乗検定、フィッシャーの正確確率検定、およびマンテル=ヘンツェル検定から得られた推定は、これらの検定がすべて観測値が独立していることを仮定しているため有効ではありません。

「マクネマーの検定」を使用して、突き合わせを行う実験に有効な推定を得ることができます。マクネマーの統計は、対称性すなわち観測値がセル $[i,j]$ に分類される確率がセル $[j,i]$ に分類される確率と同じであるという帰無仮説を検定するために使用されます。返された p 値を注意深く解釈してください。その有効性は、セル度数が少なくとも適度に大きいという仮定に依存します。セル度数が適切なときでも、カイ二乗法は帰無仮説によるマクネマーの統計量の真の分布に関する唯一の大標本近似です。

マクネマーの検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ 仮説検定 ▶ マクネマーの検定を選択します。図 8.24 に示すように、McNemar's Chi-Square Test ダイアログが開きます。

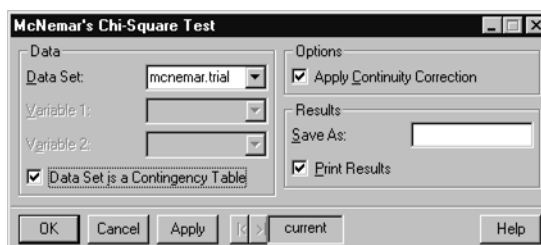


図 8.24 : McNemar's Chi-Square Test ダイアログ

例

表 8.6 に示したデータセットは、突き合わせた対のデータの分割表を含み、各度数が、突き合わせた対の個体と関連付けられます。

表 8.6 : 突き合わせた対のデータの分割表


	B,Survive	B,Die
A,Survive	90	16
A,Die	5	510

この表では、各エントリが 1 対の患者を表しており、一方の患者は治療 A を受け、他方の患者は治療 B を受けました。たとえば、左下のセルの 5 は 5 対で、治療 A を受けた個体は死亡したけれど、その個体と対になり治療 B を受けた個体が生存していることを意味します。ここでは、珍しい種類のガンの治療における治療 A と B の相対的な効果に関心があります。

表内で突き合わせた対のうち、片方の個体は生存しているけれど他方の個体が死亡している対は「一致しない対」と呼ばれます。治療 A を受けた個体が生存し、治療 B を受けた個体が死亡する一致しない対は 16 あります。治療 A を受けた個体が死亡し、治療 B を受けた個体が生存する逆の状況の一致しない対は 5 つあります。両方の治療が同じように効果的な場合は、これらの 2 つのタイプの一致しない対がほぼ等しい頻度で生じると予想されます。確率の点から、帰無仮説は $p_1 = p_2$ であり、ここで p_1 は突き合わせた対の個体に第 1 のタイプの不一致が生じる確率で、 p_2 は第 2 のタイプの不一致が生じる確率です。

データをセットアップする

表 8.6 の情報を含むデータセット `mcnemar.trial` を作成します。

1. 標準ツールバーの**新規データセット**ボタン  をクリックして、空のデータセットを開きます。
2. 表 8.5 に示した 2 列の観測値を入力します。

3. まず V1 をダブルクリックして B.Survive を入力し、V2 をダブルクリックして B.Die を入力し、列名を変更します。
4. 各行の最初のエントリの隣にある灰色のボックス内をダブルクリックして、行名を変更します。第 1 行名に A.Survive、第 2 行名に A.Die を入力します。行名を見やすくするために、灰色の列のサイズを変更することができます。これを行うときは、列の右側の境界を希望する幅までドラッグ アンド ドロップしてください。
5. データ・ウィンドウの左上角をダブルクリックして、データセットの名前を変更します。現われたダイアログの **Name** フィールドに mcnemar.trial を入力し、**OK** をクリックします。

統計的推測

マクネマーの検定を使用して、治療が同じように有効かどうか調べます。

1. P.327 の方法で **McNemar's Square Test** ダイアログを開きます。
2. **Data Set** フィールドに mcnemar.trial と入力します。
3. **Data Set is a Contingency Table** チェックボックスを選択します。
4. **OK** をクリックします。

検定のサマリーがレポート・ウィンドウに表示されます。p 値 0.0291 は、表の対称性の帰無仮説を棄却することを示しています。これは、2 つの治療の効果が異なることを示しています。

マンテル＝ ヘンツェル検定

「マンテル＝ヘンツェル検定」は、3 次元の分割表における独立性のあるカイ二乗検定を行います。これは、3 つの因子で構成された分割表に使用されます。マクネマーの検定と同じように、返された p 値を注意深く解釈してください。有効性は、期待セル度数の合計が少なくとも適度に大きいという仮定に依存します。セル度数が適切なきでも、カイ二乗法は、帰無仮説下のマンテル＝ヘンツェル統計量の真の分布に関する唯一の大標本近似です。

マンテル=ヘンツェル検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ 仮説検定 ▶ マンテル=ヘンツェル検定を選択します。図 8.25 に示すように、**Mantel-Haenszel's Chi-Square Test** ダイアログが開きます。



図 8.25 : Mantel-Haenszel's Chi-Square Test ダイアログ

例

表 8.7 に示したデータセットは、ガンの研究結果をまとめた 3 元分割表を含みます。第 1 列は、個体が Smoker であるかどうかを示します。第 2 列において、“Case” はガンにかかった個体を指し、“Control” はガンにかかっていない個体を指します。第 3 列は、個体が「Passive (受動的) な喫煙者」かどうかを示します。受動的な喫煙者は、喫煙者と生活を共にする個体です。したがって、個体が喫煙者と受動的喫煙者の両方である場合もあります。第 4 列は、Smoker、Group、Passive の値のそれぞれの組み合わせの個体の数を示します。

表 8.7 : ガンの研究結果をまとめた 3 元分割表

Smoker	Group	Passive	Number
Yes	Case	Yes	120
Yes	Case	No	111
Yes	Control	Yes	80
Yes	Control	No	155
No	Case	Yes	161
No	Case	No	117
No	Control	Yes	130
No	Control	No	124

データをセットアップする

Orthogonal Array Design ダイアログを使って、表 8.7 の情報を含む `mantel.trial` データセットを作成します。このダイアログの詳細は、341 ページの「直交要因計画」を参照してください。

1. メインメニューから**統計 ▶ 計画 ▶ 直交要因計画**を選択します。
2. **Levels** フィールドに 2, 2, 2 を入力します。このステップで、データセットの最初の 3 つの各列にそれぞれ 2 水準を指定します。
3. 3 つの因子列の水準の名前を付ける必要があります。**Factor Names** フィールドに、次の式を入力してください。
`c("Yes", "No"), c("Case", "Control"), c("Yes", "No")`
4. **Save In** フィールドに `mantel.trial` と入力し、**OK** をクリックします。
5. `mantel.trial` データ・ウィンドウの第 4 列に、表 8.7 から 8 つの Number 値を入力します。
6. 列 `Smoker`、`Group`、`Passive` および `Number` の名前を、表の名前と一致するように変更します。

データセット `mantel.trial` には、それぞれ 2 つの水準を持つ、3 つの因子の 8 つの全ての組み合わせを表す 8 行があります。しかし、**Mantel-Haenszel's Chi-Square Test** ダイアログは、データが生形であることが必要で、分割表になったデータは使用できません。**Subset** ダイアログを使って、次のように生データを作成しなおすことができます。

1. メインメニューから、**データ ▶ 行の部分集合**を選択します。
2. **Data Set** フィールドに `mantel.trial` と入力します。
3. **Subset Rows with** フィールドに `rep(1:8, Number)` と入力します。これにより、`Number` 列内の対応する度数によって示された回数だけ整数 1~8 が複製されます。
4. **Save In** フィールドに `mantel.raw` と入力し、**OK** をクリックします。

`mantel.raw` データセットの最初の 3 つの列は、分割表に相当する分類していないデータを表します。これは、**Mantel-Haenszel's Chi-Square Test** ダイアログで必要な形式です。次の解析例では、`mantel.raw` データを使用します。

統計的推測

Mantel-Haenszel's Chi-Square Test ダイアログを使って、ガン状態と受動的喫煙の間の独立性を検定します。

1. P.330 の方法で **Mantel-Haenszel's Chi-Square Test** ダイアログを開きます。
2. **Data Set** フィールドに `mantel.raw` と入力します。
3. **Variable 1** として `Group`、**Variable 2** として `Passive`、**Stratification Variable** として `Smoker` を選択します。
4. **OK** をクリックします。

検定のサマリーがレポート・ウィンドウに表示されます。p 値 0.0002 は、ガン状態と受動的喫煙の間の独立性の帰無仮説を棄却することを示しています。

カイ二乗検定

「カイ二乗検定」は、2 次元分割表でピアソンのカイ二乗検定を実行します。この検定は、行と列の統計的独立性、グループの等質性など、いくつかのタイプの帰無仮説に関連しています。特定の帰無仮説へこの検定を行うことが適当かどうか、ならびに結果の解釈は、使用しているデータの性質に依存します。特に、カイ二乗検定が適当かどうかを決める要因としては、サンプリング方法が重要です。

カイ二乗検定によって返された p 値は、注意深く解釈してください。有効性は期待セル度数が少なくとも適度に大きいという仮定に強く依存しており、しばしば最小サイズ 5 が経験則として引用されることがあります。セル度数が適切なきでも、カイ二乗法は帰無仮説のもとでカイ二乗の真の分布に関する唯一の大標本近似です。データが小さくてカイ二乗検定に適していない場合は、フィッシャーの正確確率検定が望ましいことがあります。

ピアソンのカイ二乗検定を実行する

メインメニューから、統計 ▶ 標本比較 ▶ 仮説検定 ▶ カイ二乗検定を選択します。図 8.26 に示すように、Pearson's Chi-Square Test ダイアログが開きます。

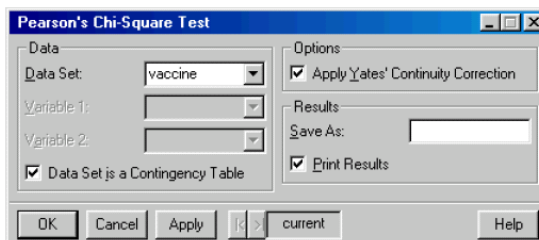


図 8.26 : Pearson's Chi-Square Test ダイアログ

例

表 8.8 に示したデータセットは、1950 年代初めのソークワクチンの試験結果による分割表を含みます。ソークの試験には、“種痘済み”と“偽薬”の 2 つの水準を有する種痘状態と、“非ポリオ”、“非麻痺性ポリオ”および“麻痺性ポリオ”の 3 つの水準を有するポリオ状態を表わす、2 つのカテゴリカル変量があります。予防接種を受けた 200,745 の個体のうち、24 が非麻痺性ポリオ、33 が麻痺性ポリオにかかり、残りの 200,688 はどのポリオにもかかりませんでした。偽薬を受けた 201,229 の個体のうち、27 が非麻痺性ポリオ、115 が麻痺性ポリオにかかり、残りの 201,087 はどのポリオにもかかりませんでした。


表 8.8 : ソークワクチンの試験結果をまとめた分割表

	None	Nonparalytic	Paralytic
Vaccinated	200,688	24	33
Placebo	201,087	27	115

分割表データを扱う際の、最も大きな関心は、表示された度数を持つ 2 つのカテゴリカル変量間で統計的に何らかの関連があるかどうかを決定することです。帰無仮説は、2 つの変量が統計的に独立しているというものです。

データをセットアップする

表 8.8 の情報を含む vaccine データセットを作成します。

1. 標準ツールバーの**新規データセット**ボタン  をクリックして、空のデータセットを開きます。
2. 表 8.8 に示した 3 列の観測値を入力します。
3. V1 をダブルクリックして None を入力し、V2 をダブルクリックして Nonparalytic を入力し、V3 をダブルクリックして Paralytic を入力して、列名を変更します。
4. 各行の第 1 のエントリの隣の灰色のボックス内をダブルクリックして、行名を変更します。第 1 行名に Vaccinated、第 2 行名に Placebo を入力します。
5. データ・ウィンドウの左上角をダブルクリックして、データセットの名前を変更します。現われたダイアログの **Name** フィールドに vaccine と入力し、**OK** をクリックします。

統計的推測

vaccine データの独立性のカイ二乗検定を実行します。

1. P.333 の方法で **Pearson's Chi-Square Test** ダイアログを開きます。
2. **Data Set** フィールドに vaccine を入力します。
3. **Data Set is a Contingency Table** チェックボックスを選択します。
4. **OK** をクリックします。

検定のサマリーが**レポート・ウィンドウ**に表示されます。p 値 0 は、独立であるという帰無仮説を棄却していることを示します。種痘状態とポリオ状態には関連があります。

検出力と標本数

研究を計画するときに生じる最初の問題の 1 つは、どれほどの大きさの標本が必要かということです。標本数は、対象の検出可能な最小差、真の帰無仮説を棄却する許容可能な確率（アルファ）、偽の帰無仮説を適切に棄却するのに必要な確率（検出力）、研究対象の母集団のばらつきに依存します。

S-PLUS は、平均あるいは二項比率の 1 標本および 2 標本検定に関する検出力と標本数を計算することができます。

- **正規検出力および標本数**：標本平均などの漸近的に正規分布した統計の標本数を計算します。また、指定数の標本検出力または検出可能な最小差を計算するために使用されます。
- **二項の検出力と標本数**：比率などの漸近的に二項分布した統計の標本数を計算します。また、指定数の標本検出力または検出可能な最小差を計算するために使用されます。

平均

Normal Power and Sample Size ダイアログを使用して、漸近的に正規分布した統計の標本数を計算することができます。また、このダイアログで、指定数の標本検出力または検出可能な最小差を計算するために使用することもできます。

平均の検出力と標本数を計算する

メインメニューから、**統計 ▶ 検出力と標本数 ▶ 平均**を選択します。図 8.27 に示すように、**Normal Power and Sample Size** ダイアログが開きます。

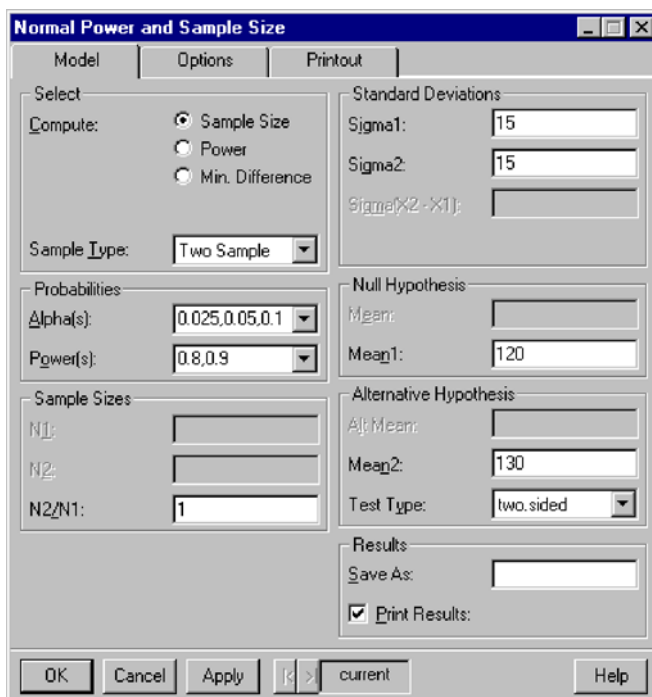


図 8.27 : Normal Power and Sample Size ダイアログ

例

ある科学者が新しい治療方法の効果を調査しています。計画では試験群の半分は治療を施し、治療した被験者の診断酵素の水準を治療しなかった対照被験者と比較することになっています。科学者は治療によって診断酵素の濃度が大きく変化するかどうかを判定するために、必要な被験者の数を決定しなければなりません。

過去の情報では、平均的な酵素の水準は 120 で、標準偏差が 15 であることを示しています。治療群と対照群の平均水準の差が 10 以上になることは、臨床的に重要であると考えられます。科学者は、アルファ（異なっていないのに両群が異なると偽って主張する確率）と検出力（異なっているときに両群が異なると正しく主張する確率）の様々な組み合わせについて、必要な標本数を決定したいと考えています。

Normal Power and Sample Size ダイアログは、アルファと検出力の様々な組み合わせの標本数の表を生成します。

1. P.335 の方法で **Normal Power and Sample Size** ダイアログを開きます。
2. **Sample Type** として Two Sample を選択します。
3. **Mean1** として 120、**Mean2** として 130、**Sigma1** と **Sigma2** の両方に 15 を入力します。
4. **Alpha(s)** に 0.025、0.05、0.1 を入力し、**Power(s)** に 0.8、0.9 を入力します。それらのアルファ値と検出力値のすべての組み合わせについて標本数を計算します。
5. **OK** をクリックします。

検出力表がレポート・ウィンドウに表示されます。表は様々なアルファと検出力の水準において、各群に必要な標本数 n1 および n2 を示します。たとえば、科学者はアルファ 0.05 と検出力 0.8 における差 10 を決定するために、群ごとに 36 の被験者を必要とします。

```

*** Power Table ***
mean1 sd1 mean2 sd2 delta alpha power n1 n2
1 120 15 130 15 10 0.025 0.8 43 43
2 120 15 130 15 10 0.050 0.8 36 36
3 120 15 130 15 10 0.100 0.8 28 28
4 120 15 130 15 10 0.025 0.9 56 56
5 120 15 130 15 10 0.050 0.9 48 48
6 120 15 130 15 10 0.100 0.9 39 39

```

二項比率

Binomial Power and Sample Size ダイアログを使用すると、比率などの漸近的に二項分布した統計の標本数を計算することができます。また、指定数の標本検出力または検出可能な最小差を計算するために使用することもできます。

比率の検出力と標本数を計算する

メインメニューから統計 ▶ 検出力と標本数 ▶ 比率を選択します。図 8.28 に示すように、**Binomial Power and Sample Size** ダイアログが開きます。

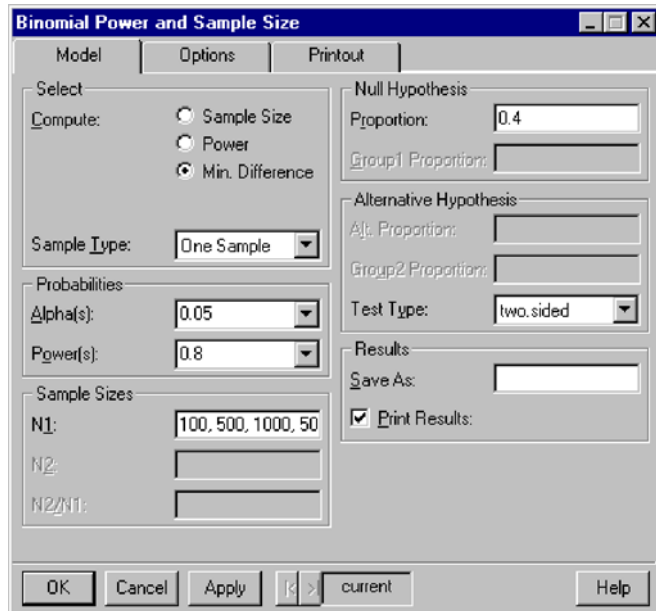


図 8.28 : Binomial Power and Sample Size ダイアログ

例

今まで、ある下院議員選挙区の有権者の 40%は、民主党議員の候補者に投票しています。世論調査員は、次の選挙の民主党投票者の比率を調べたいと思います。世論調査員は、様々な標本数で検出できる差の大きさを知りたいと考えています。すなわち民主党投票者の比率が今までの基準と大きく異なると主張するためには、標本の民主党投票者の比率が今までの比率 40%とどれだけ違わなければならないでしょうか。

1. P.337 の方法で **Binomial Power and Sample Size** ダイアログを開きます。
2. **Compute** する値として **Min. Difference** を選択します。**Proportion** として 0.4、検討する標本数 **N1** として 100, 500, 1000, 5000 を入力します。
3. **OK** をクリックします。

検出力表がレポート・ウィンドウに表示されます。この表は、それぞれの標本数の検出可能な差 δ を示します。たとえば観測値が 1000 の場合、世論調査員は、比率が 40% と比べて少なくとも 4.34% 以上異なっていれば、今までの水準とは異なる判断することができます。

```
*** Power Table ***
  p.null   p.alt   delta alpha power  n1
1    0.4 0.5372491 0.1372491 0.05 0.8 100
2    0.4 0.4613797 0.0613797 0.05 0.8 500
3    0.4 0.4434020 0.0434020 0.05 0.8 1000
4    0.4 0.4194100 0.0194100 0.05 0.8 5000
```

実験計画

一般に研究者は、実験を始める前に計画を作成します。この計画とは、観測値を得る実験変量水準の組み合わせからなるデータセットです。次に研究者は、示された実験変量の組み合わせのいくつかの結果を計測し、新しい列を既に生成済みの計画データセットに追加することによりそれを記録します。結果を記録した後で探索的プロットを使用して、結果と実験変量の関係を調べることができます。次に、ANOVA やその他の方法を使用してデータを分析することができます。

Factorial Design および **Orthogonal Array Design** ダイアログは、条件にあった実験計画を作成します。**Design Plot**、**Factor Plot** および **Interaction Plot** ダイアログは、計画のための探索的プロットを生成します。

要因計画

Factorial Design ダイアログは、「要因計画」または「一部実施要因計画」を作成します。基本的な要因計画は、変量水準のすべての可能な組み合わせを含み、反復、ランダムイズも可能です。一部実施要因計画は、対象とするモデル効果に基づいていくつかの組み合わせを除外します。

要因実験を作成する

メインメニューから、**統計 ▶ 計画 ▶ 要因計画**を選択します。図 8.29 に示すように、**Factorial Design** ダイアログが開きます。

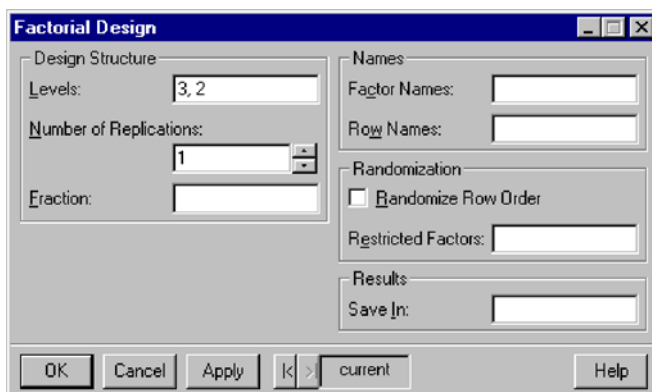


図 8.29 : Factorial Design ダイアログ

例

3水準を有する第1変量と2水準を有する第2変量を含む計画を作成します。

1. P.340の方法で **Factorial Design** ダイアログを開きます。
2. **Levels** として 3、2 を指定します。
3. **OK** をクリックします。

設計を含むデータセットが作成され、**データ・ウィンドウ**に表示されます。

直交要因計画

Orthogonal Array Design ダイアログは、直交要因計画を作成します。「直交要因計画」は、主（1番目の）効果を推測できるように構成された、もともとときわめて疎な一部実施要因計画です。2番目以降の効果を推定するのに必要な水準の組み合わせは、できるだけ少ない測定回数で求められるよう除外されます。

直交要因計画を作成する

メインメニューから**統計 ▶ 計画 ▶ 直交要因計画**を選択します。図 8.30 に示すように、**Orthogonal Array Design** ダイアログが開きます。

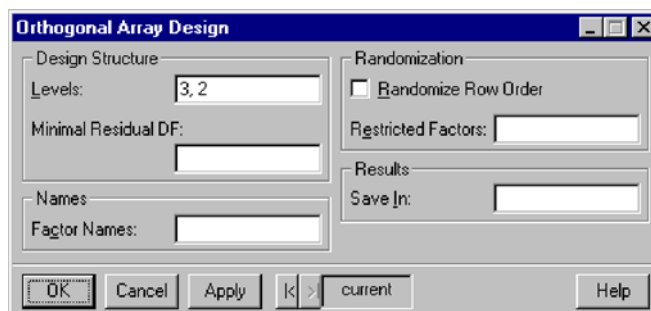


図 8.30 : Orthogonal Array Design ダイアログ

例

3水準を有する第1変量と2水準を有する第2変量を含む計画を作成します。

1. P.341の方法で **Orthogonal Array Design** ダイアログを開きます。
2. **Levels** として3、2を指定します。
3. **OK** をクリックします。

計画を含むデータセットが作成され、**データ・ウィンドウ**に表示されます。この簡単な例では、直交要因計画は前の例で作成した要因計画と同じです。

要約統計量 プロット

各因子ごとの「要約統計量プロット」は、1つまたは複数の対応する因子の各水準に対する変量の要約統計量を表示します。デフォルトで適用される関数は平均です。

要約統計量プロットを作成する

メインメニューから、**統計 ▶ 計画 ▶ 各因子ごとの要約統計量プロット**を選択します。図 8.31 に示すように、**Design Plot** ダイアログが開きます。

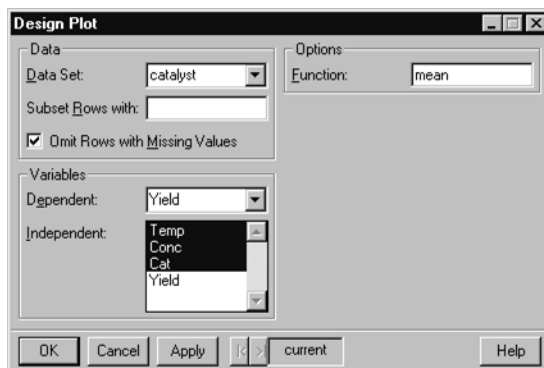


図 8.31 : Design Plot ダイアログ

例

データセット catalyst は、計画実験から得たものです。8行は、2種類の気温 (Temp)、2種類の濃度 (Conc)、および2種類の触媒 (Cat) のすべての可能な組み合わせを表します。第4列は、反応変量 Yield を表します。気温/濃度/触媒が、Yield にどのように影響を及ぼすかを調べたいと思います。ANOVA モデルをあてはめる前に、様々なプロットを使用してこれらの変数の関係を調べることができます。最初に、それぞれの因子の要約統計量プロットを作成します。

1. P.342 の方法で **Design Plot** ダイアログを開きます。
2. **Data Set** フィールドに catalyst を入力します。
3. **Dependent** 変数として Yield を選択します。
4. CTRL-クリックして、**Independent** 変数として Temp、Conc、および Cat を選択します。
5. **OK** をクリックします。

各因子ごとの要約統計量プロットがグラフシートに表示されます。このプロットには各因子ごとに縦線があり、それぞれの因子水準ごとに Yield の平均の位置を示す横線があります。

箱ひげ図

各因子ごとの「箱ひげ図」は、因子の様々な水準における統計量を比較する横並びのプロットから成ります。デフォルトでは、箱型図が使用されます。詳細は、ヘルプファイル plot.factor を参照してください。

箱ひげ図を作成する

メインメニューから、統計 ▶ 計画 ▶ 各因子ごとの箱ひげ図を選択します。図 8.32 に示すように、**Factor Plot** ダイアログが開きます。

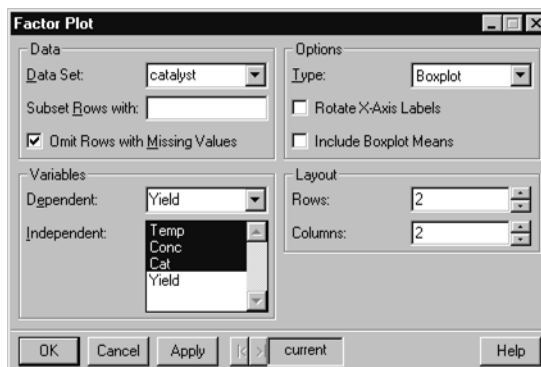


図 8.32 : Factor Plot ダイアログ

例

catalyst データセットの箱ひげ図を以下のように作成します。

1. P.340 の方法で **Factor Plot** ダイアログを開きます。
2. **Data Set** フィールドに catalyst と入力します。
3. **Dependent** 変数として Yield を選択します。
4. CTRL-クリックして、**Independent** 変数として Temp、Conc、および Cat を選択します。
5. **Rows** の数と **Columns** の数を 2 に変更します。これで、2×2 の格子のプロットが指定されます。
6. **OK** をクリックします。

グラフシートに各因子ごとの箱ひげ図が表示されます。各因子について、因子水準ごとの Yield の統計量を示す箱ひげ図となります。

交互作用の チェック

「交互作用プロット」は、*x* 軸上にある因子の水準、*y* 軸上に反応、および別の因子の特定水準に対応する点を線で表します。このタイプのプロットは、交互作用を調べたり見つけたりするのに役立ちます。

交互作用プロットを作成する

メインメニューから、**統計 ▶ 計画 ▶ 交互作用のチェック** を選択します。図 8.33 に示すように、**Interaction Plot** ダイアログが開きます。

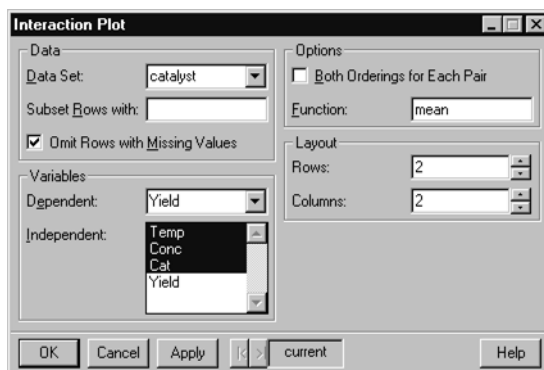


図 8.33 : Interaction Plot ダイアログ

例

catalyst データセットの交互作用プロットを以下のように作成します。

1. **Interaction Plot** ダイアログを開きます。
2. **Data Set** フィールドに catalyst と入力します。
3. **Dependent** 変量として Yield を選択します。
4. CTRL-クリックして **Independent** 変量として、Temp、Conc、および Cat を選択します。
5. **Rows** の数と **Columns** の数を **2** に変更します。これにより、 2×2 の格子のプロットが指定されます。
6. **OK** をクリックします。

グラフシートに交互作用プロットが表示されます。2 つの因子の組み合わせごとに、第 1 因子の各水準における第 2 因子の各水準の Yield の平均を示す線が作成されます。プロットの線が交わる場合は、2 つの因子間に交互作用があることを示しています。

回帰

回帰は、様々な予測変量が反応変量とどのように関係しているかを評価する標準的な方法です。この節では、**統計 ▶ 回帰**メニューから使用できる回帰法について説明します。

- **線形回帰**：最小二乗法を用いて、予測変量の線形結合関数として連続反応量を予測します。
- **ロバスト MM 回帰**：MM ベースのロバスト法を用いて、連続反応量を予測します。
- **ロバスト LTS 回帰**：刈り込み最小二乗法を用いて、連続反応量を予測します。
- **ステップワイズ線形回帰**：段階的モデル選択を使用して、線形回帰モデルで有効な変量を選択します。
- **一般化加法モデル**：予測変量のノンパラメトリック平滑 1 変量関数の加法モデルとして、一般化反応変量を予測します。
- **局所回帰**：最小二乗法を使用し、予測変量のノンパラメトリック平滑化関数として連続反応量を予測します。
- **非線形回帰**：最小二乗法を使用し、予測変量の非線形関数として連続反応量を予測します。
- **一般化線形モデル**：最尤法を使用して、予測変量の線形結合として一般化反応変量を予測します。
- **対数線形（ポアソン）回帰**：ポアソン最尤法を使用して、度数を予測します。
- **ロジスティック回帰**：ロジスティック関数にリンクする二項最尤法を使用して、2 値反応変量を予測します。
- **プロビット回帰**：プロビット関数にリンクする二項最尤法を使用して、2 値反応変量を予測します。

線形回帰

「線形回帰」は、連続変量またはカテゴリカル変量が連続反応変量に及ぼす影響を説明するために使用される、最も一般的な回帰方法です。線形回帰モデルは、予測変量を線形結合で定式化し、ランダムな変動（誤差）を加えることによって、反応変量が求まることを仮定しています。誤差は、分散が一定のガウス（正規）分布で予測変量と独立であると仮定されます。

線形回帰は、残差の2乗和を最小にする線を当てはめる「最小二乗法」です。一組 n 個の反応変量 y_i の観測値が、モデル $\hat{y} = f(\hat{x})$ による一組の予測変量 \hat{x} の値に対応すると仮定します。但し、 $\hat{y} = (y_1, y_2, \dots, y_n)$ と $\hat{x} = (x_1, x_2, \dots, x_n)$ です。 i 番目の「残差」 r_i は、 i 番目の観測値 y_i と i 番目のあてはめ値 $\hat{y}_i = \hat{f}(x_i)$ の差として定義されます。なすわち、 $r_i = y_i - \hat{y}_i$ となります。最小二乗法により、残差の合計 $\sum_{i=1}^n r_i^2$ を最小にする一組のあてはめ値が得られます。

対象の反応変量が連続でない場合には、ロジスティック回帰／プロビット回帰／対数線形（ポアソン）回帰／一般化線形モデルが適していることがあります。予測変量が反応変量に非線形に影響を及ぼす場合は、非線形回帰／局所回帰／一般化加法モデルが適していることがあります。データが外れ値を含むか、誤差がガウス分布でない場合は、ロバスト回帰が適していることがあります。カテゴリカル変量の影響を知りたい場合は、ANOVA が適していることがあります。観測値に相関関係があるかランダムな影響がある場合は、一般化最小二乗モデルまたは混合効果モデルが適していることがあります。

線形回帰に関連する他のダイアログは、**Stepwise Linear Regression**（ステップワイズ法）、**Compare Models**（モデルの比較）および**Multiple Comparisons**（分散分析メニューより選択）です。**Stepwise Linear Regression** ダイアログは、段階的モデル選択法を使用してモデルにどの変量を含めるかを示します。**Compare Models** は、いくつかのモデルのうちのどれが最も適しているかを調べる検定を行います。**Multiple Comparisons** は、線形回帰またはANOVA内のカテゴリカル予測変量の効果を計算します。

線形回帰モデルをあてはめる

メインメニューから**統計 ▶ 回帰 ▶ 線形**を選択します。図 8.34 に示すように、**Linear Regression** ダイアログが開きます。

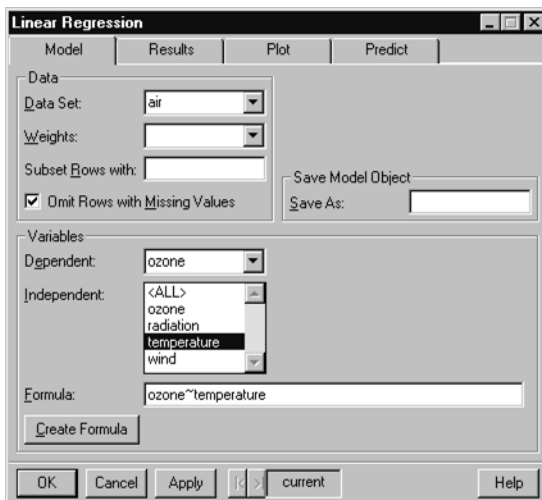


図 8.34 : Linear Regression ダイアログ

例

例のデータセット air で大気汚染データを調べます。これは、111 の観測値（行）と 4 つの変量（列）からなるデータセットです。これは、連続 111 日間にわたってオゾン／太陽放射線／気温／風速の 4 つの変量を測定した環境調査から得たものです。最初に、図 8.35 に示すように、air の変量 temperature および ozone の散布図を作成します。

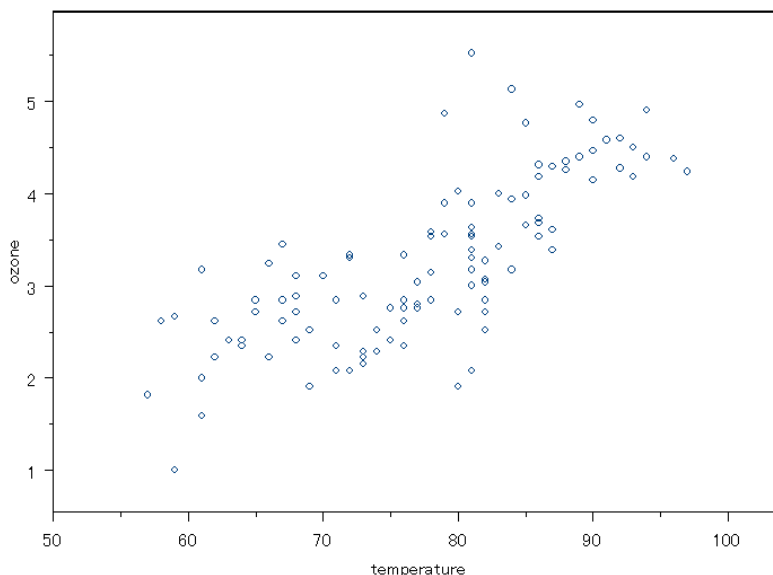


図 8.35 : ozone と temperature の関係を示す散布図

この散布図から、気温とオゾン濃度に線形関係があるという仮説をたてます。反応変数として ozone、1つの予測変数として temperature を選択します。反応変数と予測変数の選択は、統計的見地ではなくデータそのものの関係により行われます。

1. P.348 の方法で **Linear Regression** ダイアログを開きます。
2. **Data Set** フィールドに air を入力します。
3. **Formula** フィールドに ozone ~ temperature を入力します。あるいは、**Dependent** 変数として ozone、**Independent** 変数として temperature を入力します。式を生成する第 3 の方法として、**Create Formula** ボタンをクリックし、**Response** 変数として ozone、**Main Effect** として temperature を選択します。**Create Formula** ボタンを使用して複雑な線形モデルを作成し、モデル指定の表記を学習することができます。モデル式作成の詳細は、オンラインヘルプを参照してください。
4. **Linear Regression** ダイアログの **Plot** ページに進み、7つの診断プロットのチェックボックスをチェックします。
5. **OK** をクリックして、線形回帰を行います。

S-PLUS は、7 つの診断プロットを含むグラフシートを生成します。グラフシートの下にある 7 つのページタブをクリックすることによって、これらのプロットを表示させることができます。このプロットは、図 8.36 に示すものと似ています。S-PLUS は、レポート・ウィンドウに線形回帰の結果を表示します。

```

*** Linear Model ***
Call: lm(formula = ozone ~ temperature, data = air,
na.action = na.exclude)
Residuals:
    Min       1Q   Median       3Q      Max
-1.49 -0.4258  0.02521  0.3636  2.044

Coefficients:
                Value Std. Error  t value Pr(>|t|)
(Intercept)  -2.2260    0.4614   -4.8243  0.0000
temperature   0.0704    0.0059   11.9511  0.0000

Residual standard error: 0.5885 on 109 degrees of freedom
Multiple R-Squared: 0.5672
F-statistic: 142.8 on 1 and 109 degrees of freedom, the
p-value is 0

```

Coefficients の下の Value 列は線形モデルの係数を示し、次のような推定回帰式として読み取ることができます。

$$\text{ozone} = -2.2260 + 0.0704 \times \text{temperature}$$

列の見出しの Std. Error は、それぞれの係数の推定標準誤差を示しています。Multiple R-Squared の項は、このモデルが ozone のばらつきの約 57% を説明していることを示しています。F-statistic は、推定した ozone の分散と回帰による残差平均二乗の比率です。もし気温とオゾンに関係がない場合は、自由度が 1 と 109 の F 分布に従います。この比率は明らかに大きく、したがって回帰直線の真の勾配はおそらく 0 ではありません。

線形モデルの診断プロット

あてはめた線形回帰モデルはどのくらい有効でしょうか。気温はオゾン濃度の適切な予測変数でしょうか。もっと良いモデルを作成できないでしょうか。これらの疑問は、データを統計モデルで説明しようとするときには不可欠です。モデルをデータに当てはめて終わっただけでは不十分です。そのモデルがデータにどれだけ良くあてはまっているかを評価し、データを十分に説明できない場合にはモデルを修正するか、そのモデルをすべて棄却する準備をしなければなりません。

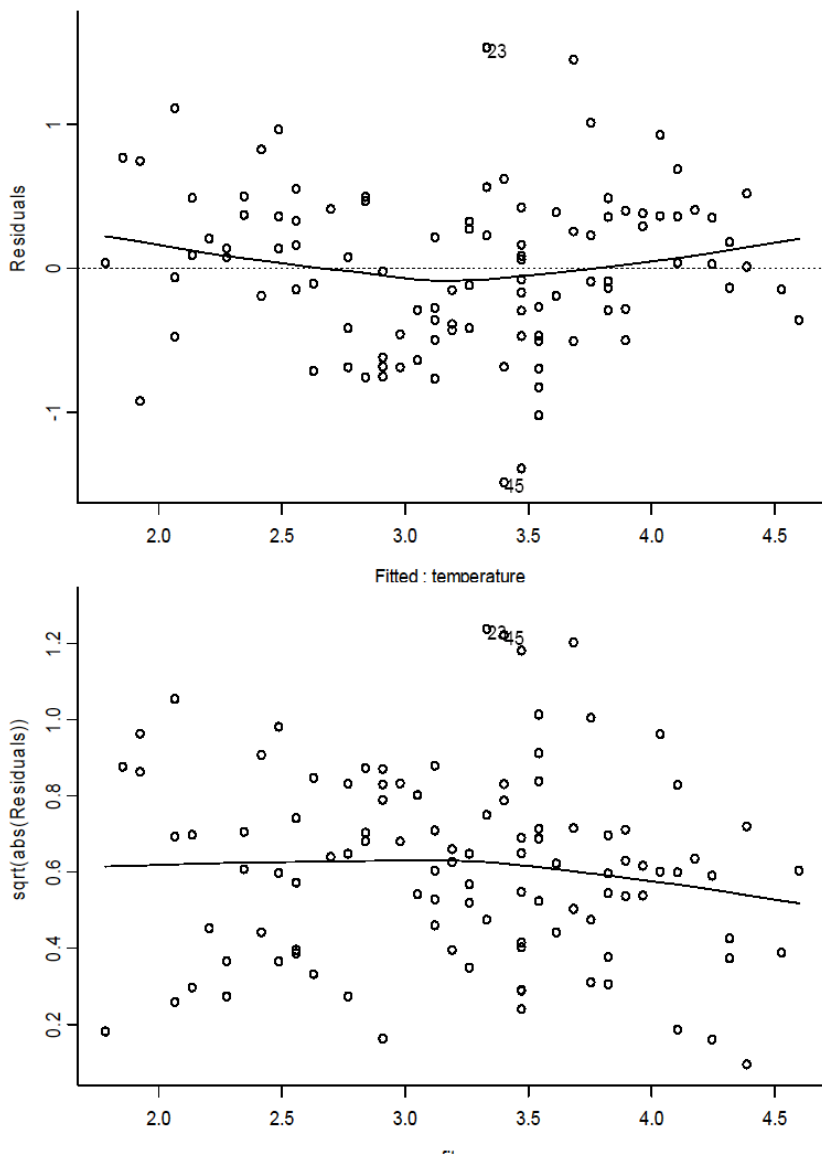


図 8.36 : Linear Regression ダイアログで作成された 7 つの診断プロット

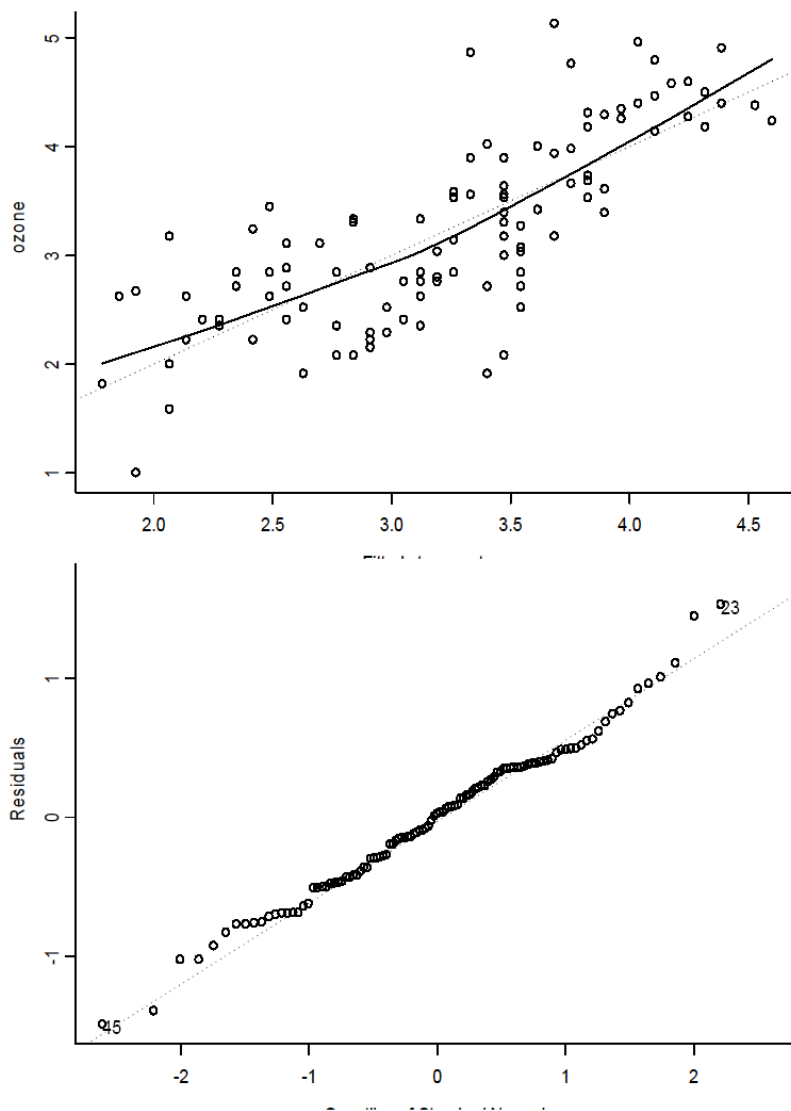


図 8.36 : Linear Regression ダイアログで作成された7つの診断プロット (続き)

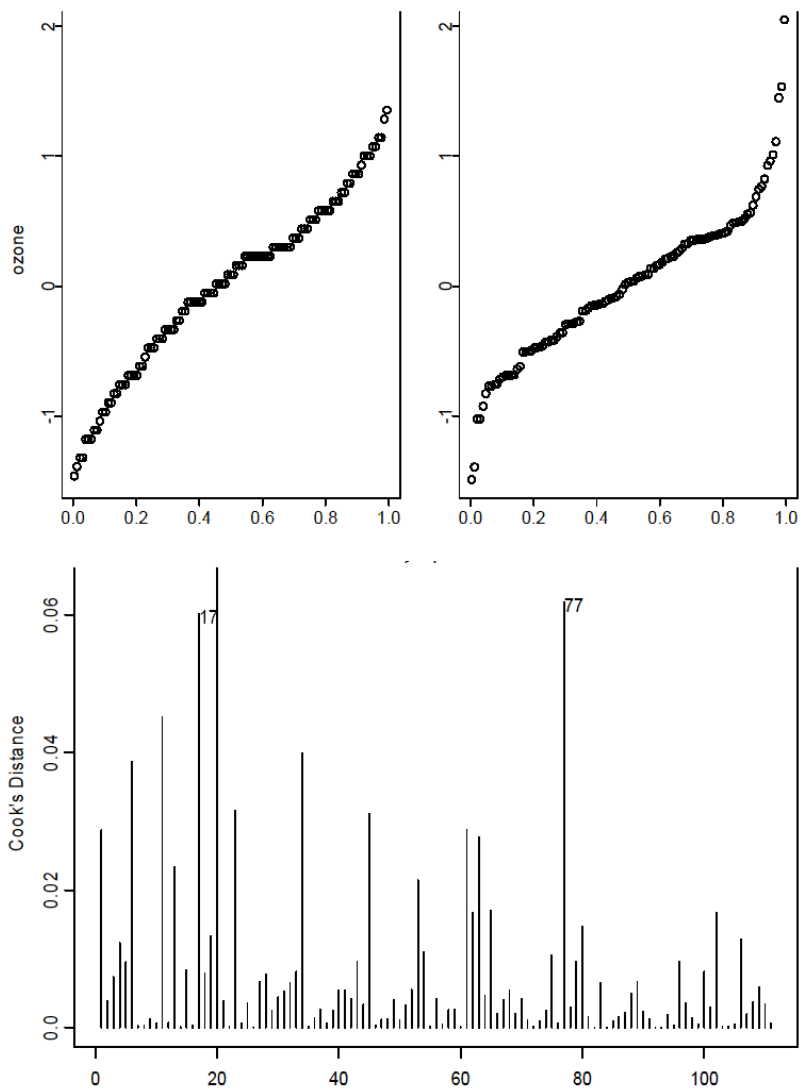


図 8.36 : Linear Regression ダイアログで作成された7つの診断プロット (続き)

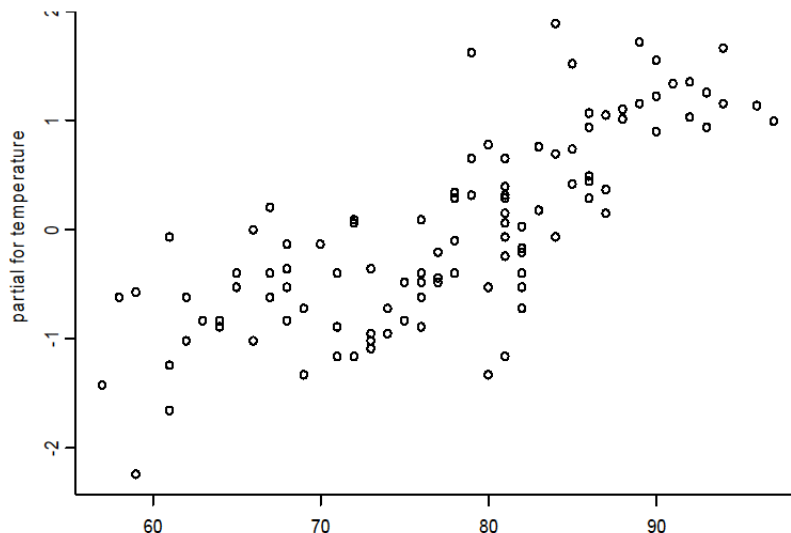


図 8.36 : Linear Regression ダイアログで作成された7つの診断プロット (続き)

あてはめを評価する最も簡単で有益な方法は、モデルの長所と短所を明らかにするプロットの組み合わせを使ってモデルをグラフィカルに調べることです。たとえば、あてはめ値に対する元の反応変量のプロットにより、モデルがデータの概略をどれだけうまく捉えているかがよくわかります。あてはめ値に対する残差のプロットを調べると、より良いモデルでは単なるノイズとして現われるべき残差に残された、説明されていない構造が明らかになります。**Linear Regression** ダイアログのプロットオプションは、これらの2つのプロットと共に次のような有益なプロットを作成します。

- 「あてはめ値に対する残差絶対値の平方根」このプロットは、外れ値を識別し残差の構造を視覚化するのに便利です。
- 「残差の正規確率点プロット」このプロットを使用すると、モデルの誤差が正規分布するという仮定を視覚的に確認することができます。順序付けられた残差が重なったQQ線に沿って集合している場合は、誤差が正規分布であるという確かな証拠になります。

- 「残差あてはめ広がりプロットまたは **r-f** プロット」このプロットは、あてはめ値の広がりや残差の広がりと比較します。モデルはデータのばらつきを説明しようとするものなので、あてはめ値の広がりや残差の広がりよりもはるかに大きいことが期待されます。
- 「クックの距離プロット」クックの距離は、個々の観測値が回帰係数に及ぼす影響の尺度です。
- 「偏残差プロット」偏残差プロットは、 $r_i = b_k x_{ik}$ と x_{ik} の関係を表すプロットで、 r_i は i 番目の観測値の残差、 x_{ik} は k 番目の予測変量の i 番目の観測値、 b_k は k 番目の予測変量の回帰係数推定値です。偏残差プロットは反応量と予想に用いられる変量間の非線形性を検出し、過度に大きい残差の原因を識別するのに役立ちます。

$y = \hat{y}$ は、図 8.36 の上段の 3 番目のプロットに点線として示されています。単回帰の場合、この線は回帰直線と視覚的に同じものです。回帰直線は、データの傾向を適切にモデル化して表しています。残差プロット (図 8.36 のはじめの 2 つのプロット) は、はっきりしたパターンを示していませんが、5 つの観測値は外れ値のように見えます。デフォルトでは、残差プロットとクックの距離プロットのそれぞれにおいて、最も極端な 3 つの値がデータ番号をつけて識別されています。

もう 1 つの有効な診断プロットは、残差の正規プロット (352 ページ下のプロット) です。正規プロットは、残差が確かに正規分布していることを示しています。一方、**r-f** プロット (353 ページ上のプロット) は、このモデルの欠点を示しています。欠点は、残差の広がりや実際にあてはめたデータの広がりよりも大きくなっているという点です。しかし 5 つの外れた残差を除けば、残差があてはめたデータよりも緊密に固まっています。

クックの距離プロットは、きわめて影響力の大きい 4 つか 5 つの観測値を示しています。回帰直線がデータをうまくあてはめ、回帰が有効で残差が正規分布しているように見えるため、与えられた気温に対するオゾン濃度を推定する方法として回帰直線を使用するのは妥当であるようですが、回帰直線がデータのばらつきの 57%しか説明していないという大きな問題が残ります。オゾン濃度に関する他の変数の影響を考慮することにより、多少うまく説明することができるかもしれません。

ロバスト MM 回帰

ロバスト回帰モデルは、データのランダムなばらつき（誤差）がガウス（正規）分布ではないときや、データが大きい外れ値を含むときに線形関係をあてはめるのに役立ちます。そのような状況では、通常の線形回帰は間違った結果を返すことがあります。

「ロバスト MM 回帰」法は、通常の線形回帰モデルの構造とほとんど同じモデルを返します。そのため、ロバストモデルにおいても分かりやすいプロットとサマリーを作成することができます。MM 法は、現在 *Insightful* が推奨しているロバスト回帰法です。

ロバスト MM 回帰を実行する

メインメニューから、**統計 ▶ 回帰 ▶ ロバスト MM** を選択します。図 8.37 に示すように、**Robust MM Linear Regression** ダイアログが開きます。

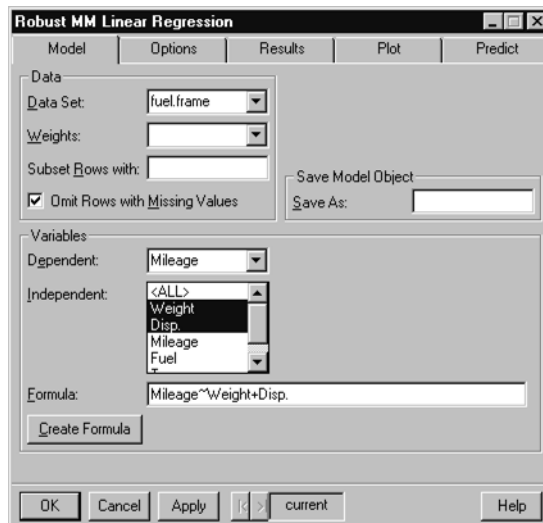


図 8.37 : Robust MM Linear Regression ダイアログ

例

データセット `fuel.frame` は、『Consumer Reports』1990年4月号から引用したものです。60の観測値（行）と5つの変量（列）を含みます。60の車のそれぞれについて、重量、エンジン排気量、走行距離、型、および燃料消費率を調べました。`fuel.frame` データにおいて、ロバストMM回帰を使って、`Weight` と `Disp.` により `Mileage` を予測します。

1. P.356の方法で **Robust MM Linear Regression** ダイアログを開きます。
2. **Data Set** フィールドに `fuel.frame` を入力します。
3. **Formula** フィールドに `Mileage~Weight+Disp.` を入力します。あるいは、**Dependent** 変量として `Mileage` を選択し、**Independent** 変量として CTRL-クリックして `Weight` と `Disp.` を選択します。式を生成する3つ目の方法として、**Create Formula** ボタンをクリックし、**Response** として `Mileage` を選択し、**Main Effects** として CTRL-クリックして `Weight` と `Disp.` を選択します。**Create Formula** ボタンを使って複雑な線形モデルを作成し、モデル指定の表記を学習することができます。式の作成の詳細は、オンラインヘルプを参照してください。
4. **OK** をクリックしてロバストMM回帰モデルをあてはめます。

モデルのサマリーがレポート・ウィンドウに表示されます。また、メッセージ・ウィンドウに、最初と最終の推定値に関する警告が表示されることがあります。この警告メッセージの詳細は、『Guide to Statistics, Volume 1』の第11章「Robust Regression」を参照してください。

ロバストLTS 回帰

「ロバストLTS回帰」法は、刈り込み最小二乗回帰を実行します。このプロットとサマリーは、通常の線形回帰やロバストMM回帰に較べると詳細なものではありません。

ロバスト LTS 回帰を実行する

メインメニューから、統計 ▶ 回帰 ▶ ロバスト LTS を選択します。図 8.38 に示すように、**Robust LTS Linear Regression** ダイアログが開きます。

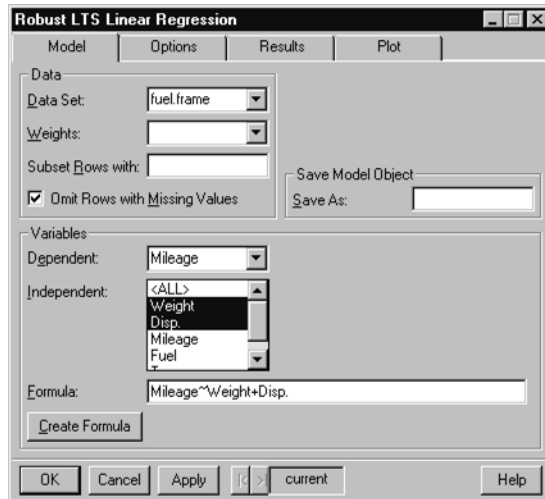


図 8.38 : Robust LTS Linear Regression ダイアログ

例

データ fuel.frame で、ロバスト LTS 回帰を使って、Mileage を Weight と Disp. で予測します。

1. **Robust LTS Linear Regression** ダイアログを開きます。
2. **Data Set** フィールドに fuel.frame と入力します。
3. **Formula** フィールドに Mileage~Weight+Disp. と入力します。あるいは、**Dependent** 変数として Mileage を選択し、**Independent** 変数として CTRL-クリックして Weight と Disp. を選択します。式を生成する 3 つ目の方法として、**Create Formula** ボタンをクリックし、**Response** 変数として Mileage を選択し、**Main Effects** として CTRL-クリックして Weight と Disp. を選択します。**Create Formula** ボタンを使って複雑な線形モデルを作成し、モデル指定の表記を学習することができます。式の作成の詳細は、オンラインヘルプを参照してください。
4. **OK** をクリックしてロバスト LTS 回帰モデルをあてはめます。

レポート・ウィンドウにモデルのサマリーが表示されます。

ステップ ワイズ線形 回帰

モデル化プロセスのステップの 1 つは、回帰モデルに含める変量を決定することです。「ステップワイズ線形回帰」は、回帰モデルに含める変量を選択するための自動化された手順です。前進ステップワイズ回帰は、適合度が改善されなくなるまで項をモデルに追加します。各ステップで、あてはめを最も改善する項が追加されます。後進ステップワイズ回帰は、項が適合度を大きく低下させない範囲でモデルから項を削除します。各ステップで、あてはめの低下が最も小さい項が削除されます。またステップワイズ回帰には、項を追加するか削除するか、前進と後退の両方を行うオプションがあります。これがデフォルトで使用されます。

ステップワイズ線形回帰を実行する

メインメニューから、統計 ▶ 回帰 ▶ ステップワイズ法を選択します。図 8.39 に示すように、**Stepwise Linear Regression** ダイアログが開きます。

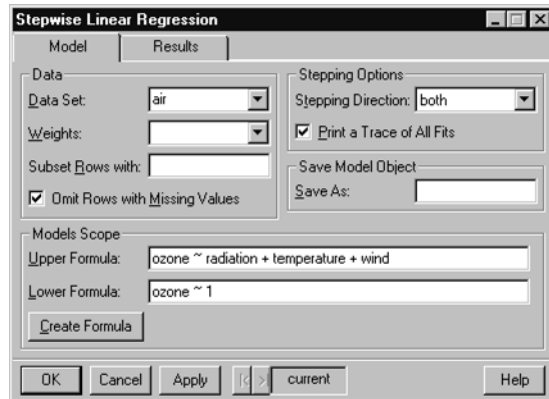


図 8.39 : Stepwise Linear Regression ダイアログ

例

P.359 の方法で air データにステップワイズ回帰を適用します。

1. **Stepwise Linear Regression** ダイアログを開きます。
2. **Data Set** フィールドに air と入力します。
3. 最も複雑なモデルを表す式を指定しなければなりません（モデルの最上限）。**Upper Formula** として、`ozone ~ radiation + temperature + wind` を指定します。

4. 最も単純なモデルを表す式を指定しなければなりません (モデルの最下限)。**Lower Formula** として、`ozone ~ 1` を指定します。1 は、単に切片を意味します。
5. **OK** をクリックします。

ステップワイズ回帰は、適合度を表す選択基準として **Cp** 規準を使用します。これは、モデルの複雑さをペナルティとし、モデルの精度を報奨として計算された統計値です。この例では、どの項を削除した場合よりもフルモデルのものが最も小さい **Cp** 統計値を持つモデルになります。したがって、フルモデルが最良のモデルとして選択されます。

ステップのサマリーがレポート・ウィンドウに表示されます。

```

*** Stepwise Regression ***

*** Stepwise Model Comparisons ***
Start:  AIC= 29.9302
       ozone ~ radiation + temperature + wind

Single term deletions

Model:
ozone ~ radiation + temperature + wind

scale:  0.2602624

           Df Sum of Sq      RSS      Cp
<none>                27.84808 29.93018
 radiation  1    4.05928 31.90736 33.46893
 temperature 1   17.48174 45.32982 46.89140
      wind   1    6.05985 33.90793 35.46950

*** Linear Model ***

Call:  lm(formula = ozone ~ radiation + temperature + wind,
data = air, 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
 radiation    0.0022   0.0006    3.9493  0.0001
 temperature  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

一般化加法 モデル

「一般化加法モデル」は、モデルに予測変量と反応変量のフレキシブルな加法非線形関係を加えることによって線形モデルを拡張します。線形モデルは各予測変量と、反応変量の関係が線形であることを仮定していますが、加法モデルは反応変量が各予測変量の滑らかな関数から影響を受けることだけを仮定しています。反応変量は予測変量の平滑関数の和としてモデル化され、平滑関数はスモーカーを使用して自動的に予測されます。加法モデルは最終的な近似を得たり、標準的な線形モデルにはどのタイプの変数変換が適しているかを探索したりするのに役立つことがあります。

加法モデルをあてはめる

メインメニューから統計 ▶ 回帰 ▶ 一般加法モデル(GAM)を選択します。図 8.40 に示すように、Generalized Additive Models ダイアログが開きます。

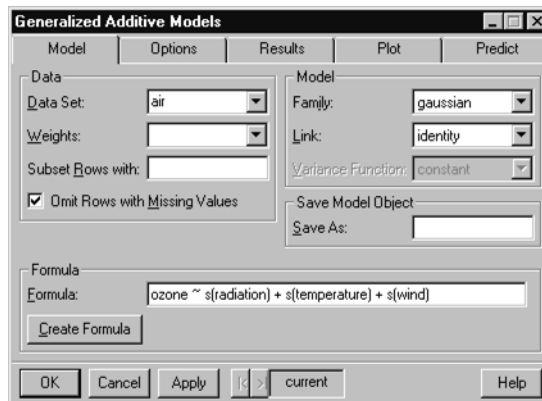


図 8.40 : Generalized Additive Models ダイアログ

例

air データに加法モデルをあてはめます。

1. P.361 の方法で **Generalized Additive Models** ダイアログを開きます。
2. **Data Set** フィールドに air と入力します。
3. **Formula** フィールドで $\text{ozone} \sim \text{s}(\text{radiation}) + \text{s}(\text{temperature}) + \text{s}(\text{wind})$ を指定します。
4. ダイアログの **Plot** ページで、**Partial Residuals** と **Include Partial Fits** のチェックボックスを選択します。これで、予測変量ごとの偏残差と偏回帰のプロットを作成することができます。
5. **OK** をクリックします。

加法モデルのサマリーが **レポート・ウィンドウ** に表示されます。各ページに偏残差プロットが 1 つあるマルチページの **グラフシート** が現われます。

局所回帰

「局所回帰」は、多変量の多項式回帰のノンパラメトリックな一般化です。一般的な平滑面をあてはめる 1 つの方法と考えることができます。面の形を指定する様々なオプションを利用することができます。

局所回帰をあてはめる

メインメニューから**統計 ▶ 回帰 ▶ 局所**を選択してください。図 8.41 に示すように、**Local (Loess) Regression** ダイアログが開きます。

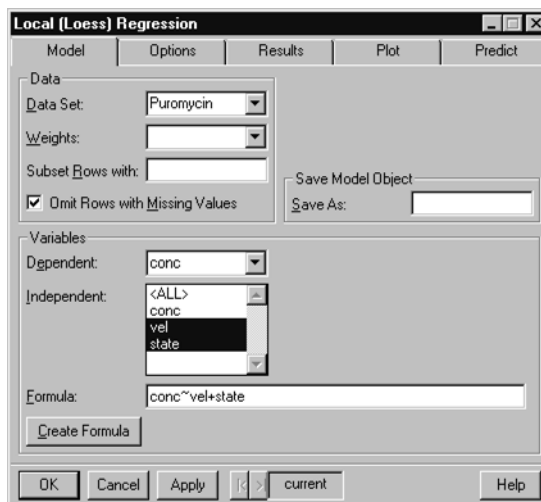


図 8.41 : Local (Loess) Regression ダイアログ

例

データセット Puromycin には、6 つの異なる基質濃度と 2 つの異なる細胞処理の生化学反応の初期速度の測定値を表す 23 行があります。364 ページの「非線形回帰」で、これらのデータについて詳細に説明し、データの理論モデルについて考察します。理論モデルをあてはめる前に、**Local (Loess) Regression** ダイアログを使用して、データにノンパラメトリック平滑曲線をあてはめることができます。

モデルは、処理群ごとの個別の曲線から成ります。vel と state で反応 conc を予測します。state は因子なので、これは state の水準ごとに vel の別々の平滑曲線をあてはめます。

1. P.363 の方法で **Local (Loess) Regression** ダイアログを開きます。
2. **Data Set** フィールドに Puromycin と入力します。

3. **Formula** フィールドに `conc~vel+state` と入力します。あるいは、**Dependent** 変量として `conc` を選択し、CTRL-クリックして **Independent** 変量として `vel` と `state` を選択します。式を生成する第3の方法として、**Create Formula** ボタンをクリックし、**Response** 変量として `conc` を選択し、CTRL-クリックして **Main Effects** として `vel` と `state` を選択します。**Create Formula** ボタンを使って複雑な線形モデルを作成したり、モデル指定の表記を学習したりすることができます。式の作成の詳細は、オンラインヘルプで説明しています。
4. ダイアログの **Plot** ページで、**Cond. Plots of Fitted vs Predictors** を選択します。このタイプのプロットは、予測変量のいくつかのサブセットに関して反応量の個別のプロットを表します。このケースでは、`state` の水準ごとに個別の曲線をプロットします。
5. **OK** をクリックします。

Loess (局所重み付き多項式回帰) モデルのサマリーがレポート・ウィンドウに表示され、**グラフシート**に条件付きプロットが表示されます。

非線形回帰

「非線形回帰」は、特定の非線形関係を使用して、1つまたは複数の予測変量から連続変量を予測します。非線形関係の形は、通常、そのデータが得られた状況に固有の理論モデルから規定されます。

Nonlinear Regression ダイアログは、非線形回帰モデルをあてはめます。非線形回帰を使用するときは、**S-PLUS** シンタックスによってモデル式を指定し、パラメータ推定のための初期値を与える必要があります。

非線形最小二乗回帰をあてはめる

メインメニューから、統計 ▶ 回帰 ▶ 非線形を選択します。図 8.42 に示すように、**Nonlinear Regression** ダイアログが開きます。

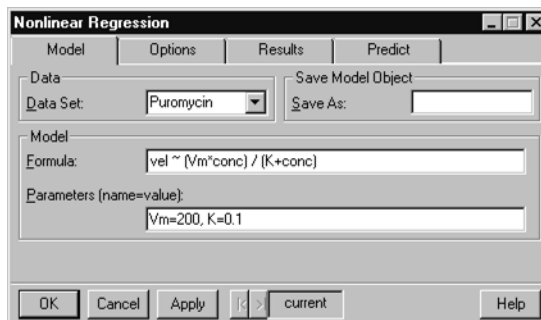


図 8.42 : Nonlinear Regression ダイアログ

例

データセット Puromycin には、6つの異なる基質濃度と2つの異なる細胞処理の生化学反応の初期速度の測定値を表す 23 行があります。図 8.43 は、2つの処理群（処理した群と処理しない群）における濃度と反応速度の関係を異なる記号でプロットしています。

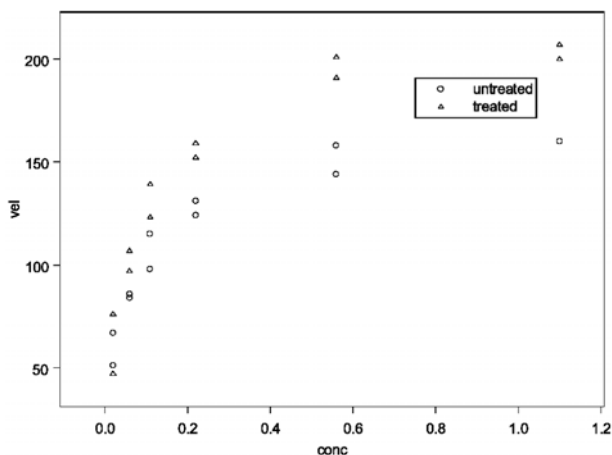


図 8.43 : Puromycin データの散布図

速度と濃度の関係は、次のようなミカエリス - メンテン (Michaelis-Menten) の関係に従うことが知られています。

$$V = \frac{V_{max}c}{K + c} + \varepsilon$$

ここで、 V は初期反応速度、 c は酵素基質濃度、 V_{max} は c を無限大にしたときの漸近速度を表すパラメータ、 K はミカエリスパラメータ、 ε は実験誤差です。薬による処理によって、 K は変化せず V_{max} が変化することを仮定しています。最適化目的関数は次の通りです。

$$S(V_{max}, K) = \sum \left(V_i - \frac{V_{max} + \Delta V_{max} I_{\{treated\}}(state) c_i}{K + c_i} \right)^2$$

ここで、 $I_{\{treated\}}$ は細胞をプロマイシン (Puromycin) で処理したかどうかを示す関数です。

最初に、1つの曲線を両群にあてはめる単純なモデルのあてはめを行います。次に、処理の影響を反映させて項を追加します。

非線形回帰モデルをあてはめるためには、非線形モデルの形、データセット名、およびパラメータ推定のための初期値を指定しなければなりません。図 8.43 の考察は、すべての観測値を1つのグループとして扱うためには初期値として $v=200$ と $K=0.1$ を示唆しています。速度と濃度の間のミカエリス - メンテンの関係をあてはめるには、

1. P.365 の方法で **Nonlinear Regression** ダイアログを開きます。
2. **Data Set** フィールドに Puromycin と入力します。
3. ミカエリス - メンテンの関係 $vel \sim (V_m * conc) / (K + conc)$ を **Formula** フィールドに入力します。
4. パラメータ開始値 $V_m=200$, $K=0.1$ を、**Parameters** フィールドに入力します。
5. **OK** をクリックします。

次のような結果がレポート・ウィンドウに現われます。

```
*** Nonlinear Regression Model ***
Formula: vel ~ (Vm * conc)/(K + conc)
Parameters:
      Value Std. Error  t value
Vm 190.8050000  8.7644700 21.77030
K   0.0603863   0.0107682  5.60785
Residual standard error: 18.6146 on 21 degrees of freedom
Correlation of Parameter Estimates:
      Vm
K 0.776
```

表示された結果は、パラメータの推定値と標準誤差と t 値、ならびに残差の標準誤差とパラメータ推定値の相関を示しています。

次に、処理の影響を含むモデルをあてはめます。

1. P.365 の方法で **Nonlinear Regression** ダイアログを開きます。
2. **Data Set** フィールドに Puromycin と入力します。
3. **Formula** フィールドに、ミカエリス - メンテンの関係

$$vel \sim ((Vm + delV * (state == "treated")) * conc) / (K + conc)$$
 を入力します。
4. 図 8.43 は初期値として $Vm=160$ と $delV=40$ を示唆しています。一方、前のモデルは $K=0.05$ でした。**Parameters** フィールドに、初期値として $Vm=160$, $delV=40$, $K=0.05$ を入力します。
5. **OK** をクリックします。

次のような結果がレポート・ウィンドウに現われます。

```
*** Nonlinear Regression Model ***
Formula: vel ~ ((Vm + delV * (state == "treated")) * conc) /
(K + conc)
Parameters:
      Value Std. Error  t value
Vm 166.6010000  5.80726000 28.68840
delV 42.0245000  6.27201000  6.70032
K    0.0579659  0.00590968  9.80863
Residual standard error: 10.5851 on 20 degrees of freedom
```

Correlation of Parameter Estimates:

```

Vm    delV
delV  -0.5410
K     0.6110  0.0644

```

表示された結果は、パラメータの推定値と標準誤差と t 値、ならびに残差の標準誤差とパラメータ推定値の相関を示しています。delV の t 統計量の大きさは、処理が最高速度に影響を及ぼしていることを示しています。

一般化線形モデル

「一般化線形モデル」は、反応量が離散であったり、モデルを標準的な線形モデルと異なる形にあてはめられない状況に対処するように、よく知られた線形回帰モデルを一般化する方法です。最も広く使用される一般化線形モデルは、2 値データが反応量となるロジスティック回帰モデルと度数データが反応量となる対数線形（ポアソン）モデルです。

一般化線形モデルをあてはめる

メインメニューから、統計 ▶ 回帰 ▶ 一般化線形モデル (GLM) を選択します。図 8.44 に示すように、**Generalized Linear Models** ダイアログが開きます。



図 8.44 : Generalized Linear Models ダイアログ

例

データセット `solder` には、プリント回路基板に部品を取り付けるウェーブはんだ付け方法に関連して、異なる 5 つの因子の実験結果である 900 の観測値（行）から成っています。反応変数 `skips` は、外観検査で見られたはんだ飛びの度数です。**Generalized Linear Models** ダイアログを使用して、どのプロセス変量のはんだ飛びの数に影響を及ぼしているかを評価することができます。

1. P.368 の方法で **Generalized Linear Models** ダイアログを開きます。
2. **Data Set** フィールドに `solder` と入力します。
3. **Dependent** 変数として `skips` を、**Independent** 変数リストで **<ALL>** を選択します。これにより、**Formula** フィールドに `skips~.` が生成されます。
4. **Family** として `poisson` を選択します。**Link** は、ポアソンモデルの自然リンク関数である `log` に変化します。
5. **OK** をクリックします。

ポアソン回帰のサマリーが **レポート**・ウィンドウに現われます。

対数線形 (ポアソン) 回帰

度数データは、しばしば対数線形回帰を使用してモデル化されます。「対数線形回帰」では、反応は共変量の値に依存する中心パラメータをもつポアソン分布から生成されると仮定されます。

対数線形の (ポアソン) 回帰をあてはめる

メインメニューから、統計 ▶ 回帰 ▶ ポアソン対数 を選択します。図 8.45 に示すように、**Log-linear (Poisson) Regression** ダイアログが開きます。

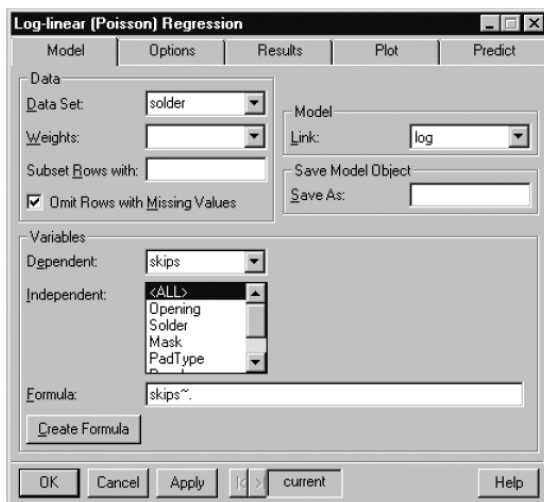


図 8.45 : Log-linear (Poisson) Regression ダイアログ

例

この例では、solder データにポアソン回帰をあてはめます。

1. P.370 の方法で **Log-linear (Poisson) Regression** ダイアログを開きます。
2. **Data Set** フィールドに solder と入力します。
3. **Dependent** 変数に skips を選択し、**Independent** 変数リストで **<ALL>** を選択します。これにより、**Formula** フィールドに skips~. が生成されます。
4. **OK** をクリックします。

対数線形 (ポアソン) 回帰のサマリーがレポート・ウィンドウに現われます。係数の結果テーブルの t 値はかなり大きく、これはすべてのプロセス変数が飛びの生じる数に大きく影響していることを示しています。

ロジスティック回帰

「ロジスティック回帰」は、2水準因子反応変量と、1つまたは複数の予測変量の関係をモデル化します。予測変量の線形結合は最尤推定法を使って求められ、この場合反応変量は、予測変量の値に依存する確率パラメータを持つ二項分布プロセスによって生成されると仮定されます。

ロジスティック回帰をあてはめる

メインメニューから、統計 ▶ 回帰 ▶ ロジスティックを選択します。図 8.46 に示すように、**Logistic Regression** ダイアログが開きます。

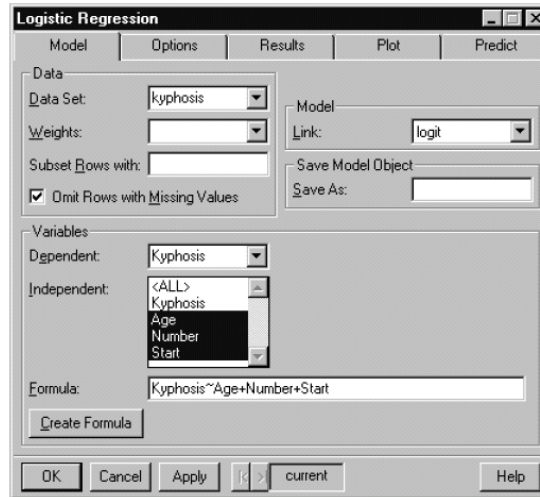


図 8.46 : Logistic Regression ダイアログ

例

データセット `kyphosis` は、脊椎矯正手術を受けた 81 人の子供のデータを表す 81 行から成ります。結果である `Kyphosis` は 2 水準因子変量であり、他の 3 つの変量 `Age`、`Number` および `Start` は数値です。図 8.47 は、`Kyphosis` の各水準に対する `Age`、`Number` および `Start` の箱型図を表します。

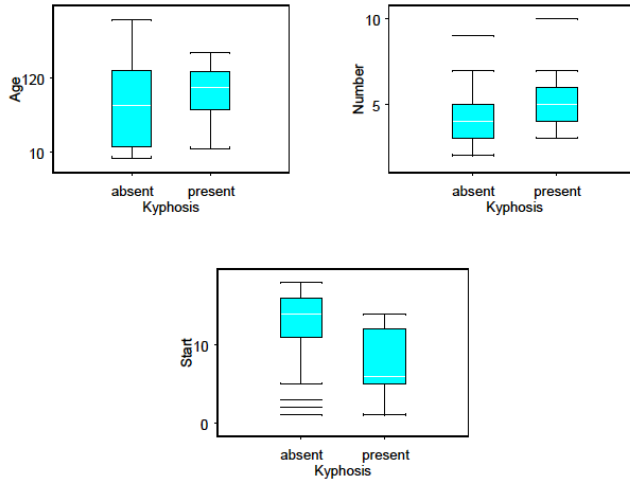


図 8.47 : 脊柱後弯症データの箱型図

脊柱後弯症は手術後の脊椎の変形です。変形が起きるかどうかに共変量がどのように影響するかを調べたいと思います。`Start` と `Number` は両方とも、`Kyphosis` の有無によって大きい位置のずれを示しています。`Age` 変量には、そのような位置のずれがありません。ロジスティック回帰を使用すると、各共変量の変形するかどうかの確率に与える影響を測定することができます。

1. P.371 の方法で **Logistic Regression** ダイアログを開きます。
2. **Data Set** フィールドに `kyphosis` と入力します。
3. **Formula** フィールドで `Kyphosis~Age+Number+Start` を指定します。
4. **OK** をクリックします。

ロジスティック回帰のサマリーがレポート・ウィンドウに表示されます。サマリーには、残差、係数および尤離度に関する情報が含まれます。`Start` に対する高い `t` 値は、`Start` が脊柱後弯症の発生に大きな影響をもつことを示しています。`Age` と `Number` の `t` 値は、反応に対して多大な影響を示すほど大きくありません。

```

*** Generalized Linear Model ***

Call: glm(formula = Kyphosis ~ Age + Number + Start,
  family = binomial(link = logit), data = kyphosis,
  na.action = na.exclude, control = list(
  epsilon = 0.0001, maxit = 50, trace = F))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.312363 -0.5484308 -0.3631876 -0.1658653  2.16133

Coefficients:
                Value Std. Error  t value
(Intercept) -2.03693225  1.44918287  -1.405573
      Age      0.01093048  0.00644419   1.696175
    Number    0.41060098  0.22478659   1.826626
      Start  -0.20651000  0.06768504  -3.051043

(Dispersion Parameter for Binomial family taken to be 1 )

Null Deviance: 83.23447 on 80 degrees of freedom
Residual Deviance: 61.37993 on 77 degrees of freedom
Number of Fisher Scoring Iterations: 5

```

プロビット 回帰

Probit Regression ダイアログは、プロビット反応モデルをあてはめます。これは、二項反応データに適したロジスティック回帰の変形です。

プロビット回帰モデルをあてはめる

メインメニューから、**統計 ▶ 回帰 ▶ プロビットモデル**を選択します。図 8.48 に示すように、**Probit Regression** ダイアログが開きます。

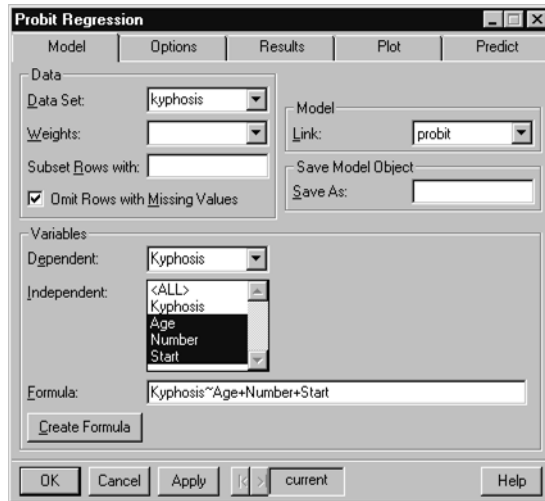


図 8.48 : Probit Regression ダイアログ

例

この例では、kyphosis データセットにプロビット回帰モデルをあてはめます。

1. **Probit Regression** ダイアログを開きます。
2. **Data Set** フィールドに kyphosis と入力します。
3. **Formula** フィールドで `Kyphosis~Age+Number+Start` を指定します。
4. **OK** をクリックします。

モデルのサマリーがレポート・ウィンドウに表示されます。

```
*** Generalized Linear Model ***
```

```
Call: glm(formula = Kyphosis ~ Age + Number + Start,
  family = binomial(link = probit), data = kyphosis,
  na.action = na.exclude, control = list(epsilon =
  0.0001, maxit = 50, trace = F))
```

```
Deviance Residuals:
```

```
      Min       1Q   Median       3Q      Max
-2.217301 -0.5440968 -0.3535132 -0.124005  2.149486
```

```
Coefficients:
```

```
              Value Std. Error t value
(Intercept) -1.063353291  0.809886949 -1.312965
```

```
Age 0.005984768 0.003507093 1.706475  
Number 0.215179016 0.121687912 1.768286  
Start -0.120214682 0.038512786 -3.121423
```

(Dispersion Parameter for Binomial family taken to be 1)

Null Deviance: 83.23447 on 80 degrees of freedom

Residual Deviance: 61.0795 on 77 degrees of freedom

Number of Fisher Scoring Iterations: 5

分散分析

分散分析 (ANOVA) は、一般に連続反応に対する 1 つまたは複数のカテゴリカル変量の影響を調べるために使用されます。

固定効果 ANOVA

ANOVA ダイアログは、伝統的な固定効果の分散分析を行います。

固定効果 ANOVA モデルをあてはめる

メインメニューから、統計 ▶ 分散分析 ▶ 固定効果を選択します。図 8.49 に示すように、ANOVA ダイアログが開きます。

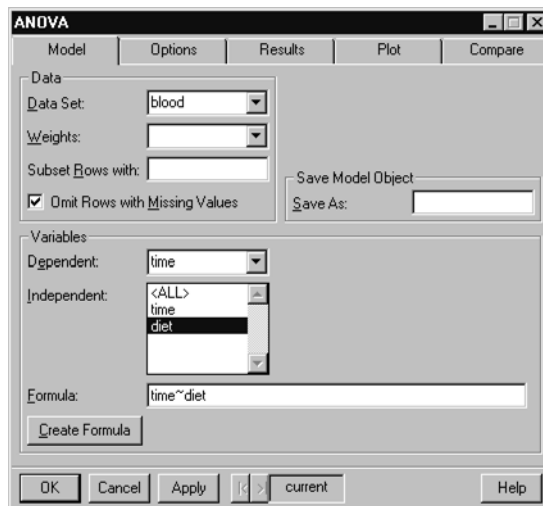


図 8.49 : ANOVA ダイアログ

例

311 ページの「一元分散分析」では、表 8.2 に示した blood データに単純な一元 ANOVA を実行しました。これらのデータは、4 つの異なる食物の血液凝固時間を示しています。一般に、ANOVA ダイアログは、一元 ANOVA ダイアログよりもはるかに複雑な計画を処理することができます。さらに診断プロットを生成して、分析結果に関するより多くの情報を提供します。ここでは、ANOVA ダイアログを使って前の例の結果を再現します。また、作成したモデルがデータにどれだけ適しているかを確認する診断プロットを生成します。

1. blood データセットをまだ作成していない場合は、313 ページの指示にしたがって blood データセットを作成してください。
2. P.376 の方法で **ANOVA** ダイアログを開きます。
3. **Data Set** として blood と入力します。
4. 行おうとしている一元 ANOVA の **モデル式** time ~ diet を入力します。あるいは、**Dependent** 変量として time、**Independent** 変量として diet を選択します。式を生成する第 3 の方法として、**Create Formula** ボタンをクリックし、**Response** 変量として time、**Main Effect** として diet を選択します。**Create Formula** ボタンを使って複雑な線形モデルを作成したり、モデル指定の表記を学習したりすることができます。式の作成の詳細は、オンラインヘルプで説明しています。
5. **Plot** ページをクリックし、可能な 7 つのプロットをすべてクリックします。
6. **OK** をクリックして解析を行います。

S-PLUS は、7 つの診断プロットを生成します。グラフシートの下にある 7 つのページタブをクリックして、これらのプロットを表示させることができます。このモデルでは、プロットに特に大きな問題は現れません。レポート・ウィンドウに ANOVA の結果が表示されます。

ランダム効果 ANOVA

「ランダム効果 ANOVA」は、処理効果がランダムになっている釣り合いのとれた計画実験に使用されます。モデルは釣り合いがとれていて、完全にランダムでなければなりません。単層計画だけが可能です。

混合効果モデルには、**Linear Mixed Effects** ダイアログを使用してください。

ランダム効果 ANOVA モデルをあてはめる

メインメニューから、統計 ▶ 分散分析 ▶ ランダム効果を選択します。図 8.50 に示すように、**Random Effects Analysis of Variance** ダイアログが開きます。

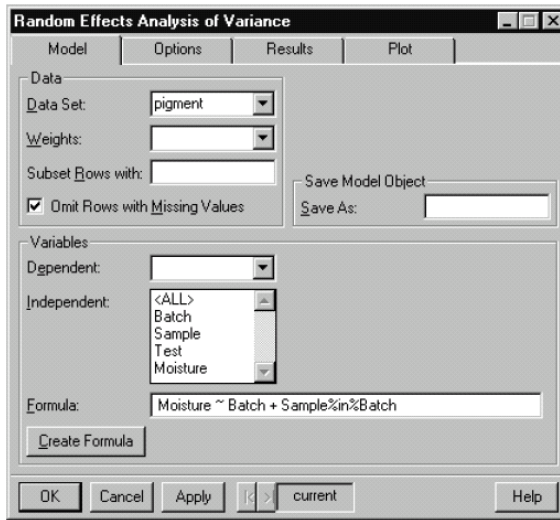


図 8.50 : Random Effects Analysis of Variance ダイアログ

例

データセット pigment は、60 行と 4 列から成っています。行は、各ロットから抜き出した 2 つの標本の 15 ロットの顔料を表し、各標本に 2 回の分析を行いました。これらのデータは、ロット内にネストされた標本に対する含水量の計画実験によるものです。ここでは、ランダム効果 ANOVA モデルをあてはめてロット内のばらつきとロット間のばらつきを評価します。

1. P.377 の方法で **Random Effects Analysis of Variance** ダイアログを開きます。
2. **Data Set** フィールドに pigment と入力します。
3. 次の**モデル式**を入力します。

Moisture ~ Batch + Sample %in% Batch

あるいは、**Create Formula** ボタンを使ってモデル式を作成します。**Term Category** から **nested effect** を選ぶことができる **Special Terms** グループを使って、項 Sample %in% Batch を作成します。

4. **OK** をクリックします。

モデルのサマリーが**レポート・ウィンドウ**に表示されます。

多重比較

分散分析モデルは、一般に反応量に対するいくつかの処理の効果を比較するために使用されます。分散分析モデルをあてはめた後で様々な処理群ごとの反応に有意な差があるかどうかを判定し、違いがある場合はその差の大きさを推定したいことがあります。「多重比較」は、処理効果の影響が等しいかどうかと、その影響の推定を検定します。

Multiple Comparisons ダイアログは、固定効果線形モデルの予想変量の各水準の差に関して、同時信頼区間または非同時信頼区間を計算します。分散分析モデル (aov) または線形モデル (lm) のオブジェクト名と対象とする効果を指定する必要があります。

また、多重比較機能は、ANOVA ダイアログの **Compare** ページでも使用することができます。

多重比較を実行する

メインメニューから、統計 ▶ 分散分析 ▶ 多重比較を選択します。図 8.51 に示すように、**Multiple Comparisons** ダイアログが開きます。

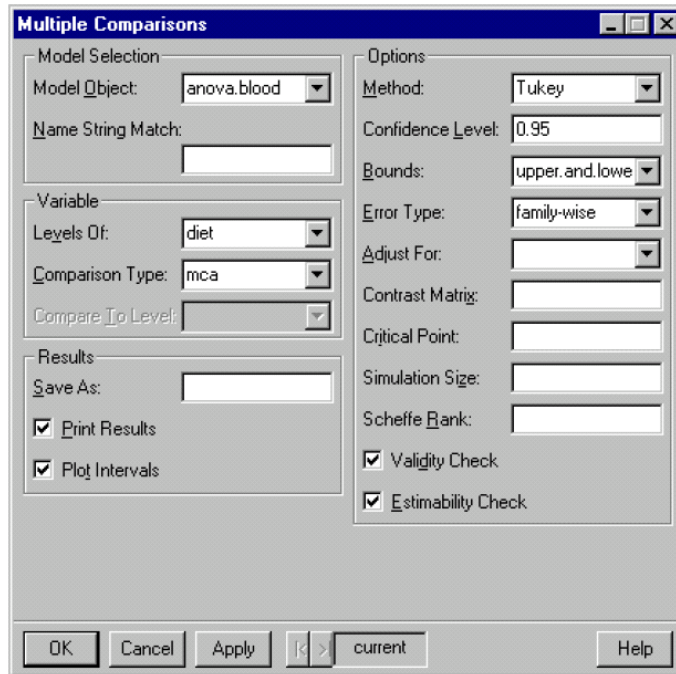


図 8.51 : Multiple Comparisons ダイアログ

例

311 ページの一元分散分析では、表 8.2 に示した blood データセットに単純な一元 ANOVA を実行しました。これらのデータは、4 つの異なる食物の血液凝固時間を示します。376 ページの固定効果 ANOVA では、blood データセットを再び取り上げて、食物が血液凝固時間に影響を及ぼすと結論づけました。次のステップは、食物ごとの凝固時間が互いに異なることを確認するためにすべての水準の組み合わせ間の同時信頼区間を生成することです。これは、ANOVA ダイアログの **Compare** ページまたは **Multiple Comparisons** ダイアログを使用して行うことができます。

1. blood データセットをまだ作成していない場合は、313 ページの指示にしたがって blood データセットを作成してください。
2. 一元分散分析をまだ行ってない場合は、311 ページの一元分散分析を実行し、その結果をオブジェクト anova.blood に保存してください。
3. P.379 の方法で **Multiple Comparisons** ダイアログを開きます。

4. プルダウンメニューから **Model Object** として `anova.blood` を選択します。
5. テューキーの多重比較法を使用して `diet` の水準を比較したいと思います。 **Levels Of** のプルダウンメニューから `diet` を選択し、 **Method** を **Tukey** に設定します。
6. **OK** をクリックして多重比較を生成します。

レポート・ウィンドウに次のような結果が表示されます。

```
95 % simultaneous confidence intervals for specified
linear combinations, by the Tukey method
```

```
critical point: 2.7987
response variable: time
```

```
intervals excluding 0 are flagged by '****'
```

	Estimate	Std.Error	Lower Bound	Upper Bound	
A-B	-5.00e+000	1.53	-9.28	-0.725	****
A-C	-7.00e+000	1.53	-11.30	-2.720	****
A-D	-8.93e-014	1.45	-4.06	4.060	
B-C	-2.00e+000	1.37	-5.82	1.820	
B-D	5.00e+000	1.28	1.42	8.580	****
C-D	7.00e+000	1.28	3.42	10.600	****

上の結果と信頼区間のプロットから、食物 A、D グループと食物 C、D グループの間では、血液凝固時間が大きく異なることが分かります。

混合効果モデル

混合効果モデルは、固定効果とランダム効果の両方を含む回帰モデル、または分散分析モデルです。

線形

Linear Mixed Effects Models ダイアログは、Laird と Ware の式 (1982) で線形混合効果モデルをあてはめますが、ネストされたランダム効果を考慮しています。

線形混合効果モデルをあてはめる

メインメニューから、統計 ▶ 混合効果モデル ▶ 線形を選択します。図 8.52 に示すように、**Linear Mixed Effects Models** ダイアログが開きます。

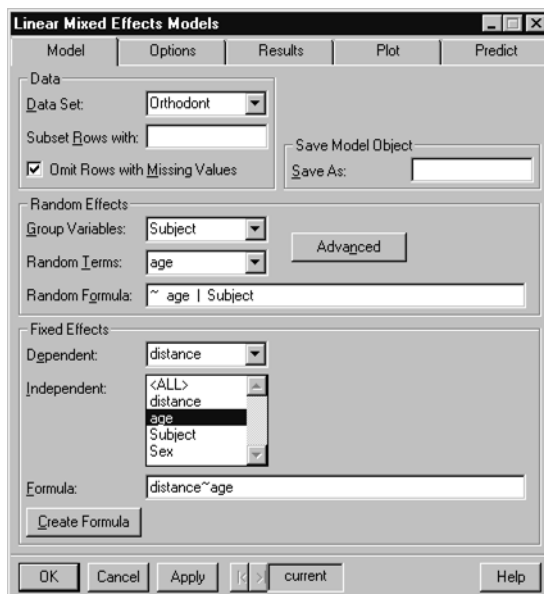


図 8.52 : Linear Mixed Effects Models ダイアログ

例

データセット Orthodont は、4 段階の年齢での少女 11 人と少年 16 人の歯列矯正の実測値を示す 108 行と 4 列から成っています。ここでは、線形混合効果モデルを使用して年齢による距離の違いを調べます。モデルは age の固定効果とランダム効果を含み、Subject は計測個体を識別するものです。

1. P.382 の方法で **Linear Mixed Effects Models** ダイアログを開きます。
2. **Data Set** フィールドに Orthodont と入力します。
3. **Formula** フィールドで distance~age を指定します。
4. **Group Variable** として Subject を、**Random Term** として age を選択します。**Random Formula** フィールドに、~age|Subject が自動的に入ります。
5. **OK** をクリックします。

モデルのサマリーがレポート・ウィンドウに表示されます。S-PLUS が、データセットを groupedData 構造と認識する場合は、**Formula** および **Random Formula** フィールドは、groupedData オブジェクトのメソッド formula によって自動的に埋められます。詳細は、『Guide to Statistics, Volume 1』の第 14 章「Linear and Nonlinear Mixed-Effects Models」を参照してください。

非線形

Nonlinear Mixed Effects Models ダイアログは、Lindstrom と Bates (1990) において記述された式に非線形混合効果モデルをあてはめますが、ネストされたランダム効果を考慮します。

非線形混合効果モデルをあてはめる

メインメニューから、統計 ▶ 混合効果モデル ▶ 非線形を選択します。図 8.53 に示すように、**Nonlinear Mixed Effects Models** ダイアログが開きます。

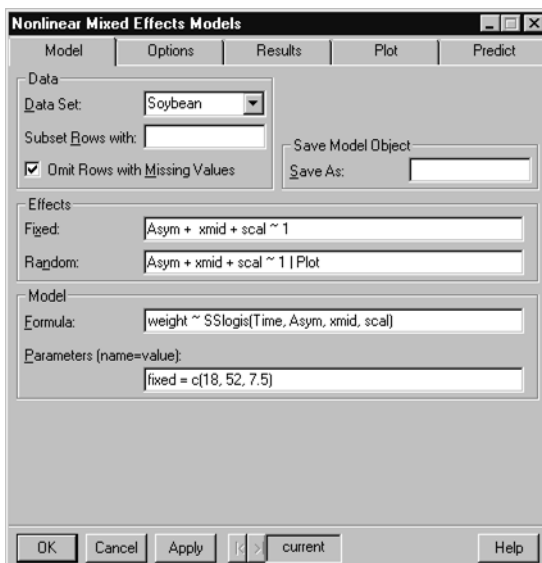


図 8.53 : Nonlinear Mixed Effects Models ダイアログ

例

データ Soybean は、2つの遺伝子型の大豆の成長パターンを比較する実験から得たものです。変量には、各地区の固有の識別子を示す因子 (Plot)、その地区の大豆の種類を示す因子 (Variety)、大豆を植えた年 (Year)、各標本を取得した時期 (time)、一株当たりの葉の平均重量 (weight) があります。パラメータ Asym、xmid および scal を持つロジスティックモデルにより、Time の関数として weight をモデル化したいと思います。これらのパラメータには、固定効果とランダム効果の両方があります。グループ化変量は Plot です。

1. P.383 の方法で **Nonlinear Mixed Effects Models** ダイアログを開きます。
2. **Data Set** フィールドに Soybean と入力します。
3. 次の**モデル式**を Formula に入力します。

```
weight ~ SSlogis(Time, Asym, xmid, scal)
```

このモデル式は、weight を変量 Time、Asym、xmid および scal を持つ関数 SSlogis によって予測したいことを指定しています。関数 SSlogis は非線形モデルを指定し、初期推定値をソルバーに提供するために使用される「セルフスターティング」関数です。

4. **Parameters (name=value)** フィールドに最初の固定効果パラメータの推定値を指定します

```
fixed=c(18, 52, 7.5)
```

5. **Effects** の下の **Fixed** フィールドに次のモデル式を指定して、**Asym**、**xmid** および **scal** が固定効果変量であることを指定します。

```
Asym + xmid + scal ~ 1
```

6. **Effects** の下の **Random** フィールドに次のモデル式を指定して、**Asym**、**xmid** および **scal** がランダム効果変量で、**Plot** がグループ化変量であることを指定します。

```
Asym + xmid + scal ~ 1 | Plot
```

7. **OK** をクリックします。

あてはめたモデルのサマリーがレポート・ウィンドウに表示されます。

一般化最小二乗モデル

一般化最小二乗モデルは、残差がより一般化された共分散構造を持つ回帰モデルまたは分散分析モデルです。サポートされる共分散構造には、相関を持つ残差と不均一分散の残差があります。

線形

Generalized Least Squares ダイアログは、一般化最小二乗法を使って線形モデルをあてはめます。誤差は、相関関係があることも不均一分散をもつこともあります。

一般化最小二乗回帰を実行する

メインメニューから、統計 ▶ 一般化最小二乗モデル ▶ 線形を選択します。図 8.54 に示すように、**Generalized Least Squares** ダイアログが開きます。

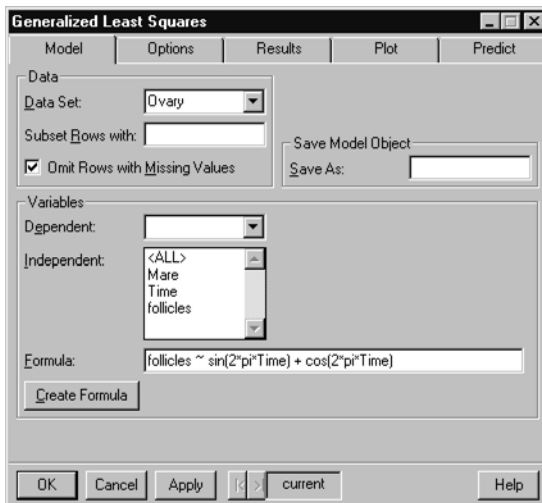


図 8.54 : Generalized Least Squares ダイアログ

例

データセット Ovary は、発情周期において異なる時期に別々の雌馬から検出された卵胞の数を示す 308 行と 3 列から成ります。生物学的モデルは、卵胞の数を $2 \cdot \pi \cdot \text{Time}$ の正弦と余弦の線形結合としてモデル化すると良いかもしれないことを示しています。ばらつきが Time と共に増えることを仮定し、そのため通常の線形回帰ではなく、べき乗分散構造 (Power variance structure) による一般化最小二乗を使用します。べき乗分散構造では、分散は、あてはめ値の絶対のべき乗で増大します。

1. P.386 の方法で **Generalized Least Squares** ダイアログを開きます。
2. **Data Set** フィールドに Ovary と入力します。
3. Formula に次のモデル式を入力します。

$$\text{follicles} \sim \sin(2\pi \cdot \text{Time}) + \cos(2\pi \cdot \text{Time})$$
4. ダイアログの **Options** ページで、**Variance Structure** グループの **Type** として **Power** を選択します。
5. **OK** をクリックします。

あてはめたモデルのサマリーが、レポート・ウィンドウに表示されます。

非線形

Generalized Nonlinear Least Squares ダイアログは、一般化最小二乗法を使用して非線形モデルをあてはめます。誤差は、相関関係があることも不均一分散をもつこともあります。

一般化非線形最小二乗回帰を実行する

メインメニューから、統計 ▶ 一般化最小二乗モデル ▶ 非線形を選択します。図 8.55 に示すように、**Generalized Nonlinear Least Squares** ダイアログが開きます。



図 8.55 : Generalized Nonlinear Least Squares ダイアログ

例

データ Soybean は、2つの遺伝子型の大豆の成長パターンを比較する実験から得たものです。変量には、各地区の固有の識別子を示す因子 (Plot)、その地区の大豆の種類を示す因子 (Variety)、大豆を植えた年 (Year)、各標本を取得した時期 (time)、一株当たりの葉の平均重量 (weight) があります。パラメータ Asym、xmid、および scal を持つロジスティックモデルにより Time の関数で weight をモデル化したいと思います。ばらつきが Time と共に増大することを仮定し、そのため通常の非線形回帰ではなく、べき乗分散構造による一般化最小二乗を使用します。べき乗分散構造では、分散が、あてはめ値の絶対のべき乗で増大します。

1. P.387の方法で **Generalized Nonlinear Least Squares** ダイアログを開きます。
2. **Data Set** フィールドに Soybean を入力します。
3. Formula に次のモデル式を入力します。

```
weight ~ SSlogis(Time, Asym, xmid, scal)
```

関数 SSlogis は非線形モデルを指定し、初期推定値をソルバーに提供するために使用される「セルフスターティング」関数です。

4. ダイアログの **Options** ページで **Variance Structure** グループの **Type** として **Power** を選択します。
5. **OK** をクリックします。

あてはめたモデルのサマリーが、レポート・ウィンドウに表示されます。

生存時間解析モデル

生存時間解析は、打ち切りがあるデータに使用されます。

ノンパラ メトリック

「ノンパラメトリック生存時間曲線」は、時間の経過による生存確率のノンパラメトリック推定値です。これは医学的試行のように、反応変数が通常打ち切りによって失われる時間を含む、故障までの時間のような場合に用いられます。最も一般的に使用されるノンパラメトリック生存時間曲線は、 Kaplan-Meier 推定です。**Nonparametric Survival** ダイアログは、様々なノンパラメトリック生存時間曲線をあてはめ、グループ化変量の含意を可能にします。

ノンパラメトリック生存時間曲線をあてはめる

メインメニューから、統計 ▶ 生存時間解析 ▶ ノンパラメトリックを選択します。図 8.56 に示すように、**Nonparametric Survival** ダイアログが開きます。

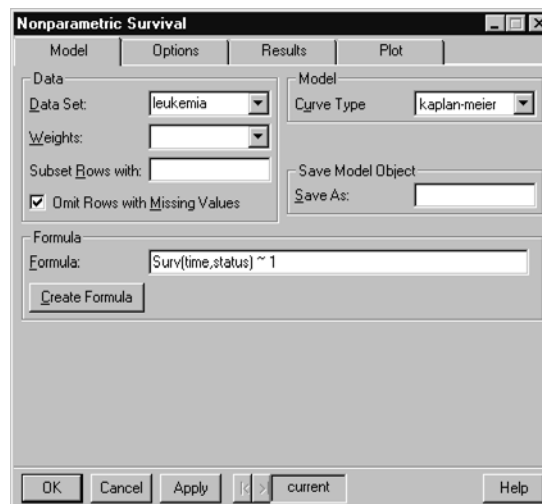


図 8.56 : Nonparametric Survival ダイアログ

例

データセット leukemia は、急性骨髄性白血病の維持化学療法の効果を評価する試行から得たデータです。すべての組のデータにカプラン=マイヤー生存時間曲線をあてはめます。

1. P.389の方法で **Nonparametric Survival** ダイアログを開きます。
2. **Data Set** フィールドに leukemia と入力します。
3. **モデル式** `Surv(time,status)~1` を入力するか、**Create Formula** ボタンをクリックしてモデル式を作成します。Surv 関数は、生存時間オブジェクトを作成します。これは、生存時間モデル式の反応変数に適しています。

モデル式ビルダーを使用するときは、**Create Formula** ボタンをクリックしてください。現われたダイアログで、time 変数をハイライトし、**Time 1** ボタンをクリックします。status 変数をハイライトし、**Censor Codes** ボタンをクリックし、次に、**Add Response** をクリックします。これで、次のモデル式が生成されます。

```
Surv(time, status, type="right") ~ 1
```

デフォルトでは、打ち切りのある引数 type は、与えられた時間変数が1つだけのときに"right"に設定されます。したがって、このモデル式は、`Surv(time,status)~1` と等価です。

4. **OK** をクリックします。

あてはめたモデルのサマリーが**レポート・ウィンドウ**に表示され、**グラフシート**に信頼区間のある生存時間曲線のプロットが現われます。

コックス比例 ハザード モデル

「コックス比例ハザードモデル」は、生存時間データに最もよく使用される回帰モデルです。このモデルを使用すると、共変量が存在する状態でノンパラメトリック生存時間曲線（カプラン=マイヤー曲線など）を推定することができます。通常は、生存時間に対する共変量の影響を知りたいときに使用します。

コックス比例ハザードモデルをあてはめる

メインメニューから、**統計 ▶ 生存時間解析 ▶ コックス比例ハザードモデル**を選択します。図 8.57 に示すように、**Cox Proportional Hazards** ダイアログが開きます。

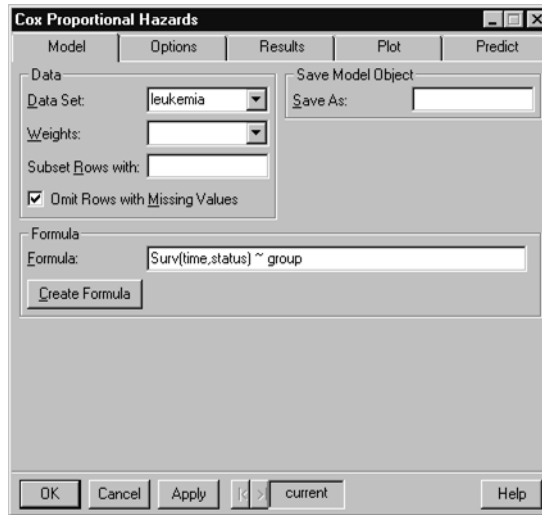


図 8.57 : Cox Proportional Hazards ダイアログ

例

group を共変量として、leukemia データセットにコックス比例ハザードモデルをあてはめます。

1. P.390 の方法で **Cox Proportional Hazards** ダイアログを開きます。
2. **Data Set** フィールドに leukemia と入力します。
3. **モデル式** `Surv(time,status)~group` を入力するか、**Create Formula** ボタンをクリックしてモデル式を作成します。Surv 関数は、生存時間オブジェクトを作成します。これは、生存時間モデル式の反応変量に適しています。

モデル式ビルダーを使用するときは、**Create Formula** ボタンをクリックしてください。現われたダイアログで、time 変量をハイライトし、**Time 1** ボタンをクリックします。status 変量をハイライトし、**Censor Codes** ボタンをクリックし、次に、**Add Response** をクリックします。最後に、group 変量をハイライトし、**Main Effect** をクリックします。これで、次のモデル式が生成されます。

```
Surv(time, status, type="right") ~ group
```

デフォルトでは、打ち切りのある引数 type は、与えられた時間変量が 1 つだけのときに "right" に設定されます。したがって、このモデル式は、`Surv(time,status)~group` と等価です。

4. **Plot** ページの **Survival Curves** チェックボックスを選択します。
5. **OK** をクリックします。

あてはめたモデルのサマリーが **レポート・ウィンドウ** に表示されます。信頼区間のある生存時間曲線のプロットが **グラフシート** に現われます。

パラメトリック

打ち切りのあるデータのパラメトリック回帰モデルは、製造業から環境汚染物質の研究までの様々な分野で使用されます。故障寿命または生存時間データのモデル化によく使用されるため、「パラメトリック生存時間モデル」と呼ばれることもしばしばあります。これらの分野では、設計した製品が故障する原因を発見するために設計全体を通して使用されます。故障寿命が早まるように製品が通常よりもかなり厳しい条件で試験されるときは、「加速故障寿命モデル」または「加速試験モデル」と呼ばれます。

Parametric Survival および **Life Testing** ダイアログは、同じタイプのモデルをあてはめます。この 2 つのダイアログは、使用できるオプションが異なります。**Life Testing** ダイアログでは、しきい値 (threshold) の推定、切り捨て分布の指定、オフセットの指定を行うことができます。さらに様々な診断プロットと予測値を得る機能があります。この機能は、**Parametric Survival** ダイアログでは使用できません。これと対照的に、**Parametric Survival** ダイアログは、フレイルティ (frailty) およびペナルティ付き尤度モデルをサポートしていますが、この機能は、**Life Testing** ダイアログでは使用できません。

パラメトリック生存時間モデルをあてはめる

メインメニューから、**統計 ▶ 生存時間解析 ▶ パラメトリック** を選択します。図 8.58 に示すように、**Parametric Survival** ダイアログが開きます。

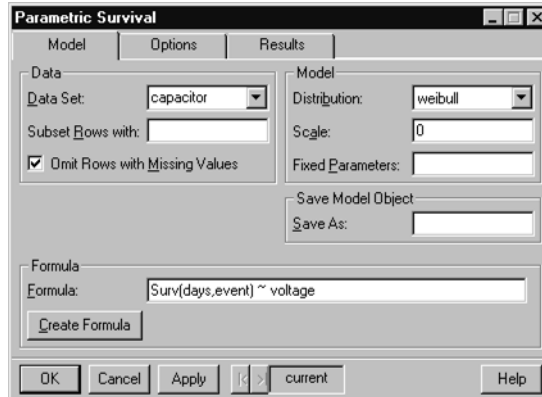


図 8.58 : Parametric Survival ダイアログ

例

データセット capacitor は、コンデンサの加速寿命試験のシミュレーションによる実測値を含みます。故障までの時間 (days)、故障または打ち切りの標識 (event)、試験を行った電圧 (voltage) を含みます。パラメトリック生存時間モデルを使用して、故障確率に電圧がどのように影響しているかを調べます。

1. P.392 の方法で **Parametric Survival** ダイアログを開きます。
2. **Data Set** フィールドに capacitor と入力します。
3. **モデル式** `Surv(days,event)~voltage` を入力するか、**Create Formula** ボタンをクリックしてモデル式を作成します。Surv 関数は、生存時間オブジェクトを作成します。これは、生存時間モデル式の反応変数に適しています。

モデル式ビルダーを使用するときは、**Create Formula** ボタンをクリックしてください。現われたダイアログで、days 変数をハイライトし、**Time 1** ボタンをクリックします。event 変数をハイライトし、**Censor Codes** ボタンをクリックし、次に、**Add Response** をクリックします。最後に、voltage 変数をハイライトし、**Main Effect** をクリックします。これで、次のモデル式が生成されます。

```
Surv(days, event, type="right") ~ voltage
```

デフォルトでは、打ち切りのある引数 `type` は、与えられた時間変量が1つだけのときに "right" に設定されます。したがって、このモデル式は、`Surv(days,event)~voltage` と等価です。

4. **OK** をクリックします。

あてはめたモデルのサマリーが、レポート・ウィンドウに表示されます。

打ち切りのあるデータの解析

Life Testing ダイアログは、打ち切りのあるデータのパラメトリック回帰モデルをあてはめます。このモデルは、製造業から環境汚染物質の研究まで様々な分野で使用されます。故障時間または生存時間のデータのモデル化によく使用されるため、「パラメトリック生存時間モデル」と呼ばれることがしばしばあります。これらの分野では、設計した製品が故障する原因を発見するために設計全体を通して使用されます。故障寿命を早めるよう製品が通常よりもかなり厳しい条件で試験されるときには、「加速故障寿命モデル」または「加速試験モデル」と呼ばれます。

Parametric Survival および **Life Testing** ダイアログは、同じタイプのモデルをあてはめます。2つのダイアログは使用できるオプションが違います。**Life Testing** ダイアログは、しきい値 (threshold) の推定、切り捨て分布の指定、オフセットの指定を行うことができます。さらに、様々な診断プロットと予測値を求める機能があります。この機能は、**Parametric Survival** ダイアログでは使用できません。これと対照的に、**Parametric Survival** ダイアログは、フレイルティ (frailty) およびペナルティ付き尤度モデルをサポートしていますが、この機能は、**Life Testing** ダイアログでは使用できません。

打ち切りのあるデータの解析を実行する

メインメニューから、統計 ▶ 生存時間解析 ▶ 打ち切りのあるデータの解析を選択します。図 8.59 に示すように、**Life Testing** ダイアログが開きます。

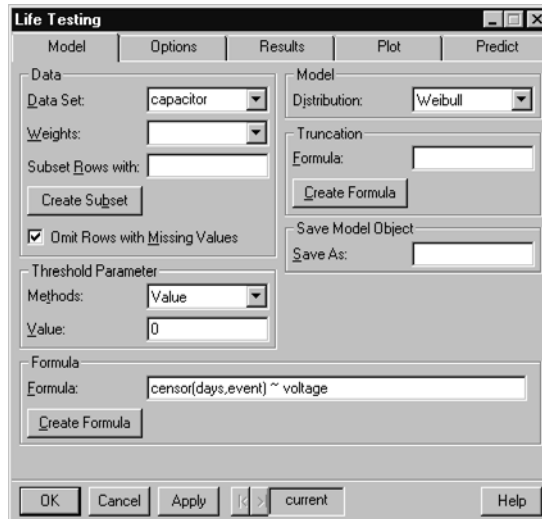


図 8.59 : Life Testing ダイアログ

例

Life Testing ダイアログを使って、capacitor データセットにおいて故障の確率に電圧がどのように影響しているかを調べます。

1. P.394 の方法で **Life Testing** ダイアログを開きます。
2. **Data Set** フィールドに capacitor と入力します。
3. **モデル式** `censor(days,event)~voltage` を入力するか、**Create Formula** ボタンをクリックしてモデル式を作成します。censor 関数は、生存時間オブジェクトを作成します。これは、生存時間モデル式に適した反応変量です。Surv 関数と同様ですが、これは、打ち切りコードを指定する多くのオプションを提供します。

モデル式ビルダーを使用するときは、**Create Formula** ボタンをクリックしてください。現われたダイアログで、days 変量をハイライトし、**Time 1** ボタンをクリックします。event 変量をハイライトし、**Censor Codes** ボタンをクリックし、次に、**Add Response** をクリックします。最後に、voltage 変量をハイライトし、**Main Effect** をクリックします。これで、次のモデル式が生成されます。

```
censor(days, event, type="right") ~ voltage
```

デフォルトでは、打ち切りのある引数 `type` は、与えられた時間変数が 1 つだけのときに "right" に設定されます。したがって、このモデル式は、`censor(days,event)~voltage` と等価です。

4. **OK** をクリックします。

あてはめたモデルのサマリーが、レポート・ウィンドウに表示されます。

樹形モデル

樹形モデルは、回帰問題に対しては線形回帰モデルや加法モデルの代わりになり、分類問題に対しては線形ロジスティックモデルや加法ロジスティックモデルの代わりに使用することができます。樹形モデルは、データを連続的に分割して等質なサブセットを形成することによってあてはめられます。結果は、予測または分類に有効な決定ルールを表す階層樹形です。

樹形モデル

Tree Models ダイアログは、樹形モデルのあてはめに使用されます。

樹形モデルをあてはめる

メインメニューから、**統計 ▶ 樹形 ▶ 樹形モデル**を選択します。図 8.60 に示すように、**Tree Models** ダイアログが開きます。

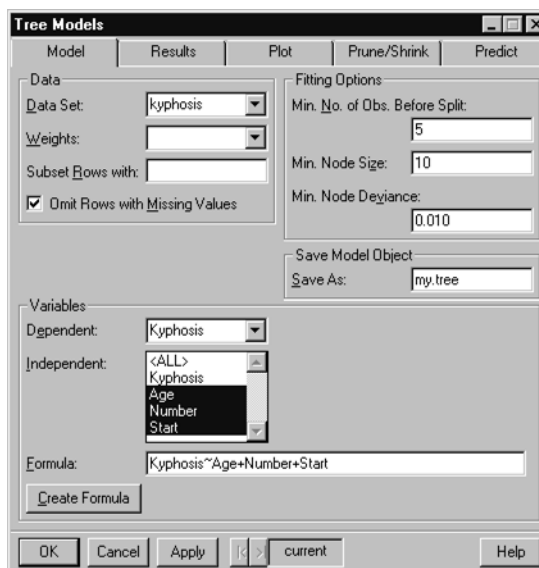


図 8.60 : Tree Models ダイアログ

例

データセット `kyphosis` は、脊椎矯正手術を受けた 81 人の子供のデータを表す 81 行から成ります。結果である `Kyphosis` は 2 値変数であり、他の 3 つの列 (`Age`、`Number` および `Start`) は数値です。脊柱後弯症は、脊椎手術を受けた一部の子供に見られる術後の変形です。子供の年齢、手術した椎

骨の数、または起点椎骨が、子供が変形を持つかどうかに影響するか調べたいと思います。

分類樹をデータにあてはめます。樹形構造は、Age、Number および Start の値に基づいて、個体を脊柱後弯症になる確率で分類するために使用されます。得られた分類樹は、個体をその変量に基づいてグループ分けします。

1. P.397 の方法で **Tree Models** ダイアログを開きます。
2. **Data Set** フィールドに `kyphosis` と入力します。
3. **Formula** フィールドで `Kyphosis~Age+Number+Start` を指定します。
4. **Save As** フィールドに `my.tree` と入力します。樹形モデルオブジェクトが、この名前で保存されます。後の例で、**Tree Tools** を使ってこのモデルを調べます。
5. **OK** をクリックします。

モデルのサマリーが、レポート・ウィンドウに表示され、樹形プロットが、グラフシートに表示されます。

樹形モデルの チェック

S-PLUS は、回帰樹を対話形式で調べるための豊富なツールを提供しています。**Tree Tools** ダイアログを使用するときは、最初に **Tree Models** ダイアログを使って樹形モデルを作成してください。このダイアログの **Save As** フィールドに名前を指定して、樹形モデルを保存してください。

Tree Tools は、指定した樹形モデルのプロットを作成することから始めます。**Browse**、**Burl**、**Histogram**、**Identify**、**Snip** ツールを使用すると、プロットの上の分割 (`split`) または節 (`node`) を選択し、その部分に関する情報を得ることができます。選択するときは左マウスボタンをクリックし、選択モードを終了するときは右マウスボタンをクリックします。これらのツールを使用する場合は、選択を行うときに必要なグラフ・ウィンドウとレポート・ウィンドウが見えているように、**OK** または **Apply** をクリックする前にウィンドウを並べ替えることが必要な場合があります。

ツールは、以下のように動作します。

- **Browse** : 樹形プロットの節を選択します。節のサマリー情報がレポート・ウィンドウに表示されます。選択モードを終了するときは右クリックしてください。

- **Burl** : 樹形プロット上の分割を選択します。樹形プロットの下に、すべての条件によって分割したときの尤離度の変化を表すプロットが表示されます。実際の分割は、最も大きく尤離度が変化したもので、グラフ上最も長い線を示しています。これらのプロットは、他の分割条件があてはめにおいて実際の分割条件による改善と同じ程度の改善をするかどうかを調べるのに役立ちます。選択モードを終了するときは右クリックしてください。候補の分割の情報を含むリストを保存するときは、**Save As** フィールドに名前を指定してください。
- **Histogram : Hist Variables** フィールドにヒストグラムを描く変量を指定してください。樹形上の分割を選択してください。樹形プロットの下に、指定した変量のヒストグラムを表すプロットが現われます。分割による2つの節（左右）ごとに、選択された変量のヒストグラムを表示します。選択モードを終了するときは右クリックしてください。ヒストグラムに対応する変量値のリストを保存するときは、**Save As** フィールドで名前を指定してください。
- **Identify** : 樹形プロット上の節を選択します。その節の観測値の行名または行番号が、**レポート・ウィンドウ**に表示されます。選択モードを終了するときは右クリックしてください。各節の観測値のリストを保存するときは、**Save As** フィールドに名前を指定してください。
- **Rug : Rug/Tile Variable** フィールドにプロットする変量を指定してください。樹形プロットの下に、各葉に分類されるデータの、指定された変量の平均値を示す高密度関数プロットがプロットされます。平均値のベクトルを保存するときは、**Save As** フィールドに名前を指定してください。このツールは対話型ではありません。
- **Snip** : いくつかの分割を削除した新しい樹形を作成するときにこのツールを使用します。樹形プロット上の節を選択すると、合計の樹形尤離度と、節を根にした部分樹が削除される場合の合計の樹形尤離度を表示します。同じ節を2度クリックすると、その部分樹が切られて見えなくなります。このプロセスは何回でも繰り返すことができます。選択モードを終了するときは右クリックしてください。切断した残りの樹を保存するときは、**Save As** フィールドで名前を指定してください。
- **Tile** : プロットする変量を **Rug/Tile Variable** フィールドに指定してください。変量の縦棒グラフが樹形プロットの下にプロットされます。因子変量には水準ごとに1つの棒グラフがあり、数値変量は、4分割されます。各葉の観測値の度数を保存するときは、**Save As** フィールドに名前を指定してください。このツールは、対話式ではありません。

樹形ツールを使用する

メインメニューから、統計 ▶ 樹形 ▶ 樹形モデルのチェックを選択します。図 8.61 に示すように、**Tree Tools** ダイアログが開きます。

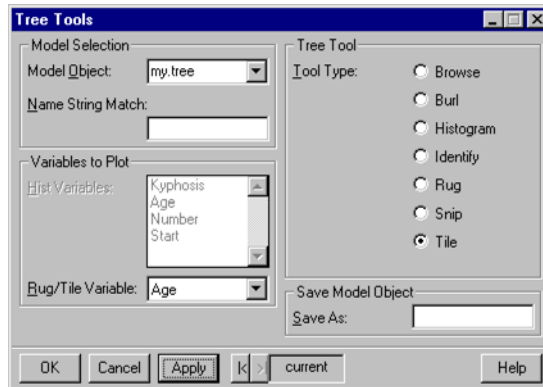


図 8.61 : Tree Tools ダイアログ

例

397 ページの「樹形モデル」において、データ `kyphosis` に分類樹をあてはめます。樹形タイルプロットを使って、各グループ内の `Age` のヒストグラムを見ることができます。

1. まだ分類樹をあてはめて `my.tree` という名前のオブジェクトで保存していない場合は、それを行ってください。この手順は、397 ページで説明しています。
2. P.400 の方法で **Tree Tools** ダイアログを開きます。
3. **Model Object** として `my.tree` を選択します。
4. **Tool Type** として **Tile** を選択します。
5. **Rug/Tile Variable** として `Age` を選択します。
6. **OK** をクリックします。

樹形タイルプロットがグラフシートに表示されます。グラフの上の部分に樹形のプロットが表示されます。下の部分には、樹形における各末端の節の `Age` のヒストグラムが表示されます。

モデルの比較

回帰と分散分析において、データアナリストは様々なモデルの候補を対象とすることがしばしばあります。データアナリストは、通常それらのモデルから予測変量と反応変量の関係を最も良く示すと思われるものを選択します。

モデルの選択には、一般的に複雑さと適合度のトレードオフが伴います。複雑なモデル（変数の数や変量の相互作用が多いモデル）ほど、観測データを単純なモデルよりうまくあてはめることができます。たとえば、観測値と同じ数のパラメータを含むモデルは、データを完全にあてはめることができます。しかしモデルが複雑になるほど、反応変量と予測変量間のより一般的な関係ではなく、得られた標本のランダムな変動の影響を受けます。そのため、新しい値を予測したりモデル構造に関する結論を引き出ししたりするのに、モデルが単純なモデルよりも有効でなくなることがあります。

回帰における一般的な方法は、適合度をあまり低下させないように、より単純なモデルを選択することです。線形回帰と分散分析では、2つのモデルを比較するために F 検定が使用されることがあります。ロジスティック回帰と対数線形回帰には、尤離度を比較するカイ二乗検定が適しています。

Compare Models ダイアログを使用すると、複数のモデルの適合度を比較することができます。一般に、モデルは単純な方が、複雑な方の特別な場合であるように「ネストされて」いることがほとんどです（単純な方に含まれる変数はすべて複雑なモデルに含まれる）。**Compare Models** ダイアログを使用する前に、まず対象とするモデルをオブジェクトとして保存してください。

モデルを比較する

メインメニューから、統計 ▶ **モデルの比較**を選択します。図 8.62 に示すように、**Compare Models (Likelihood Ratio Test)** ダイアログが開きます。

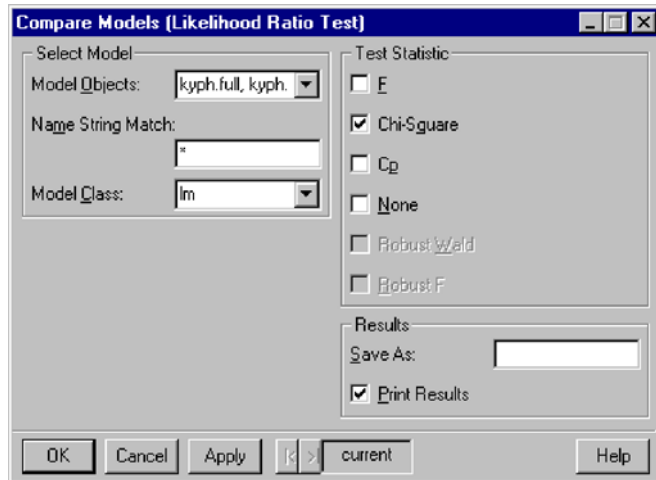


図 8.62 : Compare Models (Likelihood Ratio Test) ダイアログ

例

371 ページのロジスティック回帰分析で、Start は Kyphosis に大きな影響を及ぼすけれども、Age と Number は影響を及ぼさないことを示しました。カイ二乗検定法を使用して、Start だけによるモデルで十分かどうかを判定することができます。

1. P.371 の方法で **Logistic Regression** ダイアログを開きます。
2. **Data Set** フィールドに `kyphosis` と入力します。
3. **Formula** フィールドで `Kyphosis~Age+Number+Start` を指定します。**Save As** フィールドに `kyph.full` を入力し、**Apply** をクリックします。このモデルについての情報が、`kyph.full` と名前の付けられたオブジェクトとして保存されます。
4. **Formula** フィールドを `Kyphosis~Start` に変更します。**Save As** の名前を `kyph.sub` に変更し、**OK** をクリックします。このモデルについての情報は、`kyph.sub` と名前の付けられたオブジェクトとして保存されます。
5. P.401 の方法で **Compare Models (Likelihood Ratio Test)** ダイアログを開きます。
6. CTRL-クリックして、**Model Objects** リスト内の `kyph.full` と `kyph.sub` を選択します。
7. **Test Statistic** として **Chi-Square** を選択します。

8. **OK** をクリックします。

尤離度表の分析が、レポート・ウィンドウに表示されます。この表は、各モデルの自由度と残差尤離度を示しています。単純なモデルの方が適しているという帰無仮説では、残差尤離度の差はカイ二乗統計値として分布します。列 Pr (Chi) は、より単純なモデルが適しているという帰無仮説に対する p 値を提供します。この値が一般に 0.05 などの特定の値よりも小さい場合は、複雑なモデルは項を追加することによって尤離度を大きく変化させます。すなわちモデルの複雑さが、適合度の改善によって調整された結果です。

この例で p 値 0.035 は、Age または Number が結果を予測するのに役立つ有益な情報を追加することを示します。

Analysis of Deviance Table

Response: Kyphosis

	Terms	Resid. Df	Resid. Dev	Test
1	Age + Number + Start	77	61.37993	
2	Start	79	68.07218	-Age-Number
	Df Deviance	Pr(Chi)		
1				
2	-2	-6.692253	0.03522052	

クラスター分析

クラスター分析は、同じクラスターに属するオブジェクトは互いに類似していても、違うクラスターのオブジェクトは異なるというように、データのグループ（クラスター）を探索するものです。

非類似度の算出

クラスターリングするデータは、観測値の行から成るデータセット、あるいは観測値間の非類似性の尺度で構成された非類似度オブジェクトのいずれかです。 k -means 型、pam (partitioning around medoids)、系統分類のクラスターリングはすべて、データセットを操作するアルゴリズムです。pam アルゴリズムによるパーティショニング、ファジィ・パーティショニング、および各階層法は、データセットと非類似度オブジェクトのどちらにも使用できます。

クラスターリングルーチンは非数値変量を受け入れません。データが因子などの非数値変量を含む場合は、その変量を数値変量に変換するか非類似度を使用しなければなりません。

2つのオブジェクト間の非類似度を計算する方法は、元の変量のタイプによります。デフォルトでは、数値列は間隔尺度変量として、因子は名義変量として、順序因子は順序変量として扱われます。他の変量のタイプは、**Special Variable Types** グループのフィールドなどで指定してください。

非類似度を計算する

メインメニューから、統計 ▶ クラスター分析 ▶ 非類似度を選択します。図 8.63 に示すように、**Compute Dissimilarities** ダイアログが開きます。

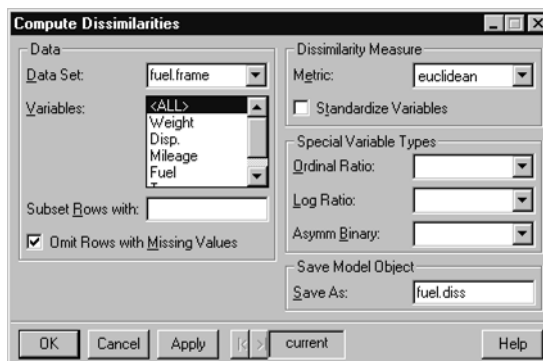


図 8.63 : Compute Dissimilarities ダイアログ

例

データセット `fuel.frame` は、1990 年 4 月発行の『Consumer Reports』からの引用です。このデータは、60 の観測値（行）と 5 つの変量（列）を含んでいます。60 台の車のそれぞれについて、重量、エンジン排気量、燃費、種類、および燃料消費率の観測値を得ました。`fuel.frame` データの非類似度を、次のように計算します。

1. P.404 の方法で **Compute Dissimilarities** ダイアログを開きます。
2. **Data Set** フィールドに `fuel.frame` と入力します。
3. **Save As** フィールドに `fuel.diss` と入力します。
4. **OK** をクリックします。

非類似度が計算され、`fuel.diss` として保存されます。これは、後のクラスタリングダイアログの例で使用します。

κ-means 型 クラスタ リング

最もよく知られている分割法の 1 つは、「κ-means 型」です。κ-means 型アルゴリズムでは、観測値が k のグループの 1 つに属するように分類されます。グループのメンバシップは各グループの重心を計算し（平均の多次元バージョン）、それぞれの観測値を最も近い重心を持つグループに割り当てることによって決定されます。

κ-means 型クラスタリングを実行する

メインメニューから、統計 ▶ クラスタ分析 ▶ κ-means 型を選択します。図 8.64 に示すように、**K-Means Clustering** ダイアログが開きます。

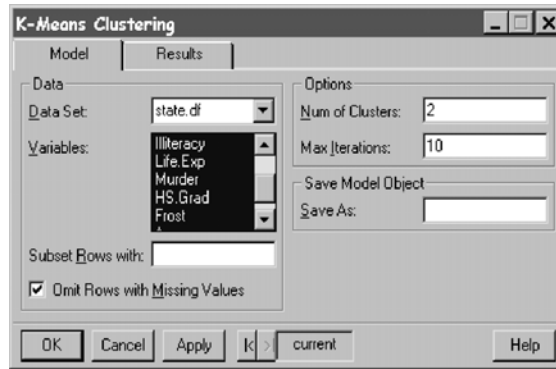


図 8.64 : K-Means Clustering ダイアログ

例

state.df データセットの情報をクラスタリングします。これらのデータは、人口、所得、非識字率、平均寿命および教育を含む、アメリカ 50 州の様々な特性を示しています。

1. **ファイル ▶ ライブラリのロード**を選択して、example5 ライブラリをロードします。**Library Name** リスト内の example5 をハイライトし、**OK** をクリックします。このライブラリは、state.df を含む、主要な S-PLUS データベースにはないいくつかのデータセットの例を含んでいます。
2. P.406 の方法で **K-Means Clustering** ダイアログを開きます。
3. **Data Set** フィールドに state.df を入力します。
4. CTRL-クリックして、**Variables** として Population から Area までを選択します。
5. **OK** をクリックします。

クラスタリングのサマリーが、**レポート・ウィンドウ**に表示されます。

pam アルゴリズムによるパーティショニング

「pam アルゴリズムによるパーティショニング」のアルゴリズムは、κ-means 型と似ていますが重心ではなく medoids を使用します。pam アルゴリズムによるパーティショニングには、次の利点があります。まず、非類似度行列を使用することができます。ユークリッド距離の二乗の和ではなく、非類似度の合計を最小にするのでより頑健です。また、新しいグラフィック表示（シルエットプロットとクラスプロット）を提供します。

pam アルゴリズムによるパーティショニングを実行する

メインメニューから、統計 ▶ クラスター分析 ▶ pam アルゴリズムによるパーティショニングを選択します。図 8.65 に示すように、**Partitioning Around Medoids** ダイアログが開きます。

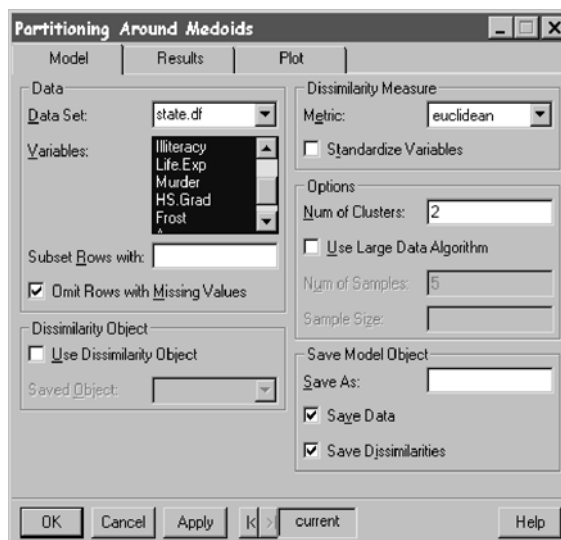


図 8.65 : Partitioning Around Medoids ダイアログ

例 1

405 ページの「 k -means 型クラスター分析」において、 k -means 型アルゴリズムを使用して `state.df` データセットの情報をクラスター分析しました。この例では、pam アルゴリズムによるパーティショニングのアルゴリズムを使用します。

1. `example5` ライブラリをまだロードしていない場合は、**ファイル ▶ ライブラリのロード**を選択して、`example5` ライブラリをロードしてください。このライブラリは、`state.df` を含む主要な S-PLUS データベースにはないいくつかのデータセットの例を含んでいます。
2. P.407 の方法で **Partitioning Around Medoids** ダイアログを開きます。
3. **Data Set** フィールドに `state.df` を入力します。
4. CTRL-クリックして、**Variables** として Population から Area までを選択します。
5. **OK** をクリックします。

クラスター分析のサマリーが、**レポート・ウィンドウ**に表示されます。

例 2

404 ページの「非類似度の算出」では、`fuel.frame` データセットの非類似度を計算しました。この例では、**pam** アルゴリズムによるパーティショニングのアルゴリズムを使って `fuel.frame` の非類似度をクラスタリングします。

1. オブジェクト `fuel.diss` をまだ作成していない場合は、405 ページの指示にしたがって作成してください。
2. P.407 の方法で **Partitioning Around Medoids** ダイアログを開きます。
3. **Use Dissimilarity Object** チェックボックスを選択します。
4. **Saved Object** として `fuel.diss` を選択します。
5. **OK** をクリックします。

クラスタリングのサマリーが、レポート・ウィンドウに表示されます。

ファジィ・ パーティ ショニング

ほとんどのクラスタリングのアルゴリズムは、明瞭なクラスタリング方法です。つまり、データの各オブジェクトは 1 つのクラスターだけに割り当てられています。たとえば 2 つのクラスターの間にあるオブジェクトは、それらの一方に割り当てられなければなりません。「ファジィなクラスタリング」では、それぞれの観測値は複数のクラスターの部分的メンバシップが与えられます。

ファジィ・パーティショニングを実行する

メインメニューから、**統計 ▶ クラスター分析 ▶ ファジィ・パーティショニング** を選択します。図 8.66 に示すように、**Fuzzy Partitioning** ダイアログが開きます。

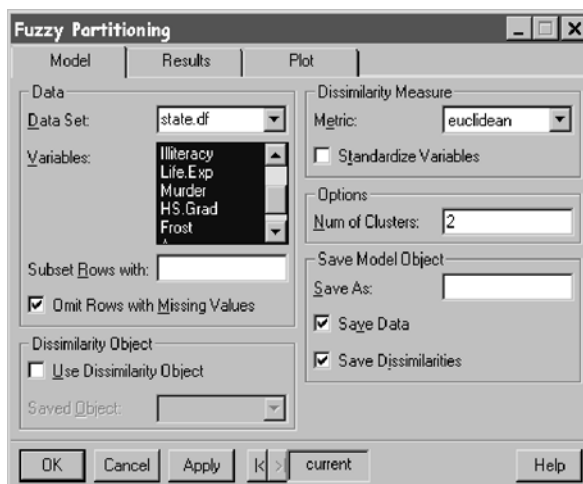


図 8.66 : Fuzzy Partitioning ダイアログ

例 1

405 ページの「 κ -means 型クラスタリング」では、 κ -means 型アルゴリズムを使って state.df データセットの情報をクラスタリングしました。この例では、ファジィ・パーティショニングを使用します。

1. example5 ライブラリをまだロードしていない場合は、**ファイル ▶ ライブラリのロード**を選択して、example5 ライブラリをロードしてください。このライブラリは、state.df を含む主要な S-PLUS データベースにはないいくつかのデータセットの例を含んでいます。
2. P.408 の方法で **Fuzzy Partitioning** ダイアログを開きます。
3. **Data Set** フィールドに state.df を入力します。
4. CTRL-クリックして、**Variables** として Population から Area までを選択します。
5. **OK** をクリックします。

クラスタリングのサマリーが、**レポート・ウィンドウ**に表示されます。

例 2

404 ページの「非類似度の算出」では、fuel.frame データセットの非類似度を計算しました。この例では、ファジィ・パーティショニングを使って fuel.frame 非類似度をクラスタリングします。

1. オブジェクト fuel.diss をまだ作成していない場合は、405 ページの指示にしたがって作成してください。

2. P.408 の方法で **Fuzzy Partitioning** ダイアログを開きます。
3. **Use Dissimilarity Object** チェックボックスを選択します。
4. **Saved Object** として fuel.diss を選択します。
5. **OK** をクリックします。

クラスタリングのサマリーが、レポート・ウィンドウに表示されます。

擬集型階層 クラスタ リング

階層アルゴリズムは、既存のグループを結合または分割することにより処理が行われ、グループを合併または分割する順序を示す階層構造を生成します。「擬集」法は最初に各観測値を別々のグループとして始め、すべての観測値が 1 つのグループになるまで結合を繰り返します。

擬集型階層クラスタリングを実行する

メインメニューから、統計 ▶ クラスタ分析 ▶ 擬集型階層を選択します。図 8.67 に示すように、**Agglomerative Hierarchical Clustering** ダイアログが開きます。

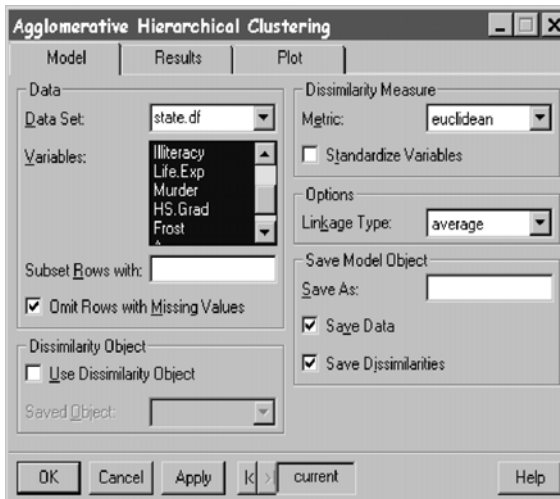


図 8.67 : Agglomerative Hierarchical Clustering ダイアログ

例 1

405 ページの「 κ -means 型クラスタリング」では、 κ -means 型アルゴリズムを使って `state.df` データセットの情報をクラスタリングしました。この例では、擬集型階層法を使用します。

1. `example5` ライブラリをまだロードしていない場合は、**ファイル ▶ ライブラリのロード**を選択して、`example5` ライブラリをロードしてください。このライブラリは、`state.df` を含む主要な S-PLUS データベースにはないいくつかのデータセットの例を含んでいます。
2. P.410 の方法で **Agglomerative Hierarchical Clustering** ダイアログを開きます。
3. **Data Set** フィールドに `state.df` を入力します。
4. CTRL-クリックして、**Variables** として Population から Area までを選択します。
5. **OK** をクリックします。

クラスタリングのサマリーが、**レポート・ウィンドウ**に表示されます。

例 2

404 ページの「非類似度の算出」では、`fuel.frame` データセットの非類似度を計算しました。この例では、擬集型階層アルゴリズムを使って `fuel.frame` の非類似度をクラスタリングします。

1. オブジェクト `fuel.diss` をまだ作成していない場合は、405 ページの指示にしたがって作成してください。
2. P.410 の方法で **Agglomerative Hierarchical Clustering** ダイアログを開きます。
3. **Use Dissimilarity Object** チェックボックスを選択します。
4. **Saved Object** として `fuel.diss` を選択します。
5. **OK** をクリックします。

クラスタリングのサマリーが、**レポート・ウィンドウ**に表示されます。

区分型階層 クラスタ リング

階層アルゴリズムは、既存のグループを結合または分割することによって処理が行われ、グループを合併または分割する順序を示す階層構造を生成します。「区分」法は最初にすべての観測値を 1 つのグループとして始め、それぞれの観測値が別々のグループになるまで分割を繰り返します。

区分型階層クラスタリングを実行する

メインメニューから、**統計 ▶ クラスタ分析 ▶ 区分型階層**を選択します。図 8.68 に示すように、**Divisive Hierarchical Clustering** ダイアログが開きます。

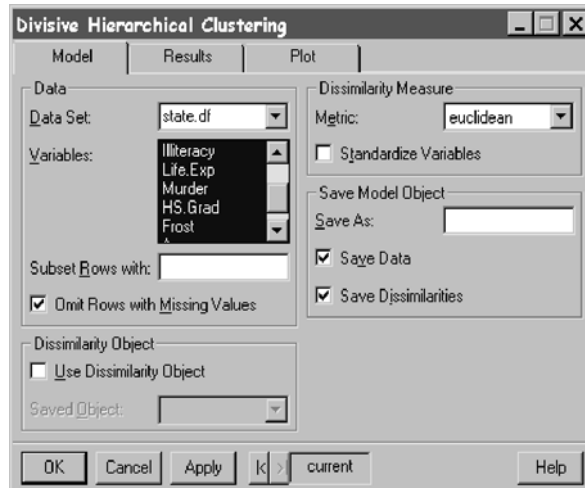


図 8.68 : Divisive Hierarchical Clustering ダイアログ

例 1

405 ページの「 κ -means 型クラスタリング」では、 κ -means 型アルゴリズムを使って state.df データセットの情報をクラスタリングしました。この例では、区分型階層法を使用します。

1. example5 ライブラリをまだロードしていない場合は、**ファイル ▶ ライブラリのロード**を選択して、example5 ライブラリをロードしてください。このライブラリは、state.df を含む主要な S-PLUS データベースにはないいくつかのデータセットの例を含んでいます。
2. P.412 の方法で **Divisive Hierarchical Clustering** ダイアログを開きます。
3. **Data Set** フィールドに state.df を入力します。
4. CTRL-クリックして、**Variables** として Population から Area までを選択します。
5. **OK** をクリックします。

クラスタリングのサマリーが、**レポート・ウィンドウ**に表示されます。

例 2

404 ページの「非類似度の算出」では、`fuel.frame` データセットの非類似度を計算しました。この例では、区分型階層アルゴリズムを使って `fuel.frame` の非類似度をクラスタリングします。

1. オブジェクト `fuel.diss` をまだ作成していない場合は、405 ページの指示にしたがって作成してください。
2. P.412 の方法で **Divisive Hierarchical Clustering** ダイアログを開きます。
3. **Use Dissimilarity Object** チェックボックスを選択します。
4. **Saved Object** として `fuel.diss` を選択します。
5. **OK** をクリックします。

クラスタリングのサマリーが、レポート・ウィンドウに表示されます。

系統分類 クラスタ リング

データ内のすべての変量が 2 水準因子であるときに観測値を分割するのに適している方法は、特定の 2 値変量の 2 つの値に基づいてデータを 2 つのグループに分けることです。「系統分類分析」は、各ステップで 2 値変量のうちの 1 つの値に基づいてグループが 2 つに分割されたクラスターの階層を生成します。

系統分類クラスタリングを実行する

メインメニューから、統計 ▶ クラスター分析 ▶ 系統分類 (2 値変量) を選択します。図 8.69 に示すように、**Monothetic Clustering** ダイアログが開きます。

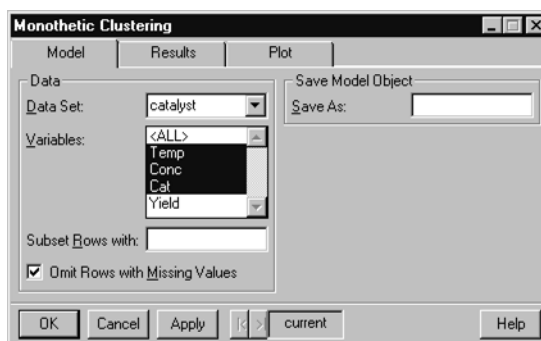


図 8.69 : Monothetic Clustering ダイアログ

例

catalyst データセットは、計画された実験から得たものです。その 8 行は、2 つの温度 (Temp)、2 つの濃度 (Conc)、および 2 つの触媒 (Cat) の可能なすべての組合せを表します。第 4 の列は、反応変量 Yield を表します。温度、濃度、および触媒が Yield にどのように影響するかを調べたいと思います。これらのデータにモデルをあてはめる前に、系統分類クラスタリングを使って、3 つの 2 水準因子予測変量にしたがい観測値をグループ化することができます。

1. **Monothetic Clustering** ダイアログを開きます。
2. **Data Set** フィールドに catalyst と入力します。
3. CTRL-クリックして **Variables** として Temp、Conc および Cat をハイライトします。
4. **OK** をクリックします。

系統分類クラスタリングのサマリーが、レポート・ウィンドウに表示されます。

多変量解析

多変量解析法は、いくつかの伝統的モデルに基づいて多変量データの構造をまとめます。

判別分析

Discriminant Analysis ダイアログを使用すると、線形または二次判別関数を 1 組の特徴的なデータにあてはめることができます。

判別分析を実行する

メインメニューから、**統計** ▶ **多変量解析** ▶ **判別分析** を選択します。図 8.70 に示すように、**Discriminant Analysis** ダイアログが開きます。

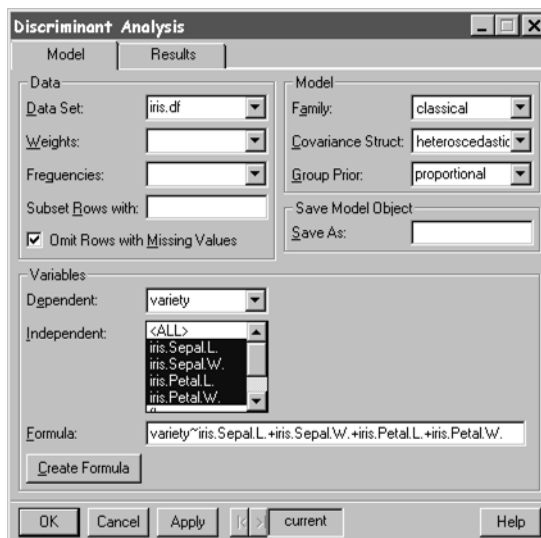


図 8.70 : Discriminant Analysis ダイアログ

例

フィッシャーの `iris` データに判別分析を実行します。このデータセットは、3 種類のアイリスについてそれぞれ 50 輪の花の 4 つの測定値を示す 3 次元配列です。測定値は、センチメートルで表したがく片の長さ、がく片の幅、花弁の長さ、および花弁の幅からなります。

アイリスの種類は、`Setosa`、`Versicolor`、および `Virginica` です。example5 ライブラリに含まれるデータフレーム `iris.df` は、`iris` データを 2 次元で表したものです。

1. **ファイル ▶ ライブラリのロード**を選択してexample5ライブラリをロードします。**Library Name** リストのexample5をハイライトし、**OK**をクリックします。このライブラリは、主要なS-PLUSデータベースにはないいくつかのデータセットの例を含みます。
2. P.415の**Discriminant Analysis** ダイアログを開きます。
3. **Data Set** フィールドにiris.dfを入力します。
4. **Dependent** 変量としてvarietyを選択します。
5. CTRL-クリックして、**Independent** 変量としてiris.Sepal.L.、iris.Sepal.W.、iris.Petal.L.、およびiris.Petal.W.を選択します。
6. **Covariance Struct** としてheteroscedasticを選択します。
7. **OK**をクリックします。

あてはめたモデルのサマリーが、レポート・ウィンドウに表示されます。

因子分析

多くの科学分野、特に心理学やその他の社会科学の分野では、知能や社会的地位などの直接測定することができない量を調べたいことがあります。その場合、対象としたい基本的変量を反映する他の量のみ測定できることがあります。「因子分析」では、それ自体を直接観測することができない基本的因子の項目と観測可能な変量間の相関を説明しようとする解析です。たとえば一連の試験の成績のような測定可能な量は、知能などの基本的因子の項目を説明することができます。

因子分析の実行

メインメニューから、**統計 ▶ 多変量解析 ▶ 因子分析**を選択します。図 8.71 に示すように、**Factor Analysis** ダイアログが開きます。

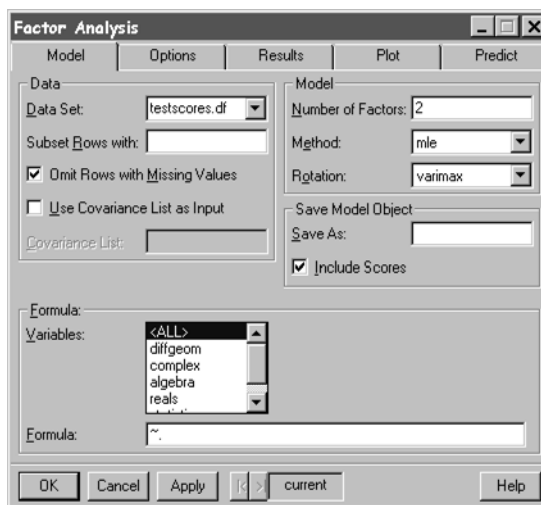


図 8.71 : Factor Analysis ダイアログ

例

example5 ライブラリにあるデータセット `testscores.df` には、学生 25 人それぞれの 5 科目の試験の点数が含まれています。因子分析を使用して点数の構造を調べます。

1. example5 ライブラリをまだロードしていない場合は、**ファイル ▶ ライブラリのロード**を選択して、example5 ライブラリをロードしてください。このライブラリは、主要な S-PLUS データベースにはないいくつかのデータセットの例を含んでいます。
2. **Factor Analysis** ダイアログを開きます。
3. **Data Set** フィールドに `testscores.df` と入力します。
4. **Number of Factors** フィールドに 2 を指定します。
5. **Variables** フィールドで **<ALL>** を選択します。
6. **OK** をクリックします。

因子分析のサマリーが、**レポート・ウィンドウ**に表示されます。

主成分分析

多数の観測変量を含む調査では、考慮する変量を減らすため、元の変量の線形結合を検討することによって分析を簡単にできることがあります。たとえば学校の学力検査は、一般に様々な教科の試験を含みます。入学を志願した学生をランク付けするために、大学運営者はすべての教科の点数を 1 指標のみの点数に換算しようとするのがよくあります。「主成分分析」は、変量の

最適な線形結合を見つける一般的な方法です。

主成分分析を実行する

メインメニューから、統計 ▶ 多変量解析 ▶ 主成分分析を選択します。図 8.72 に示すように、**Principal Components Analysis** ダイアログが開きます。

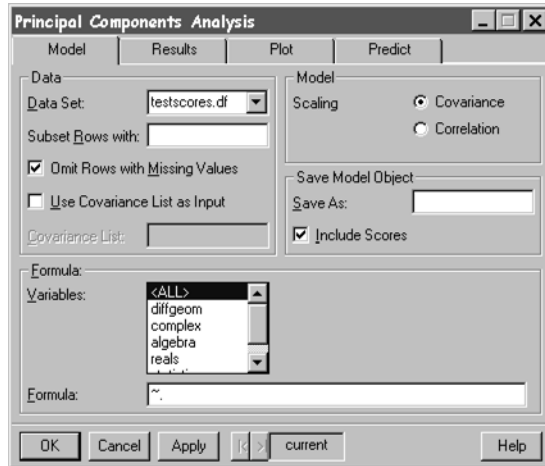


図 8.72 : Principal Components Analysis ダイアログ

例

416 ページの「因子分析」では、testscores.df データセットの因子分析を行いました。この例では、そのデータに主成分分析を実行します。

1. example5 ライブラリをまだロードしていない場合は、**ファイル ▶ ライブラリのロード**を選択して、example5 ライブラリをロードしてください。このライブラリは、testscores.df を含む、主要な S-PLUS データベースにはないいくつかのデータセットの例を含んでいます。
2. P.418 の方法で **Principal Components Analysis** ダイアログを開きます。
3. **Data Set** フィールドに testscores.df を入力します。
4. **Variables** フィールドで<ALL>を選択します。

5. **Plot** タブをクリックし、**Screplot** ボックスをチェックします。
6. **OK** をクリックします。

主成分分析のサマリーが、**レポート・ウィンドウ**に表示され、各主成分の固有値の棒グラフが、**グラフシート**に表示されます。

多変量分散分析

「多変量分散分析」は MANOVA として知られ、分散分析法を複数の反応変量まで範囲を拡張したものです。観測された複数種類の反応量は、単変量反応の集合ではなく 1 つの多変量観測値と見なされます。反応変量が独立している場合は、単変量解析だけを実行した方が賢明です。しかし反応変量に相関関係がある場合には、MANOVA の方が単変量解析よりも有益でかつ繰返しが少ないで済みます。

MANOVA を実行する

メインメニューから、**統計 ▶ 多変量解析 ▶ 多変量分散分析**を選択します。図 8.73 に示すように、**Multivariate Analysis of Variance** ダイアログが開きます。

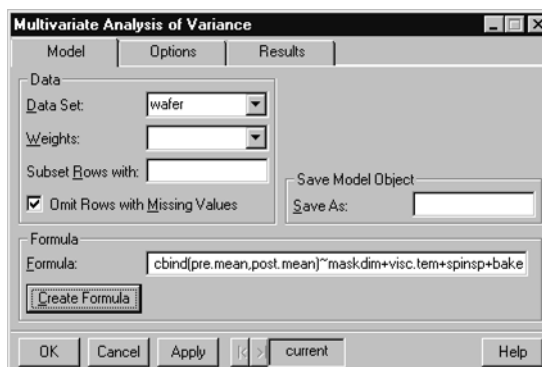


図 8.73 : Multivariate Analysis of Variance ダイアログ

例

データセット wafer は 18 行と 13 列から成っており、そのうち 8 列は因子、4 列は反応変量、1 列は補助変量 N です。これは、因子の各組み合わせにつき、2 つの集積回路ウェハを作成した実験の直交要因計画に基づく計画オブジェクトです。各ウェハについて、エッチング前とエッチング後に線幅を 5 回測定しました。反応変量は、実測値の平均値と尤離度です。ウェハのうち 3 つが壊れていたため、補助変量 N は実際に行われた実測値の数を示します。

各因子が反応変量に及ぼす影響を調べるために、MANOVA を使用して多変量反応として `pre.mean` と `post.mean` を処理したいと思います。

1. P.419 の方法で **Multivariate Analysis of Variance** ダイアログを開きます。
2. **Data Set** フィールドに `wafer` を入力します。
3. **Create Formula** ボタンをクリックして**モデル式ビルダー**を開きます。
4. CTRL キーを押しながら、**Variables** リストの `pre.mean` と `post.mean` を選択します。**Response** ボタンをクリックして、**Formula** にこれらの変量を反応として追加します。
5. `maskdim.` を選択します。**Variables** リストを `etchtime` が現われるまでスクロールします。**SHIFT** を押しながら、`etchtime` を選択します。これで、`maskdim` から `etchtime` までのすべての列が選択されます。**Main Effect** ボタンをクリックして、**Formula** に予測値としてこれらの変量を追加します。
6. **OK** をクリックして**モデル式ビルダー**を閉じます。MANOVA ダイアログの **Formula** フィールドには、構成したモデル式があります。
7. **OK** をクリックします。

MANOVA のサマリーが、**レポート・ウィンドウ**に表示されます。

QC チャート（品質管理チャート）

QC チャートは、プロセスデータの監視に有効です。「グループのあるデータ」の QC チャートは、プロセスが管理範囲内にあるかどうかを監視するのに役立ちます。「グループのないデータ」の QC チャートは、変動が群間変動ではなく連続した時間として傾向を持つ場合に適しています。また、「度数」（不良標本の数）と「比率」（不良標本の割合）の QC チャートを作成することもできます。

グループのあるデータ

Quality Control Charts (Continuous Grouped) ダイアログは、平均 (\bar{x})、標準偏差 (s)、および範囲 (r) の品質管理チャートを作成します。

QC チャートを作成する（グループのあるデータ）

メインメニューから、**統計 ▶ QC チャート ▶ グループのあるデータ**を選択します。図 8.74 に示すように、**Quality Control Charts (Continuous Grouped)** ダイアログが開きます。

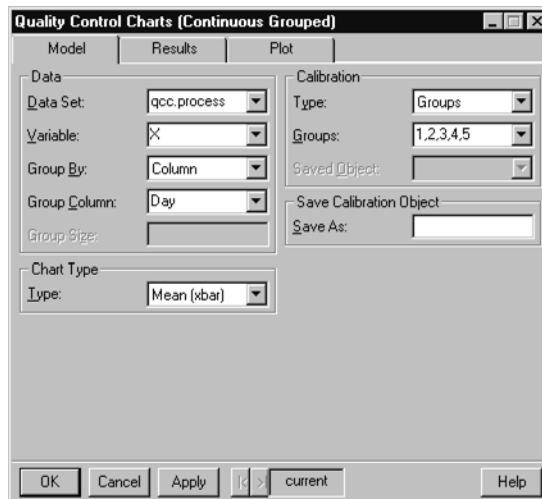


図 8.74 : Quality Control Charts (Continuous Grouped) ダイアログ

例

297 ページの「コルモゴロフ=スミルノフ適合度検定」では、プロセスをシミュレートした 200 個の測定値を含む `qcc.process` と呼ばれるデータセットを作成しました。20 日間にわたり、1 日 10 個の測定値を取得しました。この例では、プロセスが管理範囲内にあるかどうかを監視するために、`xbar` のシューハートの管理図を作成します。最初の 5 日間の観測値は、管理限界の設定値を計算するためのデータとして扱われます。

1. `qcc.process` データセットをまだ作成していない場合は、298 ページの指示にしたがって作成してください。
2. P.421 の方法で **Quality Control Charts (Continuous Grouped)** ダイアログを開きます。
3. **Data Set** フィールドに `qcc.process` を入力します。
4. **Variable** として `x` を選択します。
5. **Group Column** として `Day` を選択します。
6. **Calibration** グループの **Type** として **Groups** を選択します。
7. CTRL-クリックして、**Groups** リストボックスから **1, 2, 3, 4, 5** を選択します。
8. **OK** をクリックします。

`Day` によりグループ分けされた `x` データのシューハートの管理図が、**グラフシート**に表示されます。

グループのないデータ

Quality Control Charts (Continuous Ungrouped) ダイアログは、指数関数的に重み付けされた移動平均 (`ewma`)、移動平均 (`ma`)、移動標準偏差 (`ms`)、移動範囲 (`mr`) の QC チャートを作成します。これらのチャートは、変動が群間変動ではなく連続した時間として傾向をもつ場合に適しています。

QC チャート (グループのないデータ) を作成する

メインメニューから、**統計 ▶ QC チャート ▶ グループのないデータ**を選択します。図 8.75 に示すように、**Quality Control Charts (Continuous Ungrouped)** ダイアログが開きます。

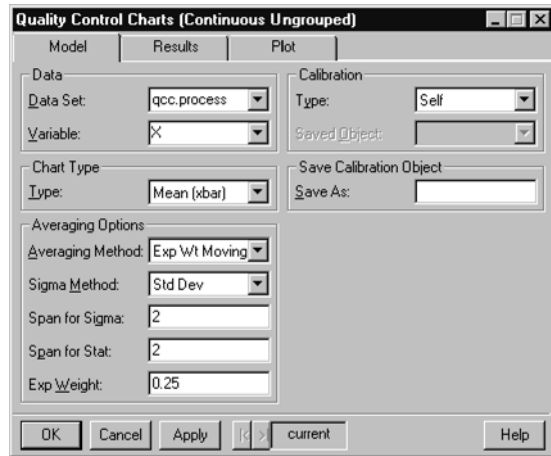


図 8.75 : Quality Control Charts (Continuous Ungrouped) ダイアログ

例

この例では、qcc.process が日次でグループ化されたデータであるという点を無視し、200 の観測値が連続する時点で得られたことにします。プロセスが管理限界内にあるかどうかを監視するために、指数関数的に重み付けした移動平均シューハートの管理図を作成します。

1. qcc.process データセットをまだ作成していない場合は、298 ページの指示にしたがって作成してください。
2. P.422 の方法で **Quality Control Charts (Continuous Ungrouped)** ダイアログを開きます。
3. **Data Set** フィールドに qcc.process と入力します。
4. **Variable** として X を選択します。
5. **OK** をクリックします。

シューハートの管理図が、**グラフシート**に表示されます。

P 管理図

Quality Control Charts (Counts and Proportions) ダイアログは、度数（不良標本の数）と比率（不良標本の比率）の QC チャートを作成します。

QC チャートを作成する（P 管理図）

メインメニューから、統計 ▶ QC チャート ▶ P 管理図を選択します。図 8.76 に示すように、**Quality Control Charts (Counts and Proportions)** ダイアログが開きます。

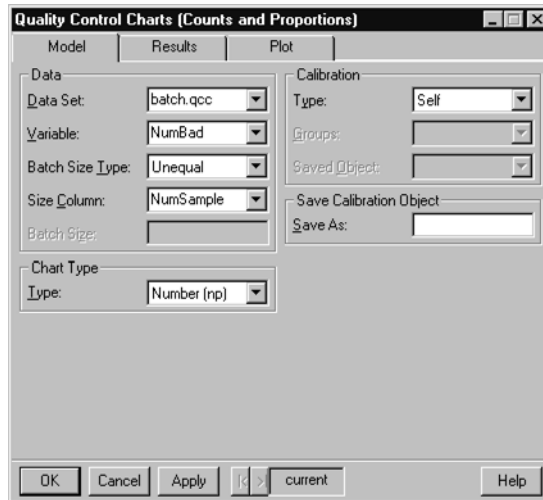



図 8.76 : Quality Control Charts (Counts and Proportions) ダイアログ

例

40 日間のロットにおける不良品目数を表すシミュレーションしたデータを含む S-PLUS データセット batch.qcc を作成します。最初の 10 日間ロットの大きさは 20 ですが、残りの 30 日は 35 ロットが選ばれました。

1. 標準ツールバーの**新規データセット**ボタン  をクリックして、新しいデータセットを開きます。
2. 第 1 列に次の 40 の数を入力します。

3 2 7 4 5 4 4 8 3 4

6 6 6 9 18 9 7 11 11 9

10 10 14 5 15 11 14 15 11 10

14 8 11 13 16 14 19 13 15 23

3. 第 1 列は、毎日の標本の中の不良標本の数を表します。v1 をダブルクリックし NumBad と入力して列名を変更します。ENTER を押すか、データ・ウィンドウのどこかをクリックして変更を確定します。
4. メインメニューからデータ ▶ 新しい列を変換式で作成を選択します。Data Set フィールドにデータセットの名前があることを確認し、Target Column フィールドに NumSample を入力します。Expression フィールドにコマンド `c(rep(20,10), rep(35,30))` を入力し、OK をクリックします。このステップで、整数 20 の複製を 10 個と、その後の整数 35 の複製を 30 個を含む NumSample という名前の列が作成されます。列 NumSample は、シミュレーションした観察値のロットサイズを表します。
5. データ・ウィンドウの左上角をダブルクリックしてデータセットの名前を変更します。現われたダイアログで、Name フィールドに batch.qcc を入力し、OK をクリックします。

これらのデータの Number (np) のシューハートの管理図を作成します。

1. P.424 の方法で Quality Control Charts (Counts and Proportions) ダイアログを開きます。
2. Data Set フィールドに batch.qcc と入力します。
3. Variable として NumBad を選択します。
4. Size Column として NumSample を選択します。
5. Chart Type として Number (np) を選択します。
6. OK をクリックします。

NumSample で示されたグループサイズの NumBad データのシューハートの管理図が、グラフシートに表示されます。

標本からのサンプリング

統計解析において、研究者は通常統計値の点推定だけでなく、その点推定の変動とパラメータの真の値の信頼区間を得たいと考えます。たとえば研究者は、標本平均だけでなく平均の標準誤差と平均の信頼区間を計算することができます。

標準誤差と信頼区間を計算する今までの方法は、一般に漸近的に正規分布した統計値またはその既知の変換により行われてきました。この正規性の仮定が維持されない場合、今までの方法は的確でないことがあります。ブートストラップやジャックナイフなどの標本作成方法は、平均など、任意の統計量の標準誤差、信頼区間、および分布の推定を行います。これらの方法を使用するために、調査対象のデータセット名と、対象とする統計量を計算するS-PLUS 関数または式を用意しなければなりません。

ブート ストラップ 推定

「ブートストラップ」では、指定された数の新しい標本が、対象のデータセットから繰り返し抽出されます。対象の統計量は各繰り返しで得られたデータごとに計算され、得られた一組の推定値は統計値の経験分布として使用されます。

ブートストラップ推定を実行する

メインメニューから、統計 ▶ 標本からのサンプリング ▶ ブートストラップを選択します。図 8.77 に示すように、**Bootstrap Inference** ダイアログが開きます。

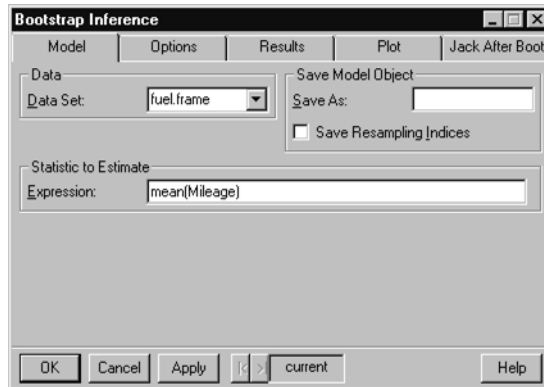


図 8.77 : Bootstrap Inference ダイアログ

例 1

データセット `fuel.frame` は、1990 年 4 月発行の『Consumer Reports』からの引用です。このデータは、60 の観測値（行）と 5 つの変量（列）を含んでいます。60 台の車のそれぞれについて、重量、エンジン排気量、燃費、種類、および燃料消費率の観測値を得ました。変数 `Mileage` の平均のブートストラップ推定値と平均の変動を求めます。

1. P.426 の方法で **Bootstrap Inference** ダイアログを開きます。
2. **Data Set** フィールドに `fuel.frame` と入力します。
3. **Expression** フィールドに `mean(Mileage)` と入力します。
4. **Options** ページで、標本からのサンプリングをデフォルトの数よりも少なく実行するために、**Number of Resamples** フィールドに **250** を入力します。これで、この例に必要な計算が早くなります。
5. **Plot** ページをクリックし、デフォルトで **Distribution of Replicates** プロットが選択されていることに注意してください。
6. **OK** をクリックします。

ブートストラップサマリーが、レポート・ウィンドウに表示され、確率密度線付きヒストグラムがグラフシートにプロットされます。

例 2

この例では、線形モデルの係数の平均と変動のブートストラップ推定値を求めます。使用するモデルは、`fuel.frame` データセットの `Weight` と `Disp.` から `Mileage` を予測します。

1. P.426 の方法で **Bootstrap Inference** ダイアログを開きます。
2. **Data Set** フィールドに `fuel.frame` を入力します。
3. **Expression** フィールドに次の式を入力します。

```
coef(lm(Mileage ~ Weight+Disp., data=fuel.frame))
```

4. **Options** ページで、標本からのサンプリングをデフォルトの数よりも少なく実行するために、**Number of Resamples** フィールドに **250** を入力します。これで、この例に必要な計算が速くなります。
5. **Plot** ページをクリックし、デフォルトで **Distribution of Replicates** プロットが選択されていることに注意してください。
6. **OK** をクリックします。

ブートストラップサマリーが、レポート・ウィンドウに表示されます。さらに、(各係数ごとに) 密度線を備えた 3 つのヒストグラムが、グラフシートにプロットされます。

ジャックナイフ推定

「ジャックナイフ」ではデータを複製し、それぞれの標本から 1 つの観測値を除外することによって新しい標本を抜き出します。対象の統計値が 1 つの観測値を除外したデータごとに計算され、このジャックナイフ法による分布から推定値が計算されます。

ジャックナイフ推定を実行する

メインメニューから、統計 ▶ 標本からのサンプリング ▶ ジャックナイフを選択します。図 8.78 に示すように、**Jackknife Inference** ダイアログが開きます。

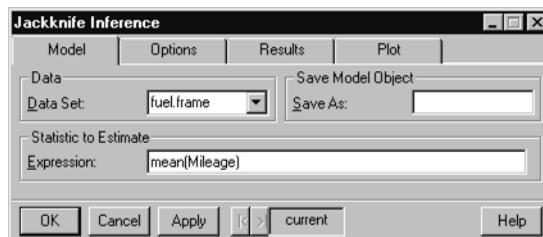


図 8.78 : Jackknife Inference ダイアログ

例 1

`fuel.frame` データについて Mileage の平均のジャックナイフ推定値と平均の変動を求めます。

1. P.428 の方法で **Jackknife Inference** ダイアログを開きます。
2. **Data Set** フィールドに `fuel.frame` と入力します。
3. **Expression** フィールドに `mean(Mileage)` を入力します。
4. **Plot** ページをクリックし、デフォルトで **Distribution of Replicates** プロットが選択されていることを確認してください。
5. **OK** をクリックします。

ジャックナイフサマリーが、**レポート・ウィンドウ**に表示され、確率密度線付きヒストグラムが、**グラフシート**にプロットされます。

例 2

この例では、線形モデルの係数のジャックナイフによる平均と変動の推定値を求めます。使用するモデルは、`fuel.frame` データセットにおいて `Weight` と `Disp.` から `Mileage` を予測します。

1. **Jackknife Inference** ダイアログを開きます。
2. **Data Set** フィールドに `fuel.frame` と入力します。
3. **Expression** フィールドに次のように入力します。


```
coef(lm(Mileage ~ Weight+Disp., data=fuel.frame))
```
4. **Plot** ページをクリックし、デフォルトで **Distribution of Replicates** プロットが選択されていることを確認してください。
5. **OK** をクリックします。

ジャックナイフサマリーが、**レポート・ウィンドウ**に表示され、(各係数ごとの) 3つの密度線付きヒストグラムが、**グラフシート**にプロットされます。

平滑化

平滑化法は、1変量反応を1変量予測変量の平滑関数としてモデル化します。標準の回帰法では、線形パラメトリック関数を散布図のデータにあてはめます。しばしば、どのパラメトリック関数を使用すべきか決定するのに十分な情報がないことがあります。そのような場合は、特定の関係を想定しない「ノンパラメトリック曲線」をあてはめることができます。

ノンパラメトリック曲線近似は、データの一般的な傾向を示す平滑曲線を作成しようとするため、「スモーカー」とも呼ばれます。最も簡単なスモーカーは「移動平均」です。特定の x 値におけるあてはめが近くの点の y 値の加重平均として計算されます。その加重は、その x 値と対象となる x 値との間の距離が大きくなるほど小さくなるように設定されます。最も簡単な種類の移動平均スモーカーでは、対象となる点から一定の距離（またはウィンドウ）内にあるすべての点が、その点の平均を求めるのに等しい重みを持ちます。ウィンドウ幅は、スモーカーの「帯域幅」と呼ばれ、通常、データ点の総数に対する割合で与えられます。帯域幅を広くするとより滑らかな近似を得ることができますが、局所的に変化する特徴をとらえることができなくなることがあります。帯域幅を狭くすると、スモーカーは局所的に変化する特徴をよりの確に追跡できるようになりますが、全体的な傾向をつかみにくくなります。

複雑なスモーカーは計算区間の移動にいろいろな変更を加えたものです。たとえば、滑らかに減少する重みや局所回帰直線などを使用することができます。しかしスモーカーにはすべて、曲線の滑らかさを制御する滑らかさパラメータ（帯域幅）があります。適切な帯域幅を選択する問題は複雑で、多くの統計に関する研究の論文で扱われてきました。しかし、実際のデータにいくつかのスモーカーを試すことによって、帯域幅を変化させて得られる実際の結果をつかむことができます。

この節では、4つの異なるタイプのスモーカーを使用する方法を説明します。

- **核関数による平滑化**：様々な重み関数すなわち「核」を使用することができる、移動平均の一般化です。重み関数は、点の推移を単純な移動平均手法よりも滑らかにします。
- **局所平滑化**：データに対する局所線形回帰または局所二次近似に基づくノイズ削減手法です。
- **スプライン平滑化**：滑らかな曲線を得るために一連の多項式をつなぎ合わせた方法。
- **スーパースモーカー**：高度に自動化された可変スパンスモーカー。様々な帯域幅で重み付けしたスモーカーの組み合わせにより、あてはめ

値を求めます。

核関数による平滑化

「核関数による平滑化」は、様々な重み関数すなわち「核」を使用することができる移動平均の一般化です。重み関数は、点の推移を単純な移動平均手法よりも滑らかにします。デフォルトでは、**S-PLUS** 核関数による平滑化の帯域幅は 0.5 であり、これは、各平滑化ウィンドウ内のデータが点全体の約半分であることを表しています。

デフォルトのカーネルは、「ボックス」または「ボックスカーブスモーカー」で、平滑化ウィンドウ内の各点に同じように重み付けします。他の選択肢には、三角形カーネル、**Parzen** カーネル、およびガウスカーネルがあります。「三角形カーネル」では、対象の点から距離が離れるにつれて重みが線形に減少し、したがって、平滑化ウィンドウの端の点の重みはゼロに近くなります。

「**Parzen** カーネル」は、三角形で畳み込まれたボックスです。「正規」または「ガウスカーネル」の場合、重みは、対象の点から遠ざかるにつれてガウス（正規）分布に従って減少します。

局所回帰 (Loess)

「局所回帰」または「loess」は、ベル研究所の W.S.Cleveland らによって開発されました。これは、本質的にノイズ削減アルゴリズムである平滑化の優れた手法です。局所平滑化は、データに対する局所線形回帰または局所二次近似に基づいています。各点において、線または放物線が、平滑化ウィンドウ内の点にあてはめられ、予測値は、対象の点に関する y 値になります。各ウィンドウ内の線または放物線を計算するために加重最小二乗法が使用されます。計算した y 値を結ぶと滑らかな曲線ができます。

局所平滑化では、帯域幅はスモーカーの「スパン」と呼ばれます。スパンは、特定の平滑化ウィンドウのあてはめに用いられる点の、全体に対する割合を示す 0~1 の値です。値が小さいほど滑らかでなくなり、0 に近い小さい値は推奨されません。スパンを指定しない場合は、交差確認法を使って適切な値が計算されます。

標本が少ない場合 ($n < 50$) または x の値に近傍観測値間で大きい系列相関がある場合は、あらかじめ指定した固定のスパンスモーカーを使用しなければなりません。

スプライン平滑化

「スプライン平滑化」は、一連の多項式をつなぎ合わせることによって計算されます。三次式近似は、このクラスの平滑化に最も幅広く使用されており、局所的な三次多項式も含まれます。局所的な多項式は、ペナルティ付きの残差の平方和を最小にすることによって計算されます。滑らかさは、交わる点において隣り合った多項式の値、勾配、および曲率を一致させることによって保証されます。多項式をつなぎ合わせると、データに対するあてはめが滑らかになります。スプライン平滑化のデータ値へのあてはめが正確になるほど曲線が粗くなり、また、その逆も言えます。

スプラインの平滑化パラメータは、「自由度」と呼ばれます。自由度は、あてはめにおける曲率の大きさを制御し、局所的多項式の自由度に対応します。自由度が低いほど曲線は滑らかになります。自由度によって、あてはめの滑らかさとデータ値に対するあてはまりの良さとの間のトレードオフを調整することにより、自動的に平滑化ウィンドウを決定します。データ点が n 個の場合、自由度は、1 から $n-1$ の間でなければなりません。自由度を $n-1$ と指定すると、各データ点を曲線が正確に通るようになります。S-PLUS は、デフォルトで、三次スプラインに相当する自由度 3 を使用します。

スーパー スムーザー

「スーパー smoother」は、高度に自動化された可変スパン smoother です。様々な帯域幅で重み付けした smoother の組み合わせによって、あてはめ値を求めます。スーパー smoother の平滑化パラメータは、「スパン」と呼ばれます。スパンは、特定の平滑化ウィンドウのあてはめに用いられる点の全体に対する割合を表す 0~1 の値です。値が小さいほど滑らかでなくなり、0 に近い小さい値は推奨されません。スパンが指定されない場合は、交差確認法を使って適切な値が計算されます。標本が少ない場合 ($n < 50$) または x の値に近傍観測値間で大きい系列相関がある場合は、あらかじめ指定した固定のスパン smoother を使用しなければなりません。

例

データセット air は、111 の観測値 (行) と 4 つの変量 (列) を含みます。このデータは、連続 111 日間にわたって、オゾンと太陽放射線と気温と風速の 4 つの変量を実測した環境調査から得たものです。ここでは、オゾンと放射線の関係を示す平滑化プロットを作成します。

1. 統計 ▶ 平滑化 ▶ 核関数による平滑化を選択します。Data Set として air を入力します。x Columns として radiation を、y Columns として ozone を選択し、次に OK をクリックします。核関数による平滑化線とともに、radiation に対する ozone の関係を含むグラフシートが作成されます。
2. 統計 ▶ 平滑化 ▶ 局所回帰を選択します。Data Set として air を入力します。x Columns として radiation を、y Columns として ozone を選択し、次に OK をクリックします。局所平滑化線とともに、radiation に対する ozone の関係を含むグラフシートが作成されます。
3. 統計 ▶ 平滑化 ▶ スプライン平滑化を選択します。Data Set として air を入力します。x Columns として radiation を、y Columns として ozone を選択し、次に OK をクリックします。滑らかなスプライン平滑化線とともに、radiation に対する ozone のプロットを含むグラフシートが作成されます。

4. **統計 ▶ 平滑化 ▶ スーパー smoother**を選択します。**Data Set**として `air` を入力します。**x Columns**として `radiation` を、**y Columns**として `ozone` を選択し、次に **OK** をクリックします。スーパー smoother 平滑化線とともに、`radiation` に対する `ozone` のプロットを含む **グラフシート** が作成されます。

時系列解析

時系列解析手法は、日々の実測値のような連続的な観測値に適用されます。線形回帰のようにほとんどの統計手法では、データの観測値（行）同士の相関性は重要ではありません。これと対照的に、時系列手法は隣り合った観測値の相関性を調べます。

この節では、**統計 ▶ 時系列解析**メニューから利用できる時系列について説明します。

- **自己相関**：連続する観測値の自己相関／自己共分散／偏自己相関を計算します。
- **ARIMA**：連続する観測値に自己回帰和分移動平均モデル（ARIMA）をあてはめます。これらは、自己回帰／移動平均／季節成分を含むきわめて一般的なモデルです。
- **ラグプロット**：時系列とその時系列をずらした時系列との関係をプロットします。
- **スペクトル**：スペクトル推定の結果をプロットします。

これらの方法を使用して、環境データセットの構造を調べます。

自己相関

「自己共分散関数」は、1変量時系列の系列（または一時的）依存構造を説明するための重要なツールです。遅れのある観測値の間にどれだけ相関があるかを反映します。

自己相関をプロットする

メインメニューから、**統計 ▶ 時系列解析 ▶ 自己共分散・相関**を選択します。図 8.79 に示すように、**Autocorrelations and Autocovariances** ダイアログが開きます。

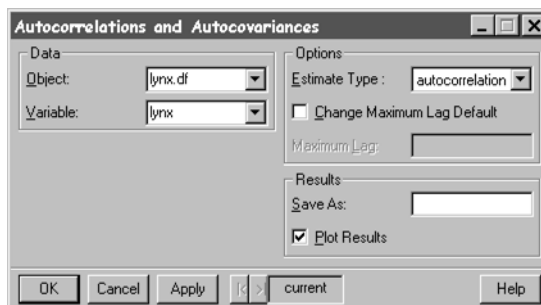


図 8.79 : Autocorrelations and Autocovariances ダイアログ

例

データセット例 lynx.df は、1821 年から 1934 年のデータはライブラリ example5 にあります。カナダ北西部マッケンジー川流域の年間のオオヤマネコ捕獲数を含みます。図 8.80 は、そのデータを表します。

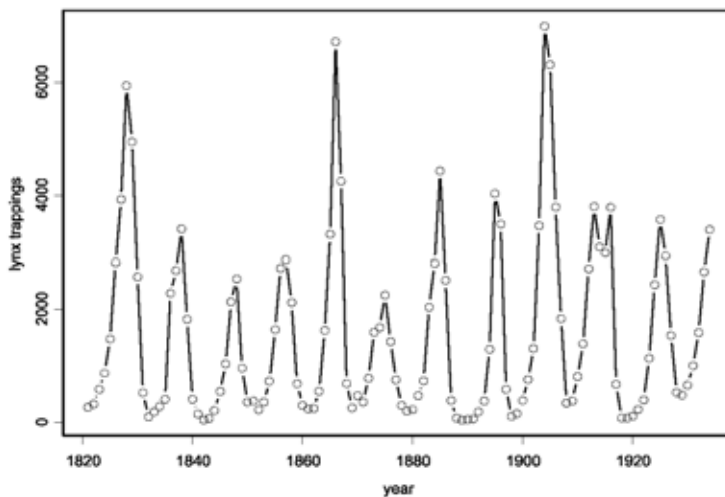


図 8.80 : カナダ北西部マッケンジー川流域のオオヤマネコ捕獲

データには明らかな周期があります。自己相関を使用して周期の長さを調べることができます。

1. **ファイル ▶ ライブラリのロード**を選択して、example5 ライブラリをロードします。**Library Name** リストの example5 をハイライトし、**OK** をクリックします。このライブラリは、主要な S-PLUS データベースにはないいくつかのデータセットの例を含んでいます。
2. P.435 の方法で **Autocorrelations and Autocovariances** ダイアログを開きます。
3. **Data Set** フィールドに lynx.df と入力します。
4. **Variable** として lynx を選択します。
5. **OK** をクリックします。

図 8.81 は、得られた自己相関プロットを表します。10 近辺の山と 5 近辺の谷は、10 年周期を表します。

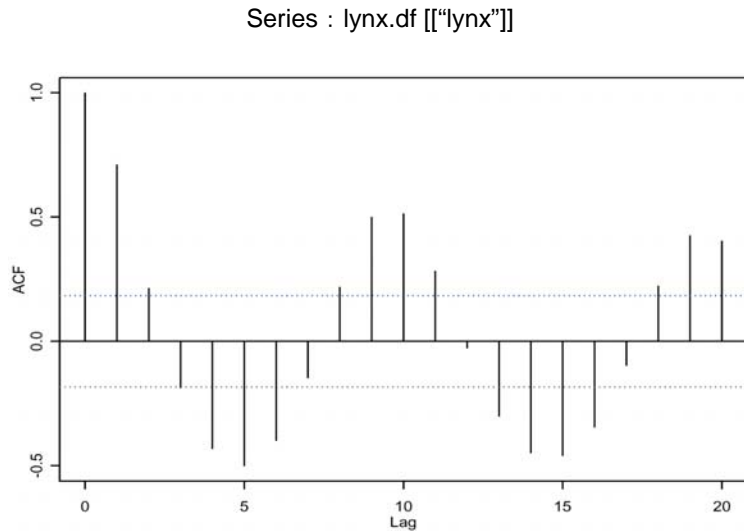


図 8.81 : lynx.df データの自己相関プロット

ARIMA

「自己回帰和分移動平均」(ARIMA) モデルは、予測／品質管理／季節調整／スペクトル推定を含む幅広い時系列解析に有効であり、データのサマリーを提供します。

ARIMA モデルをあてはめる

メインメニューから、統計 ▶ 時系列解析 ▶ ARIMA モデルを選択します。
図 8.82 に示すように、ARIMA Modeling ダイアログが開きます。

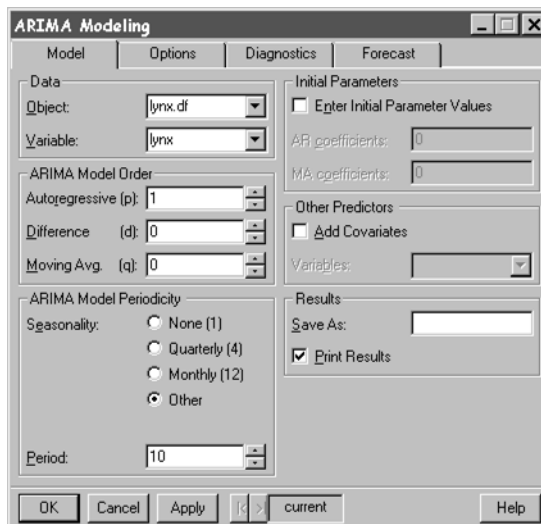


図 8.82 : ARIMA Modeling ダイアログ

例

434 ページの「自己相関」では、lynx.df 時系列の自己相関を計算しました。図 8.81 の自己相関プロットは、相関が 10 年周期の lynx.df データの観測値間の相関関係を表します。10 年周期の自己回帰モデルとして、これをモデル化することができます。

1. example5 ライブラリをまだロードしていない場合は、ファイル ▶ ライブラリのロードを選択して、example5 ライブラリをロードしてください。example5 ライブラリは、lynx.df データを含みます。
2. P.437 の方法で ARIMA Modeling ダイアログを開きます。
3. **Data Set** フィールドに lynx.df と入力します。
4. **Variable** として lynx を選択します。
5. **Model Order** グループの **Autoregressive** を 1 に指定します。
6. **Seasonality** として **Other** を選択します。

7. ARIMA Model Periodicity の **Period** を 10 に指定します。

8. **OK** をクリックします。

ARIMA モデルのサマリーが、レポート・ウィンドウに表示されます。

```

*** ARIMA Model Fitted to Series lynx.df[["lynx"]] ***

Call: arima.mle(x = lynx.df[["lynx"]], model = model,
  xreg = xreg, max.iter = nlmin.max.iter, max.fcal =
  nlmin.max.fcal)
Method: Maximum Likelihood
Model : 1 0 0
Period: 10

Coefficients:
  AR : 0.73883

Variance-Covariance Matrix:
      ar(10)
ar(10) 0.004366605

Optimizer has converged
Convergence Type: relative function convergence
AIC: 1793.16261
    
```

ラグプロット

Lag Plot ダイアログは、時系列とその時系列の遅れの間をプロットします。

ラグプロットを作成する

メインメニューから、**統計 ▶ 時系列解析 ▶ ラグプロット** を選択します。

図 8.83 に示すように、**Lag Plot** ダイアログが開きます。



図 8.83 : Lag Plot ダイアログ

例

434 ページの「自己相関」では、lynx.df 時系列の自己相関を計算しました。この例では、ラグプロットを使用して様々な遅れの観測値間の相関関係の例を示します。

1. **example5** ライブラリをまだロードしていない場合は、**ファイル ▶ ライブラリのロード**を選択して、**example5** ライブラリをロードしてください。**example5** ライブラリは、lynx.df データを含んでいます。
2. P.438 の方法で **Lag Plot** ダイアログを開きます。
3. **Data Set** フィールドに lynx.df と入力します。
4. **Variable** として lynx を選択します。
5. **Lag** を **4** にします。
6. **Rows** として **2**、**Columns** として **2** を選択し、2 行 2 列の配置にします。
7. **OK** をクリックします。

lynx.df データのラグプロットが、**グラフシート**に表示されます。

スペクトル

Spectrum Plot ダイアログは、スペクトル推定の結果をプロットします。このプロットは平滑化されたピリオドグラムか自己回帰パラメータのどちらかを使用して、時系列の推定スペクトルを表します。

スペクトルプロットを作成する

メインメニューから、**統計 ▶ 時系列解析 ▶ スペクトル**を選択します。図 8.84 に示すように、**Spectrum Plot** ダイアログが開きます。

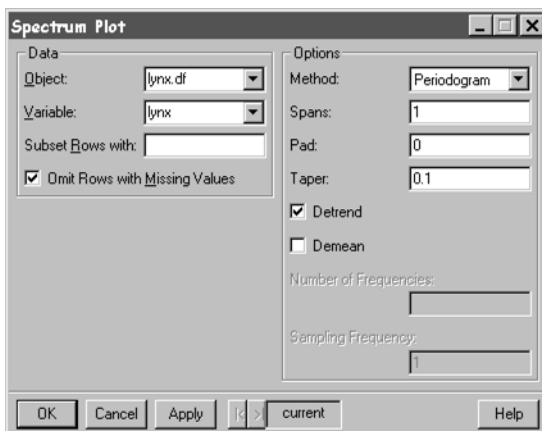


図 8.84 : Spectrum Plot ダイアログ

例

434 ページの「自己相関」では、lynx.df 時系列の自己相関を計算しました。この例では、時系列における周期性を調べるために lynx.df データの平滑化されたピリオドグラムをプロットします。

1. example5 ライブラリをまだロードしていない場合は、**ファイル ▶ ライブラリのロード**を選択して、example5 ライブラリをロードしてください。example5 ライブラリは、lynx.df データを含んでいます。
2. P.439 の方法で **Spectrum Plot** ダイアログを開きます。
3. **Data Set** フィールドに lynx.df と入力します。
4. **Variable** として lynx を選択します。
5. **OK** をクリックします。

lynx.df データのスペクトルプロットが、**グラフシート**に表示されます。

乱数と分布

データメニューには、乱数を生成し理論分布に関する値を計算するためのツールがあります。そのような方法には、次のようなものがあります。

- **ランダムサンプリング**：データセットからサンプリングします。
- **分布関数**：特定の分布に関する計算を行います。
- **乱数**：指定された分布から乱数のデータセットを作成します。

あまり統計的でないその他のデータ処理については、第 2 章の「データによる作業」で説明しています。

ランダム サンプリング

データセットがあり、そのデータから何行かをランダムに選択したいことがあります。**Random Sample of Rows** ダイアログでこれを行います。またこのダイアログを使用して、データセットの行をランダム化する（並び替える）ことができます。

行のランダムサンプリングを行う

データ ▶ ランダム・サンプリングを選択します。図 8.85 に示すように、**Random Sample of Rows** ダイアログが開きます。

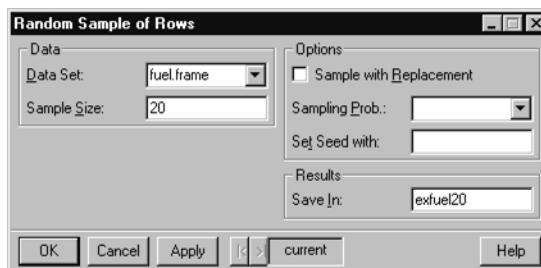


図 8.85 : Random Sample of Rows ダイアログ

例

データセット `fuel.frame` は、1990 年 4 月発行の『Consumer Reports』からの引用です。このデータは、60 の観測値（行）と 5 つの変数（列）を含んでいます。60 台の車のそれぞれについて、重量、エンジン排気量、燃費、種類、および燃料消費率の観測値を得ました。`fuel.frame` データにおいて、60 台のなかから 20 台の車をランダムにサンプリングし、それを新しく `exfuel20` と名前を付けたデータセットに入れます

1. P.441 の方法で **Random Sample of Rows** ダイアログを開きます。
2. **Data Set** フィールドに `fuel.frame` と入力します。
3. **Sample Size** に 20 と入力し、**Save In** フィールドに `exfuel20` と入力します。
4. **OK** をクリックします。

新しい標本を含むデータセット `exfuel20` が作成され、**データ・ウィンドウ** に表示されます。

Random Sample of Rows ダイアログを使用して、行をランダム化する（並び替える）ことができます。これを行うために、**Sample Size** をデータセットの観測値と同じ数に設定し、**Sample with Replacement** オプションをチェックしてください。すべての行が新しいデータセット内にもありますが、その順序はランダムに変更されます。

分布関数

Distribution Functions ダイアログは、指定した分布から確率密度／累積確率／確率点を計算します。このダイアログの目的は、データまたは連続する数から分布情報を生成することです。このダイアログを使用して、検定の p 値と棄却域を計算したり、様々な分布をプロットして視覚化することができます。

分布関数を計算する

メインメニューから、**データ ▶ 分布関数** を選択します。図 8.86 に示すように、**Distribution Functions** ダイアログが開きます。

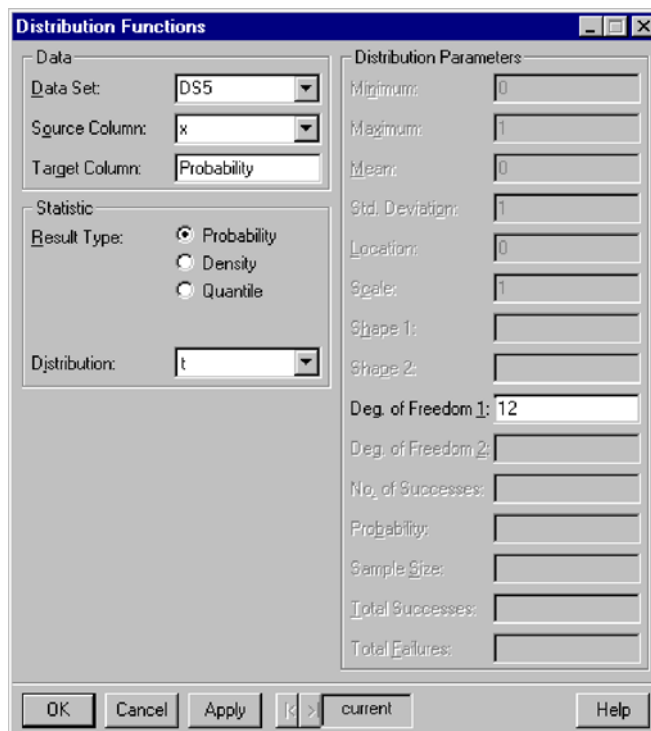


図 8.86 : Distribution Functions ダイアログ


例 1 : p 値を計算する

Distribution Functions ダイアログを使用して、統計検定のための棄却域と p 値を生成することができます。たとえば、両側の t 検定を行うとします。帰無仮説と対立仮説は、次の通りです。


$$H_0 : \mu_1 = \mu_2$$

$$H_A : \mu_1 \neq \mu_2$$

t 統計値は、自由度 12 で -2.10 です。この検定の p 値はいくつでしょうか。

1. 標準ツールバーの**新規データセット**ボタン  をクリックして、新しいデータセットを作成します。

2. 第 1 列に値-2.10 および 2.10 を入力します。t 分布がゼロに関して対称であるため、-2.10 と 2.10 の両方における確率を計算して両端の値を求めます。
3. 第 1 列の名前を **x** に変更します。
4. P.442 の方法で **Distribution Functions** ダイアログを開きます。
5. **Data Set** フィールドにデータセットの名前が現われたことを確認し、**Source Column** として **x** を選択します。
6. p 値を計算するときは、**Result Type** を **Probability** と選択します。**Distribution** プルダウンメニューから **t** を選択し、**Deg. of Freedom 1** に **12** を入力します。
7. **OK** をクリックします。

データセットに新しく **Probability** と名前のついた列が現われます。列は、**x** の値に対応する累積確率 **0.02877247** および **0.97122753** を含んでいます。少数桁数を多くして表示させるためには、**Probability** 列をハイライトし、**データ・ウィンドウのツールバーの小数点表示桁上げボタン**  をクリックします。

計算した確率を p 値に変換するときは、次のように計算します。

$$p = 2(1 - F_t(|t|))$$

これは両側検定であるため、結果を 2 倍にしなければなりません。累積確率関数から p 値を計算する方法は、常に対立仮説に依存します。

$$H_A : \mu_1 < \mu_2 \quad p \text{ 値} = F_t(t)$$

$$H_A : \mu_1 > \mu_2 \quad p \text{ 値} = 1 - F_t(t)$$

$$H_A : \mu_1 \neq \mu_2 \quad p \text{ 値} = 2(1 - F_t(|t|))$$

ここで $F_t(t)$ は、自由度 ν の t 分布の累積分布関数です。t 統計量が -2.10 の場合、これは、p 値が $2(1 - 0.97122753) = 0.05754494$ であることを示します。


例 2 : 棄却域を計算する


水準 α 検定の棄却域の計算は簡単です。たとえば、ある母集団の分散が別の母集団の分散より大きいかどうかを検定します。すなわち、帰無仮説と対立仮説は次のようになります。

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_A : \sigma_1^2 > \sigma_2^2$$

標本データからの $f=3.9$ の F 統計量を計算し、2 つの標本の数が $m=12$ と $n=10$ であると仮定します。水準 0.05 の検定では、 $f \geq F_{\alpha, m-1, n-1} = F_{0.05, 11, 9}$ の場合に帰無仮説を拒否します。棄却域を求めるときは、自由度が 11 と 9 の F 分布の $1-0.05=0.95$ における分位数を計算しなければなりません。

1. **標準** ツールバーの **新規データセット** ボタン  をクリックして、新しいデータセットを開きます。
2. 第 1 列の第 1 行に値 0.95 を入力します。第 1 列の名前を **x** に変更します。
3. P.442 の方法で **Distribution Functions** ダイアログを開きます。
4. **Data Set** フィールドにデータセットの名前があることを確認し、**Source Column** として **x** を選択します。
5. 確率点を計算するときには、**Result Type** を **Quantile** と選択します。
6. **Distribution** プルダウンメニューから **f** を選択します。**Deg. of Freedom 1** と **Deg. of Freedom 2** フィールドにそれぞれ 11 と 9 を入力します。
7. **OK** をクリックします。

データセットに **Quantile** と名前が付けられた新しい列が現われます。列は、**x** の値に対応する確率点 3.102485 を含んでいます。したがって、水準 0.05 における自由度 11 および 9 の F 分布の棄却域は 3.102485 です。少数桁数をさらに多く表示させるときは、**Quantile** 列をハイライトし、**データ・ウィンドウ** のツールバーの **小数点表示桁上げ** ボタン  をクリックします。

計算した F 統計量 3.9 が棄却域にあるため、帰無仮説を否定し、母集団 1 の分散が母集団 2 の分散よりも大きいと結論づけられます。同様の手順によって、多くの一般的な検定の棄却域を計算することができます。

例 3 : 正規分布をプロットする

Distribution Functions ダイアログを使用すると、図 8.87 に示す正規分布のような確率分布を簡単にグラフにすることができます。

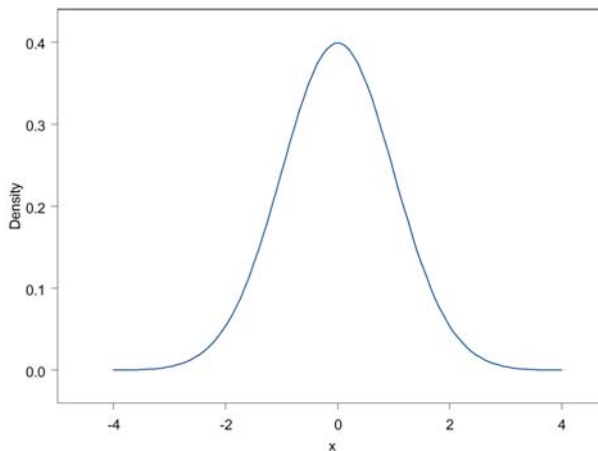



図 8.87 : 平均 0 および標準偏差 1 の正規分布

以下のステップで、図 8.87 の正規分布のプロットを生成します。

1. 標準ツールバーの**新規データセット**ボタン  をクリックして、新しいデータセットを開きます。
2. メインメニューから、**データ ▶ 埋める**を選択します。
3. **Columns** フィールドに **x** を、**Length** として 100 を入力します。**Start** 値として -4 を、**Increment** として 0.0808 を指定します。
4. **OK** をクリックします。これにより、-4 と 4 の間に 100 個の等間隔の値を含む **x** と名前が付けられた列が生成されます。
5. P.442 の方法で **Distribution Functions** ダイアログを開きます。

6. **Data Set** フィールドにデータセットの名前があることを確認し、**Source Column** として **x** を選択します。
7. 密度値を計算するときは、**Density** オプションをチェックします。
8. **Distribution** プルダウンメニューから **normal** を選択します。**Mean** と **Std.Deviation** がデフォルト値 **0** と **1** になっていることを確認してください。
9. **OK** をクリックします。各 **x** 値に対応する密度が生成され、それが新しい **Density** 列に挿入されます。
10. **データ**・ウィンドウの **x** 列をハイライトし、次に **CTRL**-クリックして **Density** 列を同時にハイライトします。
11. **2D** プロットパレットを開き、**線グラフ** ボタンをクリックします。**グラフシート** に **Density** と **x** の関係がプロットされます。

以上の手順で様々な分布のプロットを生成することができます。

乱数

Random Numbers ダイアログを使用して、様々な分布から乱数を生成することができます。

乱数を生成する

メインメニューから、**データ ▶ 乱数** を選択します。図 8.88 に示すように、**Random Numbers** ダイアログが開きます。

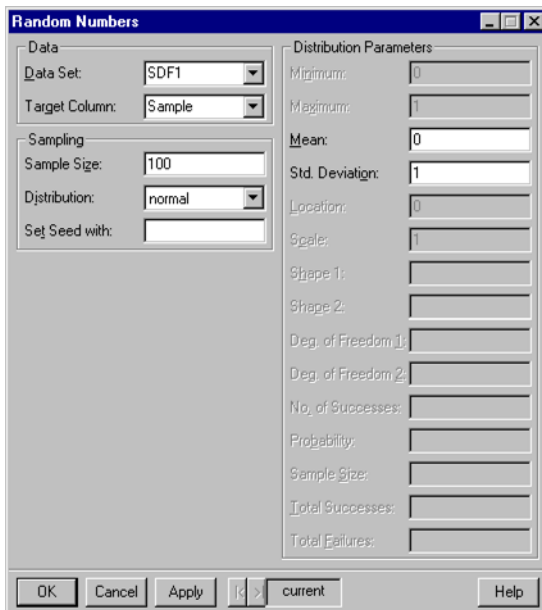


図 8.88 : Random Numbers ダイアログ

例

分布の直感的理解を深める 1 つの方法は、サンプリングしたデータのヒストグラムを繰り返しプロットすることです。図 8.87 の正規分布の密度関数を検討します。母集団が正規分布であるということは、サンプリングされたデータもその形を持つということです。すなわちこの母集団から 100 の観測値をサンプリングし、そのデータからヒストグラムを作成する場合は、ヒストグラムが図 8.89 と似ていると予想します。これが実際に有効かどうか検定してみましよう。

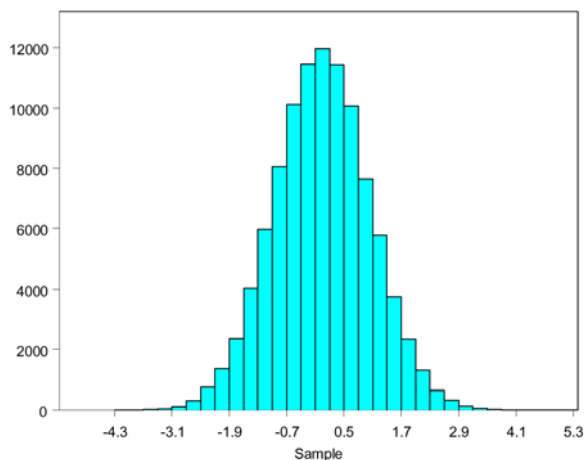




図 8.89 : 正規データのヒストグラム

1. 標準ツールバーの**新規データセット**ボタン  をクリックして、新しいデータセットを開きます。
2. P.447 の方法で **Random Numbers** ダイアログを開きます。
3. **Data Set** フィールドにデータセットの名前が現われたことを確認します。
4. **Target Column** のフィールドで **Sample** と入力します。
5. **Sample Size** に 100 を入力し、**Distribution** プルダウンメニューから **normal** を選択します。**Mean** と **Std.Deviation** がデフォルト値 0 と 1 であることを確認してください。
6. **Apply** をクリックします。乱数の標本が生成され、それがデータセットの **Sample** 列に入ります。
7. **データ・ウィンドウ** の **Sample** 列をハイライトします。
8. **2D プロットパレット** を開き、**ヒストグラム** ボタン  をクリックします。100 の観測値のヒストグラムがプロットされます。
9. 棒上にカーソルを移動させ、右クリックします。**Options** を選択し、ダイアログの **Lower Bound** に -5、**Upper Bound** に 5、**Interral Width** に 0.5 と入力し、**OK** をクリックします。
10. ヒストグラムが図 8.87 や図 8.89 のものとあまりにも似ていないことに驚かされるでしょう。**Random Numbers** ダイアログがまだ開いているので、**Apply** を再びクリックするだけでデータの新しい標本を生成することができます。新しいデータセットが生成され、新しいデー

タのヒストグラムが自動的に再描画されます。**Apply** を何度かクリックして、ヒストグラムの違いを調べてください。

11. **Sample Size** を変化させて、ヒストグラムがどのように変化するか観察してください。

標本数が多いほど、ヒストグラムは図 8.89 のようになるはずですが、データが大きい場合でも、ヒストグラムが図 8.89 とほぼ完全に同じになることはめったにありません。標本数が少ないと、得られる形が予想するような正規形のヒストグラムに近づかないことがあります。しかし、いずれのデータの母集団も正規です。このことは、正規分布からデータを生成しているため明らかです。

ここで大切なのは、データが小さいときには、ヒストグラムが正しく正規分布に見えないことにあまりとらわれない方がよいということです。データが正規分布であると確信できる理由がある場合に、その考えを変えるためにはヒストグラムが大きく歪んでいる必要があります。データが大きいほど、ヒストグラムで正規曲線がはっきりするはずですが、その場合に予想する正規曲線から少しずれていても、あまり気にしてはいけません。

図 8.89 は、100,000 の観測値を使用して生成されたものです。大きい標本数（たとえば観測値が 1,000 や 10,000）を使って、いくつかの異なる分布をプロットしてみてください。分布の形状の違いに注意してください。

参考資料

- Box, G.E.P., Hunter, W.G., & Hunter, J.S. (1978). 『Statistics for Experimenters』. New York: Wiley.
- Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A. (1983). 『Graphical Methods for Data Analysis』. Belmont, California: Wadsworth.
- Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. 『Journal of the American Statistical Association』. 74:829–836.
- Cleveland, W.S. (1985). 『The Elements of Graphing Data』. Monterrey, California: Wadsworth.
- Fleiss, J.L. (1981). 『Statistical Methods for Rates and Proportions』 (2nd ed.). New York: Wiley.
- Friedman, J.H. (1984). 『A Variable Span Smoother』. Technical Report No. 5, Laboratory for Computational Statistics. Department of Statistics, Stanford University, California.
- Laird, N.M. & Ware, J.H. (1982). Random-Effects Models for Longitudinal Data. 『Biometrics』. 38: 963–974.
- Lindstrom, M.J. & Bates, D.M. (1990). Nonlinear Mixed Effects Models for Repeated Measures Data. 『Biometrics』. 46: 673–687.
- Snedecor, G.W. & Cochran, W.G. (1980). 『Statistical Methods』 (7th ed.). Ames, Iowa: Iowa State University Press.
- Venables, W.N. & Ripley B.D. (1999). 『Modern Applied Statistics with S-PLUS』 (3rd ed.). New York: Springer.

第9章

オブジェクトおよび データベースによる作業

はじめに	454
オブジェクトタイプとデータベース	455
S-PLUS オブジェクトタイプ	455
データベース	457
オブジェクト・エクスプローラについて	461
エクスプローラ・ページの挿入と削除	464
フォルダの挿入と削除	465
オブジェクト・エクスプローラをカスタマイズする	466
オブジェクトによる作業	472
オブジェクトを検索する	472
オブジェクトのフィルタリング	473
オブジェクトを操作する	478
作業の構成	482
プロジェクトフォルダを使用する	482
チャプターによる作業	484
チャプターを追加する	484

はじめに

S-PLUS の環境は、オブジェクト指向です。これはデータセット／グラフシート／関数からメニュー／ダイアログ／ツールバー／ツールバーボタンにいたるまで、S-PLUS のすべての要素が別々の編集可能なオブジェクトであるという意味です。データや関数などのいくつかのオブジェクトは、内部データベースに自動的に保存されます。しかし、グラフシートやスクリプトなど他のタイプのオブジェクトは、現在のセッション内にもみ存在しているので、永久的に保存するためにはディスクに保存しなければなりません。

この章では、まず初めに3タイプのS-PLUS オブジェクトについて説明し、これらのオブジェクトタイプがデータベースやファイルとどのように関連しているかを検証します。次に、オブジェクト・エクスプローラについて詳細に説明します。これは、オブジェクトを操作し、その構造を見たり意味のある構造にするための強力なツールです。そしてこの章の最後では、プロジェクト・フォルダとチャプターについて考察し、ユーザの作業を別々なプロジェクトごとに構成するために、これらのツールをどのように使用するかを説明します。

オブジェクトタイプとデータベース

S-PLUS オブジェクト タイプ

S-PLUS には、次の 3 つの基本のタイプのオブジェクトがあります。

- データフレーム／関数／リストなどの「エンジンオブジェクト (Engine objects) 」
- メニュー項目／ツールバー／ダイアログなどの「インタフェースオブジェクト (Interface objects) 」
- グラフシート／レポート／スクリプトなどの「ドキュメントオブジェクト (Document objects) 」

エンジン オブジェクト

エンジンオブジェクトは、コード実行中に S-PLUS インタープリタによって作成または使用されます。これらのオブジェクトは、S-PLUS によって内部データベースに自動的に保存されます。

ベクトル

ベクトルは、S-PLUS の最も基本的なオブジェクトです。ベクトルは、単一「型」のデータ値の 1 次元配列であり、通常は数字です。またベクトルは、文字列や論理変数などの他の型のデータとなることがあります。いくつかの統計手法ではベクトルが生成され、プログラミングにも利用されます。

行列

S-PLUS のもう 1 つの重要なオブジェクトは行列です。これは、すべての要素が同じ型であるデータ値の縦横の 2 次元配列です。ベクトルと同様に、いくつかの統計手法では行列が生成され、プログラミングにも利用されます。

データフレーム

S-PLUS において、2 次元データを保存する基本構造はデータフレームです。

注意

この『ユーザーズガイド』全体を通して、2 次元データを「データセット」と呼んでいます。しかし S-PLUS では、すべてのオブジェクトに特定のクラスがあり、これらの「データセット」と呼んでいるオブジェクトのクラスは `data.frame` です。

「クラス」と「メソッド」は、S-PLUS のプログラミングで重要な役割を果たします。これらの概念の詳細については、『Programmer's Guide』を参照してください。

第9章 オブジェクトおよびデータベースによる作業

データフレームは、異なる型の列を含むことができます。たとえば 2 列のデータフレームにおいて、1 つの列は数値データよりなり、第 2 の列は文字データよりなることができます。データフレームでは、各行が 1 つの実験データなどの単位を表します。

リスト

リストは、S-PLUS でデータを保持するための最も一般的かつ柔軟なオブジェクトです。1 つのリストは複数の「成分」の集まりです。各成分はどのようなデータオブジェクトでもよく、また成分ごとに型が異なっても構いません。たとえば、文字列のベクトルと数字の行列と他のリストの 3 つの成分を含むことも可能です。

関数

関数は、データや別の関数オブジェクトを使って解析作業を実行する S-PLUS インタープリタ型コードを含むオブジェクトです。S-PLUS の組込み関数を利用する他に、S-PLUS 言語でユーザ独自の関数を作成することができます。

その他のエンジンオブジェクト

式/名前/モデル式など、主にプログラミングのために使用されるより限定的なオブジェクトもあります。これらのオブジェクトの詳細は、『Programmer's Guide』を参照してください。

インタフェース オブジェクト

インタフェースオブジェクトは、S-PLUS アプリケーションの実行中メモリに常駐しています。このオブジェクトは起動時にロードされ、アプリケーションを閉じるときに保存されます。この保存されたファイルは、基本設定（.Prefs）フォルダにあります。

メニュー項目やツールバーなど、いくつかのインタフェースオブジェクトを使って、ユーザ・インタフェースをカスタマイズすることができます。詳細は、『Application Developer's Guide』の第 11 章「Extending the User Interface」を参照してください。

ドキュメント オブジェクト

ドキュメントオブジェクトは、S-PLUS アプリケーションの“子”ウィンドウとして表示されます。ドキュメントオブジェクトは、エンジンオブジェクトとは違い、データベースには保存されません。これらのタイプのオブジェクトを永久的に保存するときは、ファイルに保存してください。ドキュメントオブジェクトのファイルフォーマットが、特定のタイプごとに固有のものであることに注意してください。

グラフシート

グラフシート (*.sgr ファイル) は、グラフを構成するドキュメントオブジェクトです。S-PLUS がオブジェクト指向であるため、グラフシートの各要素は、それ自体編集可能なオブジェクトです。オブジェクト・エクスプローラでは、各グラフシートオブジェクトが階層で表示され、一番上の親オブジェクトがグラフシート自体です。

スクリプト

スクリプト (*.ssc ファイル) は、S-PLUS スクリプト記述コード (プログラム) で構成されるドキュメントオブジェクトです。

レポート

レポート (*.rtf または *.srp ファイル) は、出力を構成するドキュメントオブジェクトです。

オブジェクト・エクスプローラ

次の節で説明するように、オブジェクト・エクスプローラは、S-PLUS オブジェクトの操作や視覚的な構成を容易にするための便利なインターフェースです。オブジェクト・エクスプローラ (*.sbf ファイル) はそれ自体がドキュメントオブジェクトであり、ファイルに保存することができます。


データベース

前の節で説明したように、エンジンオブジェクトは S-PLUS 内部の「データベース」に保存されます。「システムデータベース」には、関数やサンプルデータセットを含めて、何千もの S-PLUS 組込みエンジンオブジェクトが含まれます。さらに、ユーザが独自のデータセットと関数を作成するときは、S-PLUS は自動的にそれらのオブジェクトを「作業データベース (working data)」と呼ばれる特別のデータベースに保存します。

サーチパス

「サーチパス (search path)」は、ユーザがオブジェクトを必要とするときに S-PLUS データベースを検索する順序で、現在参照されているすべてのデータベースを表示しています。

サーチパスでデータベースを見るには、以下の操作を行います。

1. 標準ツールバーのオブジェクト・エクスプローラボタン  をクリックして、オブジェクト・エクスプローラを開きます。(オブジェクト・エクスプローラについては、次の節で詳しく説明します。)
2. オブジェクト・エクスプローラの左パネル内の SearchPath オブジェクトアイコンをクリックします。

第9章 オブジェクトおよびデータベースによる作業

図 9.1 に示すように、オブジェクト・エクスプローラの右パネルに、参照可能なすべての S-PLUS データベースの名前（作業データの場合はフルパス名）とサーチパスの順序（Pos 列）が表示されます。

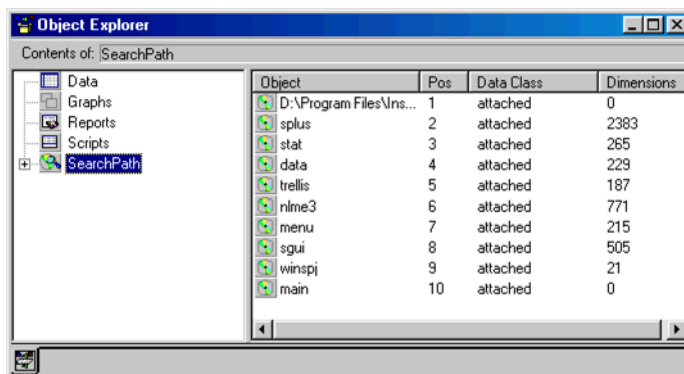


図 9.1 : オブジェクト・エクスプローラの右パネルに表示されたサーチパス

作業データ

図 9.1 で注目すべき最も重要な点は、サーチパスの位置 1(Pos 1)のデータベースです。定義上、位置 1 のデータベースは作業データベースであり、このデータベースには、ユーザが作成または変更したすべてのデータと関数オブジェクトが自動的に保存されます。

注意

サーチパスに表示されるフルパス名を参照して、その作業データベースに対応する **.Data** という名前のついた Windows フォルダを見つけることができます。

図 9.1 のサーチパスにある他のデータベースは、S-PLUS のシステムデータベースです。システムデータベースに保存されているオブジェクトを見るには、以下の操作を行います。

1. オブジェクト・エクスプローラの左パネルにある SearchPath オブジェクトを展開するには、このオブジェクトのアイコンの左側にある “+” をクリックします。

- 左パネル上でアイコンをクリックすることにより、データベース（たとえば **data**）を選択します。選択したデータベースの内容が、図 9.2 に示すように右パネルに表示されます。

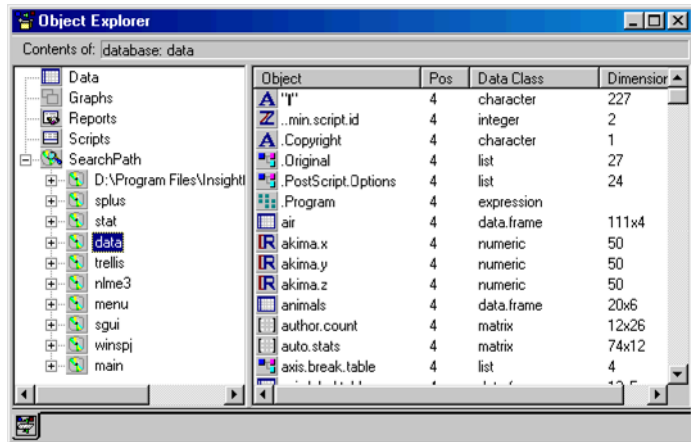


図 9.2 : 右パネルにデータベースの内容を表示する

オブジェクトを呼び出したとき（たとえば、**Select Data** ダイアログを使って）、作業データベースがサーチパスの最初のデータベースであるため、まず最初に作業データベース内でその名前のオブジェクトを探します。作業データベース内にそのオブジェクトが見つからない場合、S-PLUS は位置 2 のデータベースを探し、さらに次々とデータベースの検索を行います。これは、システムデータベースに保存されている組込みの S-PLUS オブジェクトと同じ名前のオブジェクトが作成された場合には、そのオブジェクトを削除するかまたは名前を変更しない限り、そのオブジェクトがシステムオブジェクトを“隠して (mask)” しまうことを意味しています。

オブジェクト・エクスプローラの右パネルに、データクラスとサーチパスでの位置だけでなく、オブジェクトの詳細を表示するように設定すると（468 ページを参照）、他のオブジェクトによって隠されてしまうそのサーチパス上のオブジェクトは、アイコンに赤い“X”が描かれた状態で表示されます。たとえば **air** という名前のデータセットを作成すると、このデータは自動的に作業データベースに保存され、サーチパスの位置 1 に入ります。しかし、サーチパスの位置 4 にある **data** データベースには、すでに **air** という名前の組み込みデータオブジェクトがあります。したがって、後から作成した **air** が同じ名前のシステムデータオブジェクトを隠してしまい、そのシステムデータオブジェクト **air** のアイコンは、図 9.3 に示すように赤い“X”が描かれた状態で表示されます。

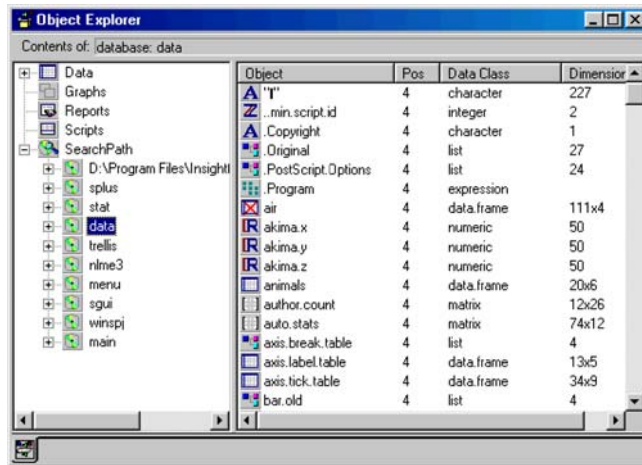


図 9.3 : 赤い“X”が描かれたアイコンはそのオブジェクトが隠されていることを示す

また、データベースに保存されている作業データ以外のオブジェクトは、ほとんどの場合、読み取り専用になっていることに注意してください。サーチパス内のデータベースに保存されているオブジェクトはどれでも変更することができますが、変更を加えたオブジェクトは作業データベースに保存され、元のオブジェクトは変更されずに元の場所に残ります。元のオブジェクトは新規作成したオブジェクトに隠されます。

オブジェクト・エクスプローラについて

オブジェクト・エクスプローラは、S-PLUS オブジェクトを操作し視覚的に構成を確認するためのシンプルかつ強力なインタフェースです。

図 9.4 の例で分かるように、オブジェクト・エクスプローラ・ウィンドウは、オブジェクトやそれらの成分および属性を種々の形で表示する 2 つのパネルに分かれています。

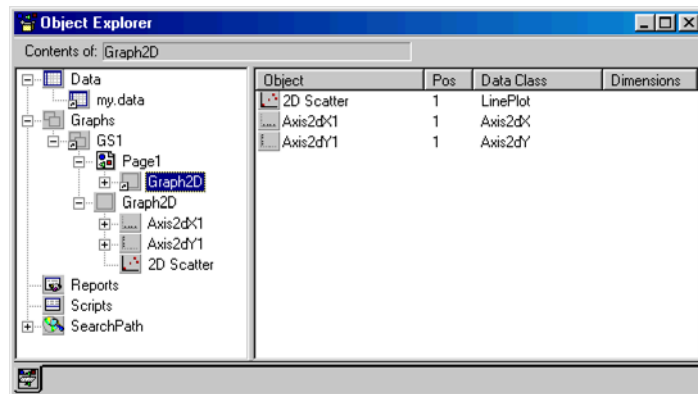



図 9.4 : オブジェクト・エクスプローラ・ウィンドウ

オブジェクト・エクスプローラの左パネルには、1 つのエクスプローラ・ページが表示されます。オブジェクト・エクスプローラは任意の数のエクスプローラ・ページを含むことができ、各ページはウィンドウ左下角のタブで表されます。それぞれのエクスプローラ・ページはまた、任意の数のフォルダを含むことができ、そのフォルダ自体には、様々なタイプのオブジェクトの参照、すなわち「ショートカット (short cuts)」を含めることができます。

注意

Windows のエクスプローラと違い、オブジェクト・エクスプローラのフォルダは実際にオブジェクトが保存されている場所を反映していません。フォルダに表示されるのはオブジェクトへのショートカットであり、これによってユーザはオブジェクトを「視覚的に」構成することができます。

オブジェクト・エクスプローラを開くには、標準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。オブジェクト・エクスプローラを閉じるときは、同じボタンを再度クリックしてください。

オブジェクト・ エクスプローラ のツールバー

オブジェクト・エクスプローラ・ウィンドウを開くと、図 9.5 に示すようなオブジェクト・エクスプローラのツールバーが自動的に現れます。ツールバーには、右パネルの表示を変更するボタンの他に、いくつかの共通のタスクを素早く実行するためのボタンがあります。

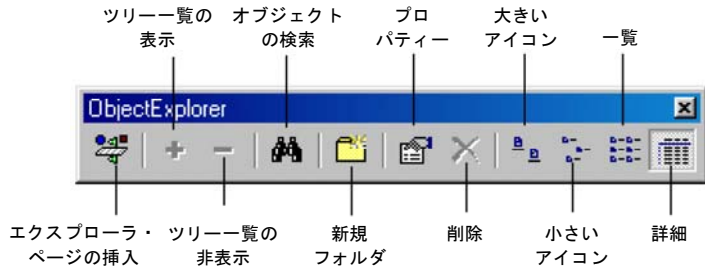


図 9.5 : オブジェクト・エクスプローラのツールバー


左パネルと 右パネルの表示

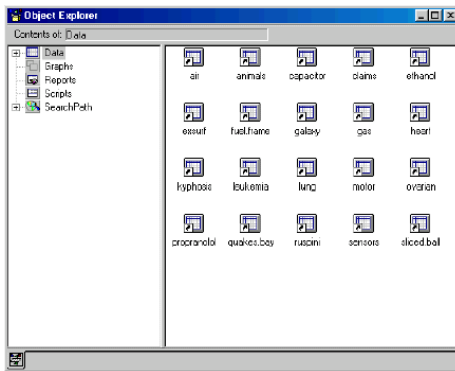
図 9.4 で分かるように、オブジェクト・エクスプローラの左パネルには、フォルダとオブジェクトが階層的に表示されます。フォルダやオブジェクトを展開することによって、その基本構造を任意の詳細レベルまで“掘り下げて”表示させることができます。

フォルダやオブジェクトを展開したり隠したりするには、以下のいずれかの操作を行います。

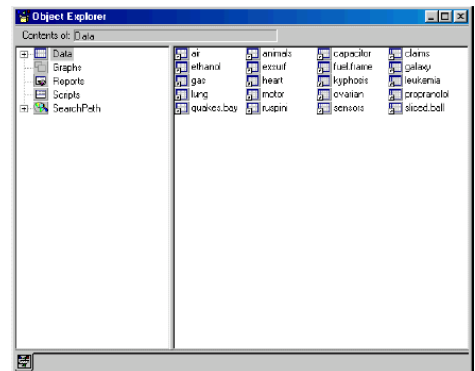
- フォルダかオブジェクトアイコンの左側にあるプラス/マイナス記号をクリックします。
- フォルダかオブジェクトを選択して、オブジェクト・エクスプローラのツールバーのツリー一覧の表示ボタン **+** またはツリー一覧の非表示ボタン **-** をクリックします。
- フォルダかオブジェクトを選択して、メインメニューから表示 ▶ 選択したものを展開、または表示 ▶ 選択のみ非表示を選択します。

オブジェクト・エクスプローラの左パネルのオブジェクトを選択すると、右パネルに選択したオブジェクトのすぐ下の階層のオブジェクトが表示されます。オブジェクト・エクスプローラのツールバーの対応するボタンをクリックすることによって、図 9.6 に示すように、右パネルの 4 つの表示方法から選択することができます。

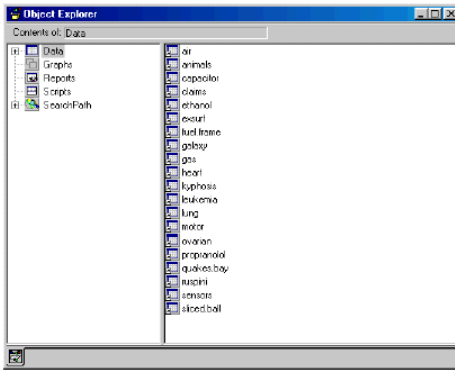
大きいアイコン 




小さいアイコン 



一覧 



詳細 

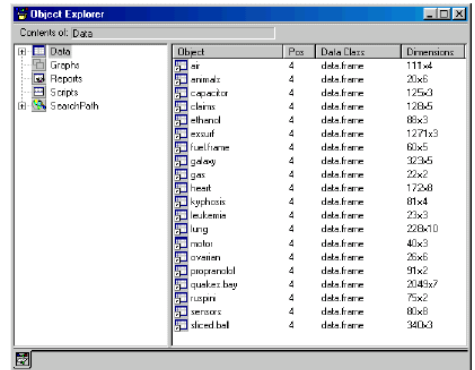


図 9.6 : オブジェクト・エクスプローラの右パネルの表示方法

右パネルの表示でオブジェクトの名前の一部が省略されている場合は、カーソルをそれぞれのパスの上に置くだけで全文より成るツールティップが表示されます。ツールティップは、オブジェクト・エクスプローラの両方のパネルとも、現在の列またはパネルの幅では完全に表示できないオブジェクトを表示します。

さらに、**オブジェクト・エクスプローラ**の右パネルにある列のサイズを変更するには、以下の操作を行います。

1. 列見出し右側の縦線の上にカーソルを置きます。マウスポインタがサイズ変更ツールになります。

Object	Pos	Data Class	Dimensions
D:\Program Files\Ins...	1	attached	0
splus	2	attached	2383
stat	3	attached	265
data	4	attached	226
trelis	5	attached	187
nime3	6	attached	771
menu	7	attached	215
sgui	8	attached	505
winspj	9	attached	21
main	10	attached	0


2. 以下のいずれかの操作を行います。
 - ダブルクリックすると、その列内の最も長いエントリの幅に合わせて、自動的に列が広がります。
 - 列の幅を広くするときは右に（幅を狭くするときは左に）ドラッグします。

右パネルに表示されている任意の情報をある列でソートするときは、その列の見出しをクリックします。

エクスプローラ・ページの挿入と削除


前述のように、**オブジェクト・エクスプローラ**に**エクスプローラ・ページ**をいくつでも追加することができます。特定のページに移動するときは、**オブジェクト・エクスプローラ**・ウィンドウの左下角のタブをクリックするだけです。

エクスプローラ・ページを挿入するには、以下のいずれかの操作を行います。

- **オブジェクト・エクスプローラ**のツールバーの**エクスプローラ・ページの挿入**ボタン  をクリックします。
- メインメニューから、**挿入 ▶ エクスプローラ・ページ**を選択します。
- **オブジェクト・エクスプローラ**のどちらかのパネルの白い部分を右クリックし、ショートカットメニューから **Create Explorer Page** を選択します。

以上のいずれかを行うと、**Explorer Page** ダイアログが開きます。このダイアログを使用することにより、469 ページで説明するように、**エクスプローラ・ページ**をフォーマットしたり、または **OK** をクリックしてデフォルト値を確定してページを挿入したりすることができます。


エクスプローラ・ページを削除するときは、まずそのタブをクリックし、次に以下のいずれかの操作を行います。

- 左パネルのオブジェクトが選択されていない状態で、**オブジェクト・エクスプローラ**のツールバーの**削除ボタン**  をクリックしてください。
- **オブジェクト・エクスプローラ**の左パネルの白い部分を右クリックし、ショートカットメニューから **Delete Explorer Page** を選択します。


フォルダの挿入と削除

エクスプローラ・ページに複数のフォルダを作り、各フォルダのフィルタリングパラメータを変更することによって、特定のオブジェクトのみで構成し表示することができます。フィルタリングはとても重要なツールなので、この章の後の節で詳しく説明します。473 ページの「オブジェクトのフィルタリング」を参照してください。

フォルダを挿入するときは、以下のいずれかの操作を行います。


- 左パネルのオブジェクトが選択されていない状態で、**オブジェクト・エクスプローラ**のツールバーの**新規フォルダボタン**  をクリックするか、メインメニューから**挿入 ▶ フォルダ**を選択します。
- **オブジェクト・エクスプローラ**の左パネルの白い部分を右クリックし、ショートカットメニューから **Insert Folder** を選択します。

既存のフォルダにフォルダを挿入するときは、以下のいずれかの操作を行います。

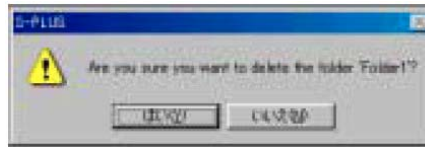
- 新しいフォルダを挿入したいフォルダを選択し、次に**オブジェクト・エクスプローラ**のツールバーの**新規フォルダボタン**  をクリックするか、メインメニューから**挿入 ▶ フォルダ**を選択します。
- 新しいフォルダを挿入したいフォルダのアイコンを右クリックし、ショートカットメニューから **Insert Folder** を選択します。

以上のいずれかの操作を行うと、デフォルト名 **Folder x** (x は連続番号) の新しいフォルダが挿入されます。

フォルダを削除するときは、以下のいずれかの操作を行います。

- フォルダを選択し **DELETE** を押します。
- フォルダを選択し、**オブジェクト・エクスプローラ**のツールバーの**削除ボタン**  をクリックします。
- フォルダのアイコンを右クリックし、ショートカットメニューから **Delete** を選択します。

削除してよいかを確認するダイアログが表示されます。**はい**をクリックするとフォルダが削除されます。



オブジェクト・エクスプローラをカスタマイズする

オブジェクト・エクスプローラは、完全にカスタマイズ可能なインタフェースです。設定を行うには、以下のいずれかの操作を行って **Object Explorer** ダイアログを開いてください。

- メインメニューから、**書式** ▶ **オブジェクト・エクスプローラ**を選択します。
- **オブジェクト・エクスプローラ**の右パネルの白い部分をダブルクリックします。
- **オブジェクト・エクスプローラ**の右パネルの白い部分を右クリックし、ショートカットメニューから **Explorer** を選択します。

Explorer ページ

以上のいずれかの操作を行うと、図 9.7 に示すように、**Explorer** ページが表示された状態で **Object Explorer** ダイアログが開きます。

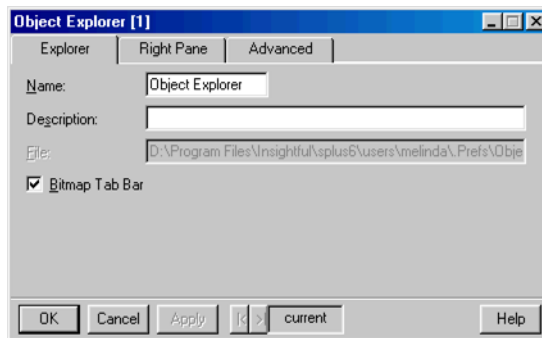


図 9.7 : Object Explorer ダイアログの Explorer ページ

Name オブジェクト・エクスプローラの名前は、他のドキュメントオブジェクトと同じように、それ自体が保存されるファイルの名前です。たとえば、この **Object Explorer** という名前を変更してファイルに保存したい場合は、このフィールドに新しい名前を入力します。

注意

標準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すると、S-PLUS は、S-PLUS プロジェクトフォルダの.Prefs フォルダ内で、**Object Explorer.sbf** という名前のファイルを探します。したがって、通常はこのツールバーのボタンをクリックして開いたオブジェクト・エクスプローラの名前を変更しないようお勧めします。

Description 必要に応じて、テキストボックスにオブジェクト・エクスプローラに関する詳しい情報をを入力します。

File オブジェクト・エクスプローラがファイルに保存されると、ここにファイルのパス名が表示されます。

Bitmap Tab Bar このオプションを選択すると、オブジェクト・エクスプローラ・ウィンドウ左下角のそれぞれのタブにエクスプローラ・ページのビットマップ画像が表示されます。その代わりにタブにエクスプローラ・ページの名前を付けるときは、チェックボックスのチェックを外してください。(エクスプローラ・ページのビットマップと名前の指定方法については、470 ページを参照してください。)

Right Pane ページ

図 9.8 に **Object Explorer** ダイアログの **Right Pane** ページを示します。

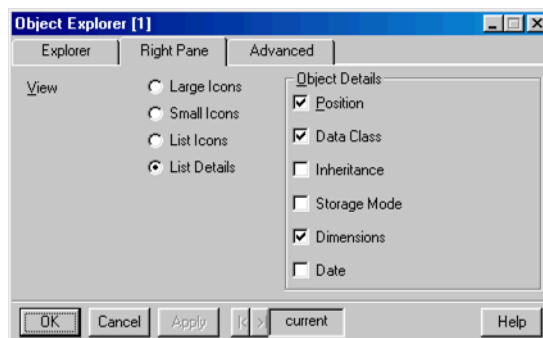


図 9.8 : Object Explorer ダイアログの Right Pane ページ

View ラジオボタンの1つをクリックして希望の右パネル表示を選択します (図 9.6 を参照)。**オブジェクト・エクスプローラ**のツールバーボタンをクリックするか、メインメニューから**表示**を選択して希望の表示を選択しても、右パネルの表示を変更することができます。



Right Pane ページ、Object Details グループ

右パネル表示に **List Details** を選択すると、**Object Details** グループが使用できるようになります。ここで、表示したい詳細情報の種類を選択することができます。チェックボックスを選択すると、右パネルにその情報の列が表示されます。

次に、各オプションについて簡単に説明します。詳細は、『Programmer's Guide』を参照してください。

Position このオプションは、リストにあるオブジェクトの型に依存します。データ、関数などのエンジンオブジェクト (内部オブジェクト) の場合、**Pos** はオブジェクトが保存されているデータベースのサーチパス番号を示します。リストやデータフレームの要素の場合、**Pos** は親オブジェクト内でのそれらの位置を示します。ツールバーボタンとメニュー項目の場合、**Pos** はツールバーとメニュー上のそれらの相対的な位置をそれぞれ示します。

Data Class `data.frame`、`design` または `lm` などのデータオブジェクトのクラスです。

Inheritance オブジェクトのデータクラスと継承する任意のクラスのことです。

Storage Mode データオブジェクトのモードです。

Dimensions オブジェクトの次元を表し、ベクトルならば長さ、データフレームや行列ならば行数と列数です。

Date オブジェクトが最後に修正された日付です。

Advanced ページ

Object Explorer ダイアログの **Advanced** ページを図 9.9 に示します。

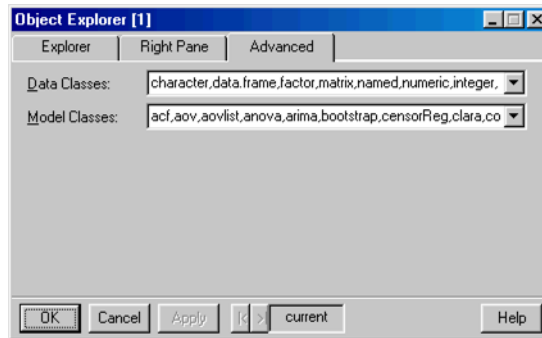


図 9.9 : Object Explorer ダイアログの Advanced ページ

Data Classes エンジンオブジェクト（内部オブジェクト）のどのクラスを表示させるデータオブジェクトとして定義したいかを指定します。

Model Classes エンジンオブジェクト（内部オブジェクト）のどのクラスを表示させるモデルオブジェクトとして定義したいかを指定します。

データクラスとモデルクラスの両方にデフォルトの定義がありますが、ユーザの必要に合わせて、これらの定義を修正することができます。データベースをフィルタリングするとき、**S-PLUS** はこれらの定義によってフォルダ内に表示するオブジェクトを決定します。

重要なフォルダのフィルタリング機能の詳細は、473 ページの「オブジェクトのフィルタリング」を参照してください。

エクスプローラ・ページをフォーマットする

Explorer Page ダイアログを使って、エクスプローラ・ページをフォーマットすることができます。新しいエクスプローラ・ページを挿入すると、ダイアログが自動的に表示されます（464 ページを参照）。既存の **Explorer Page** ダイアログを開くときは、まずタブをクリックし、次に以下のいずれかの操作を行います。

- メインメニューから、**書式 ▶ エクスプローラ・ページ** を選択します。
- **オブジェクト・エクスプローラ** の左パネルの白い部分をダブルクリックします。
- **オブジェクト・エクスプローラ** の左パネルの白い部分を右クリックし、ショートカットメニューから **Properties** を選択します。

以上のいずれかの操作を行うと、図 9.10 に示すように、**Explorer Page** ダイアログが開きます。

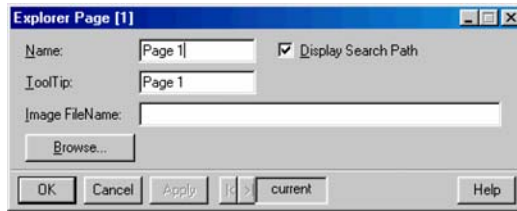


図 9.10 : Explorer Page ダイアログ


Name デフォルトでは、新しいエクスプローラ・ページは Page x (x は連続番号) という名前がつけられています。希望により、このフィールドに新しい名前を入力することができます。オブジェクト・エクスプローラが、ビットマップやアイコンではなく名前を表示するように設定されている場合は、その名前が **Explorer Page** タブに表示されます。詳細は、468 ページを参照してください。

ToolTip マウスのカーソルを **Explorer Page** タブの上に置くと、ツールティップが表示されます。デフォルトでは、ツールティップに表示されるテキストはそのページの名前ですが、希望により様々なツールティップを指定することができます。

Image FileName オブジェクト・エクスプローラが、**Explorer Page** タブにビットマップまたはアイコンを表示するように設定されている場合は (468 ページを参照)、使用したいビットマップファイルまたはアイコンのファイル名を絶対パスで入力してください。ファイルを参照するときは、**Browse** ボタンをクリックします。ファイルを指定しない場合は、デフォルトの画像が使用されます。

Display Search Path このオプションを選択すると、SearchPath オブジェクトが自動的にエクスプローラ・ページに挿入されます。

希望のデフォルトを設定する

標準ツールバーのオブジェクト・エクスプローラボタン  をクリックすると、「デフォルト」のオブジェクト・エクスプローラが開きます。オブジェクト・エクスプローラをカスタマイズした後、以下のいずれかの操作を行うことによって、変更を新しいデフォルトの設定として保存することができます。

- オブジェクト・エクスプローラを表示し、どちらのパネルでもオブジェクトが選択されていない状態で、メインメニューから、**オプション ▶ ウィンドウサイズ・プロパティの現在の状態をデフォルトにする**を選択します。
- オブジェクト・エクスプローラの右パネルの白い部分を右クリックし、シ

ショートカットメニューから **Save Object Explorer as default** を選択します。

- オブジェクト・エクスプローラの右パネルの白い部分を右クリックし、ショートカットメニューから **Save** を選択します。
- メインメニューから、**ファイル ▶ 保存** を選択します。
- オブジェクト・エクスプローラ・ウィンドウの右上隅の閉じるボタン **✕** をクリックします。そのオブジェクト・エクスプローラをファイルに保存するように促すダイアログが表示されたら、**Yes** をクリックします。

ユーザが作成した新しいオブジェクト・エクスプローラをカスタマイズし、それを新しいデフォルトの設定として保存することもできます。上記の 1 番目または 2 番目の方法を使用してください。

注意

オブジェクト・エクスプローラはドキュメントオブジェクトであり、ドキュメントオブジェクトの名前は、そのオブジェクトが保存されているファイルの名前です。S-PLUS は、デフォルトで使用する **Object Explorer.sbf** という名前のファイル (S-PLUS プロジェクトフォルダの **.Prefs** フォルダにある) を探すので、オブジェクト・エクスプローラの名前を変更し、その名前を変更したものをデフォルトとして使用することはできません。

起動時に オブジェクト・ エクスプローラ を開く

起動時にデフォルトのオブジェクト・エクスプローラを自動的に開きたい場合は、以下の操作を行います。

1. メインメニューから、**オプション ▶ 設定** を選択します。
2. **General Settings** ダイアログの **Startup** タブをクリックします。
3. **Open at Startup** グループで、**Object Explorer** チェックボックスを選択し **OK** をクリックします。


オブジェクトによる作業

この章のはじめに述べたように、S-PLUS ではすべてのものがオブジェクトです。オブジェクト・エクスプローラは、その名の示す通り S-PLUS オブジェクトの構造を探索する手段を与えてくれます。またオブジェクト・エクスプローラは、データベースに保存されているオブジェクトを検索したり、エクスプローラ・ページフォルダのフィルタリング機能を使用してオブジェクトを選択したりするための便利なツールでもあります。さらに、オブジェクト・エクスプローラを使用すると、オブジェクトおよびオブジェクトショートカットを作成／選択／表示／編集／コピー／移動／削除することができます。


オブジェクトを検索する

Find Objects ダイアログは、現在のサーチパスにある任意のデータベースに保存されたオブジェクトを検索できる強力な探索ツールです。

オブジェクトを検索して現在のフォルダ上に表示させるには、以下のいずれかの操作を行います。

- フォルダのアイコンをクリックして、次にオブジェクト・エクスプローラのツールバーのオブジェクトの検索ボタン  をクリックするか、またはメインメニューから編集 ▶ S-PLUS オブジェクトの検索を選択します。
- フォルダのアイコンを右クリックして、ショートカットメニューから **Find** を選択します。

オブジェクトを検索して **Found Objects** という名前の新しいフォルダに入れるには、以下のいずれかの操作を行います。

- 左パネルのオブジェクトが選択されていない状態で、オブジェクト・エクスプローラのツールバーのオブジェクトの検索ボタン  をクリックするか、またはメインメニューから編集 ▶ S-PLUS オブジェクトの検索を選択します。
- オブジェクト・エクスプローラの左パネルの白い部分を右クリックして、ショートカットメニューから **Find** を選択します。

Find Objects ダイアログの **Pattern** フィールド (図 9.11 に示す) は、パターンを入力として取り込みます。ワイルドカードを使用することができ、正規表現も使用することができます。前の検索のパターンが保存されており、ドロップダウンリストから選択することができます。

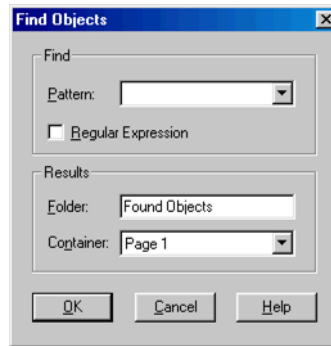


図 9.11 : Find Objects ダイアログ

サーチパスの追加されたデータベースがすべて検索され、**Folder** フィールドに指定されたフォルダに、条件に合うすべてのオブジェクトへのショートカットが入ります。**Container** フィールドには、検索結果のフォルダを含むエクスプローラ・ページの名前が反映されます。


注意

オブジェクトの検索機能は、インタフェースオブジェクトと追加されたデータベース内のオブジェクトだけを検索します。オブジェクトが保存されているデータベースがサーチパスにないと、そのオブジェクトは検索されません。

オブジェクト のフィルタ リング

独自の「フィルタリング」機能を使うことにより、フォルダに表示させるオブジェクトのタイプを特定のものに制限することができます。さらに、フォルダを使ってオブジェクトを管理することにより、データ解析や統計モデル作成作業などが容易になります。

フォルダのフィルタリングプロパティを設定するには、以下の操作のいずれかを行います。

- フォルダのアイコンをクリックしてフォルダを選択し、次に**オブジェクト・エクスプローラ**のツールバーの**プロパティ**ボタン  をクリックします。
- フォルダのアイコンをクリックしてフォルダを選択し、次にメインメニューから、**書式 ▶ 選択された Folder** を選択します。
- フォルダのアイコンを右クリックし、ショートカットメニューから**Folder** を選択します。

Folder ページ

以上のいずれかの操作を行うと、図 9.12 に示すように、**Folder** ページが表示された状態で **Folder** ダイアログが開きます。この **Folder** ページで、フィルタリングのきわめて概略的なレベルを設定することができます。

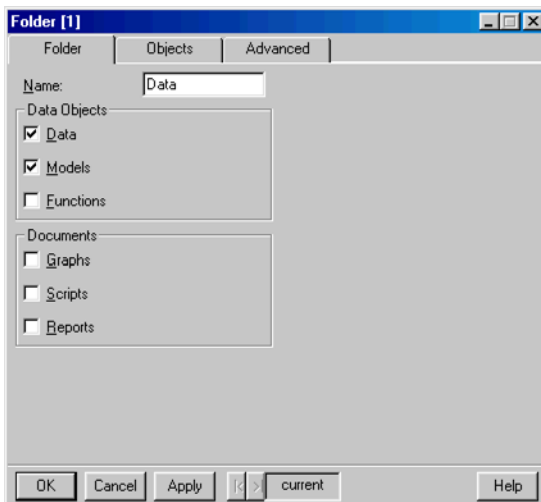


図 9.12 : Folder ダイアログの Folder ページ

Name フォルダの名前。フォルダの名前を変更するときは、このフィールドに新しい名前を入力します。

Folder ページ、Data Objects グループ

Data Objects グループは、このダイアログの **Advanced** ページ（476 ページを参照）で選択したデータベースに保存されたエンジンオブジェクト（内部オブジェクト）をフィルタリングするためのものです。

Data このオプションを選択すると、フォルダはデータオブジェクトをフィルタリングします。対象となるデータオブジェクトのほとんどは、データフレーム（`data.frame` タイプのオブジェクト）であり、行列（`matrix`）とベクトル（`vector`）の場合もあります。一般にデータオブジェクトは、そのようなオブジェクトのクラスのうちの 1 つか、あるいはそこから派生したものです。

Models このオプションを選択すると、フォルダはモデルオブジェクトをフィルタリングします。対象となるモデルオブジェクトのほとんどは、リストと構造（それぞれ、`list` と `structure` クラスのオブジェクト）から派生したものです。

注意

Object Explorer ダイアログの **Advanced** ページを使って、データとモデルオブジェクトとして定義したいオブジェクトを明示的に指定することができます。詳細は、469 ページを参照してください。

Functions このオプションが選択されると、フォルダは関数オブジェクトをフィルタリングします。フィルタリングするデータベースとして作業データを選択することにより、S-PLUS 言語で記述したユーザ定義関数、または組み込み関数を修正することによって作成した関数を表示させることができます。

Folder ページ、Documents グループ

Documents グループは、セッションで現在開いているグラフ/スクリプト/レポートを対象としてフィルタリングするためのものです（ドキュメントオブジェクトはデータベースには保存されていません）。必要に応じて、これらのチェックボックスのどれか、またはすべてを選択してください。

Objects ページ

図 9.13 に示す **Folder** ダイアログの **Objects** ページは、フォルダのフィルタリングプロパティの例外的なオブジェクトを表示します。

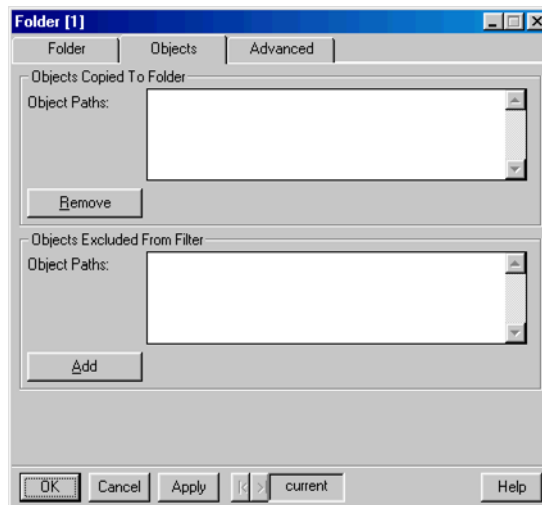


図 9.13 : Folder ダイアログの Objects ページ

Objects ページ、Objects Copied To Folder グループ

Object Paths このテキストボックスは、フォルダのフィルタリングプロパティ（検索条件）と一致しないけれども、このフォルダ内にショートカットが作成されているオブジェクトを表示します。フォルダからオブジェクトショートカットを削除するときは、それを選択して **Remove** をクリックします。

注意

Remove をクリックすると、フォルダからオブジェクトのショートカットだけが削除され、オブジェクト自体は削除されません。

Objects ページ、Objects Excluded From Filter グループ

Object Paths このテキストボックスは、フォルダのフィルタリングプロパティ（検索条件）と一致するけれども、ショートカットが削除されているオブジェクトを表示します。フォルダにオブジェクトショートカットを追加するときは、それを選択して **Add** をクリックします。

Advanced ページ

図 9.14 に示す **Folder** ダイアログの **Advanced** ページを使用すると、フィルタリングの基準を調整することができます。

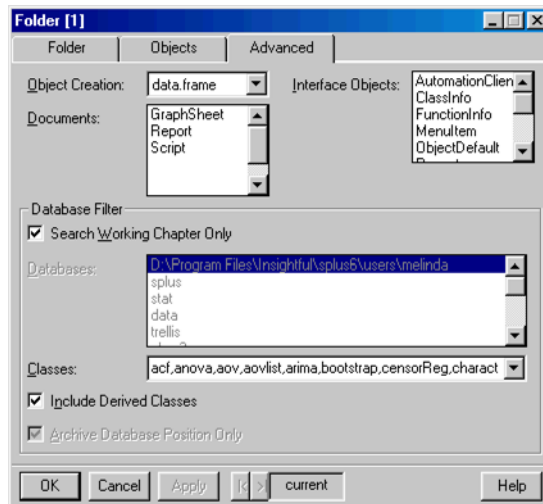


図 9.14 : Folder ダイアログの Advanced ページ

Object Creation フォルダに作成されるオブジェクトのデフォルトのクラスを選択します。ここでデフォルトのクラスを選択しても、フォルダ内にそのクラスのオブジェクトだけが作成されるよう制限しているわけではないことに注意してください。フォルダのショートカットメニューに便利なメニュー選択項目が提供されるだけです。フィルタリングしたクラスと一致しないオブジェクトを作成すると、ダイアログの **Objects** ページの **Objects Copied To Folder** ボックスにその参照が現れます。

Documents フォルダに表示させたいドキュメントオブジェクトのクラスを選択します。

Interface Objects フォルダに表示させたいインタフェースオブジェクトのクラスを選択します。

Advanced ページ、Database Filter グループ

Search Working Chapter Only これが選択された場合、S-PLUS は、作業データだけをフィルタリングし、**Search Path** 内の他のデータベースはフィルタリングしません。他のデータベースをフィルタリングするときは、まずこのチェックボックスのチェックを外し、次に **Databases** フィールドでフィルタリングするデータベースを選択してください。

Databases 現在サーチパスにあるデータベースの中から探索したいデータベースを選択します。そのデータベース内にあるオブジェクトだけがフォルダ内に表示されます。

Classes フォルダに表示するオブジェクトのクラスを選択します。このフィールドには、選択したクラスがカンマで区切ったリストとして現れます。**(Folder ページの Data Objects グループで選択を行うと、このフィールドの内容が変更されます。)** すべてのオブジェクトを含めるときは、専用キーワード **(All)** を選択してください。

Include Derived Classes このオプションを選択すると、**Classes** フィールドで指定されたクラスから派生したオブジェクトも表示されます。たとえば **design** オブジェクトは **data.frame** オブジェクトから派生しています。このオプションが選択されていて、フォルダも **data.frame** オブジェクトをフィルタリングするように設定されている場合は、**design** オブジェクトも表示されます。

Archive Database Position Only データベースのパスがマシン固有の場合があるので、**オブジェクト・エクスプローラ・ファイル**（ドキュメントファイル）を共有したいときは、このオプションを選択してください。このオプションが選択されていない場合、フォルダがフィルタリングしているデータベースのパスが保存されます。この場合、データベースのパスがファイルから読み取られ、パスが現在のサーチパスにないとき、ユーザはデータベースを追加するように要求されます。

希望のデフォルトを設定する

フォルダのフィルタリングプロパティをカスタマイズした後で、以下のいずれかの操作を行うことによって変更を新しいデフォルトの設定として保存できます。

- フォルダのアイコンをクリックして選択し、メインメニューから**オプション ▶ デフォルトの Folder** として**保存**を選択します。
- フォルダのアイコンを右クリックし、ショートカットメニューから**Save Folder as default** を選択します。

オブジェクトを操作する

オブジェクト・エクスプローラは、いくつかのタイプのオブジェクトを操作するのに最も便利な方法を提供します。オブジェクト・エクスプローラを使って、オブジェクトおよびオブジェクトショートカットを作成／選択／表示／修正／コピー／移動／削除することができます。

オブジェクトを作成する

エクスプローラ・ページ内にある各フォルダのショートカットメニューは、オブジェクトやその他のフォルダを作成するためのオプションを提供します。フォルダ内に作成するオブジェクトのデフォルトクラスを指定した場合は、477 ページで説明したように、そのオブジェクトタイプを作成するための追加のメニュー選択項目が現われます。オブジェクトを選択すると、そのオブジェクトのショートカットがフォルダ上に表示され、オブジェクト自体は常に作業データ内に保存されます。

オブジェクトを作成するには、以下の操作を行います。

1. フォルダのアイコンを右クリックして、ショートカットメニューから**Insert** を選択します。図 9.15 に示すように、**Create Object** ダイアログが開きます。

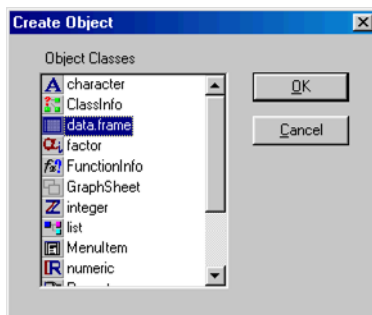


図 9.15 : Create Object ダイアログ

2. 作成したいオブジェクトのタイプを選択し、**OK** をクリックします。

注意

作業データベースが、サーチパスの最初のデータベースであるため、S-PLUS のシステムオブジェクトと同じ名前のオブジェクトを作成すると、そのオブジェクトがシステムオブジェクトを“隠して (mask)” します。そのサーチパスにあり、同じ名前の別のオブジェクトによって隠されたシステムオブジェクトは、アイコンに赤い“X”が描かれた状態で表示されます。重複するのを避けるため、オブジェクトの名前を変更してください。

**オブジェクトの
選択**

オブジェクト・エクスプローラ内のオブジェクトを選択するときは、そのアイコンをクリックしてください。オブジェクト・エクスプローラは、グラフシートを重ね合わされたグラフ要素など、他の画面で選択できないオブジェクトを選択するのに特に役立ちます。


オブジェクト・エクスプローラの右パネルのグラフ要素オブジェクトが選択されている場合、それらはそのオブジェクトがあるグラフシート内でも選択されています。同様にデータ・ウィンドウにデータセットが開かれている場合は、右パネルの列を選択するとデータ・ウィンドウ内の列も選択されます。右パネル内の列を選択することにより、プロットパレットを使用してそれをグラフにすることができます。

**オブジェクトの
表示と修正**

オブジェクト・エクスプローラにおいて、異なるクラスのオブジェクトをダブルクリックすると、その動作はクラスによりそれぞれ異なります。たとえば、クラス `data.frame` や `matrix` のオブジェクトをダブルクリックすると、データ・ウィンドウが表示されますが、クラス `lm` のオブジェクト (`lm` 関数で作成されたオブジェクト) をダブルクリックすると、レポート・ウィンドウ内のオブジェクトのサマリーが表示されます。

注意

データウィンドウで編集する場合は、クラス `numeric`、`integer`、`character`、`logical`、`single`、`complex`、`factor`、および `timeDate` の特定の列オブジェクトしか開くことができません。データウィンドウで他の特定クラスのオブジェクトを開こうとすると、オブジェクトが読み専用 (`read only`) モードで開かれるという警告が現れます。(読み専用モードでは、データウィンドウ内の列は少し暗くなります。)

そのようなオブジェクトを編集可能なデータフレームに変換するには、まずそのオブジェクトを読み専用モードで開き、次にデータウィンドウのツールバーにある **Convert to Data Frame** ボタン  をクリックするか、メインメニューから **データ ▶ 列形式の変換** を選択します。

データフレームや行列などのオブジェクトの**データ**・ウィンドウも、そのオブジェクトを右クリックしてショートカットメニューから **Edit** を選択することによってウィンドウに表示させることができます。他のタイプのオブジェクトを修正するには、そのオブジェクトのプロパティダイアログを使用します。プロパティダイアログを開くときは、オブジェクトを右クリックし、ショートカットメニューから **Properties** を選択してください。複数のページを含むプロパティダイアログの場合は、ショートカットメニューからページの名前を選択します。**オブジェクト・エクスプローラ**で行われた変更は、オブジェクトに直ちに反映されます。

オブジェクトの移動とコピー

CTRL-C と CTRL-V を使用するか、**編集**メニューまたはショートカットメニューの切り取り (Cut)、コピー (Copy)、および貼り付け (Paste) コマンドを選択することによって、オブジェクトのショートカットをコピー/移動することができます。またフォルダ間でオブジェクトをドラッグ アンド ドロップすることもできます。

ヒント

オブジェクト・エクスプローラ・ウィンドウ内でオブジェクトをドラッグ アンド ドロップすると、オブジェクトが移動します。オブジェクトをコピーするときは、CTRL を押しながらドラッグしてください。


一般に、ほとんどすべての S-PLUS オブジェクトを他のオブジェクトにドラッグ アンド ドロップして結果を得ることができます。たとえば、

- データオブジェクトをグラフオブジェクトにドロップすると、グラフ表示されます。
- 任意のオブジェクトをスクリプト・ウィンドウにドロップすると、実行するとそのオブジェクトを再現するスクリプトが作成されます。
- オブジェクトをレポート・ウィンドウにドロップすると、オブジェクトのサマリーが生成されます。
- フォルダまたは**エクスプローラ**・ページをツールバーにドロップすると、それぞれフォルダまたは**エクスプローラ**・ページを開くツールボタンが作成されます。

オブジェクトとオブジェクトショートカットを削除する

オブジェクト・エクスプローラによるオブジェクトの削除は、多数のオブジェクトを削除するときに特に便利です。たとえばプロット上の矢印をすべて削除したい場合は、その矢印をグラフ上で直接選択するよりも**オブジェクト・エクスプローラ**で選択したほうがはるかに簡単です。また、データフレームから列を簡単に削除することができます。右パネルの削除したい列を選択し、DELETE を押すだけです。

作業データベースからオブジェクトを削除するには、オブジェクト・エクスプローラから以下のいずれかの操作を行います。

- オブジェクトを選択し、**DELETE** を押します。
- オブジェクトを選択し、**オブジェクト・エクスプローラ**のツールバーの**削除**ボタン  をクリックします。
- オブジェクトを選択し、メインメニューから**編集 ▶ クリア**を選択します。
- オブジェクトを右クリックし、ショートカットメニューから **Delete** を選択します。

オブジェクトショートカットを削除するときは、以下のいずれかの操作を行います。

- オブジェクトを選択し、**CTRL-DELETE** を押します。
- オブジェクトを選択し、メインメニューから**編集 ▶ ショートカットの削除**を選択します。
- オブジェクトを右クリックし、ショートカットメニューから **Delete Short Cut** を選択します。

注意

作業データからオブジェクトを削除することができますが、システムデータベースに保存されたオブジェクトを削除することはできません。システムオブジェクトでは、ショートカットメニューで **Delete Short Cut** しか選択することができません。

作業の構成

いくつかの異なるプロジェクトを同時に処理している場合には、データと結果を各プロジェクトごとに別々に保存しておくのが便利だと思われるでしょう。S-PLUS のプロジェクトフォルダとチャプターは、そのための方法を提供し、プロジェクトの管理を容易にします。

プロジェクトフォルダを使用する

S-PLUS を起動するたびに、図 9.16 に示すダイアログを表示するように指定することができます。

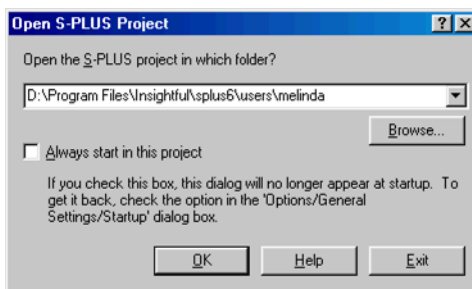


図 9.16 : Open S-PLUS Project ダイアログ

この動作をデフォルトに設定すると、プログラムを起動するたびに、始めようとしているセッションで使いたい「プロジェクトフォルダ」を指定するよう要求されます。S-PLUS プロジェクトフォルダは、セッション中に作成し変更したデータやドキュメントを保存するための中心となる Windows フォルダです。

Windows フォルダを S-PLUS プロジェクトフォルダとして指定すると、以下に示す 2 つの重要なサブフォルダが S-PLUS プロジェクトフォルダ内に自動的に作成されます。

- 特定のプロジェクトの作業データベースに対応する **.Data** フォルダ。
- S-PLUS がユーザの基本設定（例えば **General Settings** ダイアログで行った選択）およびカスタマイズの結果（例えば新しくデフォルトの **オブジェクト・エクスプローラ** を保存したとき）を保存しておく **.Prefs** フォルダ。

さらにプロジェクトフォルダは、永久的に保存するときはユーザが手動でファイルに保存しなければならない **グラフシート** / **レポート** / **スクリプト** などのドキュメントオブジェクトを保存するためのデフォルトフォルダになります。各プロジェクトフォルダにはそれぞれ独自の **.Data** および **.Prefs** フォルダ

があり、プロジェクトフォルダが、ドキュメントオブジェクトを保存するためのデフォルトフォルダになっているため、**S-PLUS** で行う作業を簡単に管理することができます。複数のプロジェクトフォルダを状況に応じて使用することにより、各プロジェクトのデータやドキュメントを独立させたまま、行っている作業をまた別なプロジェクトに構築することができます。

プロジェクト フォルダを 指定する

プロジェクト・フォルダの指定を選択すると（第 13 章、**S-PLUS** セッションのカスタマイズを参照）、プログラムを起動するたびに **Open S-PLUS Project** ダイアログが現れ、始めようとするセッションで使用したいプロジェクトフォルダを指定することができます。

注意

起動するたびに同じプロジェクトフォルダを使用したい場合は、このダイアログの **Always start in this project** チェックボックスを選択することによって、ダイアログ表示をオフにすることができます。**S-PLUS** 内でダイアログ表示をオフにするには、メインメニューから**オプション ▶ 設定**を選択し、**Startup** タブをクリックし、**Prompt for project folder** チェックボックスの選択を解除します。

プロジェクトフォルダを指定するには、以下のいずれかの操作を行い、表示されるダイアログで **OK** をクリックします。

- デフォルトのプロジェクトフォルダを確定します。**S-PLUS** を一番最初に起動したときは、**S-PLUS** プログラムフォルダ内の **users** フォルダにあるシステムデフォルトです。その後は、前のセッションで使用したプロジェクトフォルダがデフォルトになります。
- テキストボックスにそのフォルダのパス名を入力するか、または **Browse** をクリックしてそのフォルダを指定することにより、既存のプロジェクトフォルダを指定します。
- テキストボックスにそのフォルダのパス名を入力することにより、新しいプロジェクトフォルダを作成します。

注意

S-PLUS プロジェクトフォルダとして使用するフォルダには、**.Data** フォルダと **.Prefs** フォルダが含まれていなければなりません。**Open S-PLUS Project** ダイアログを使用して新しいプロジェクトフォルダを作成したときは、これらのフォルダが自動的に作成されます。

チャプターによる作業

S-PLUS では、データベースは「チャプター」と関連付けられています。各「チャプターフォルダ」は、それぞれデータベースオブジェクトを保持する **.Data** フォルダを含んでいます。特定のプロジェクトフォルダに関連付けられた作業データベース (**.Data** フォルダ) の他に、セッション中にアクセスしたい他のデータベースがあるかも知れません。データベースに含まれているオブジェクトにアクセスするときは、チャプターを追加してデータベースを追加するだけです。

チャプターを追加する

チャプターを追加するとき、または新しいチャプターを同時に作成し追加するときは、以下のいずれかの操作を行って、**Attach/Create Chapter** ダイアログを開きます。

- オブジェクト・エクスプローラの左パネルで、SearchPath オブジェクトのアイコンまたはデータベースアイコンを右クリックし、ショートカットメニューから **Attach/Create Chapter** を選択します。
- メインメニューから、ファイル ▶ チャプター ▶ チャプターのアタッチを選択します。

図 9.17 に示すように、**Attach/Create Chapter** ダイアログが開きます。

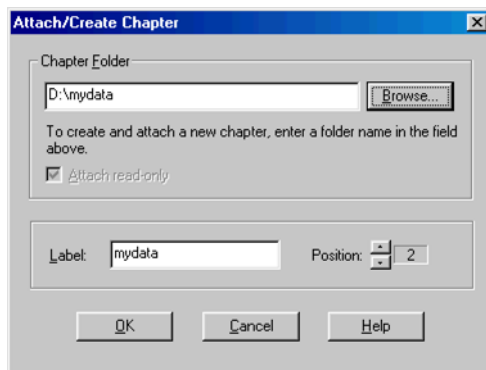



図 9.17 : Attach/Create Chapter ダイアログ

1. **チャプターフォルダ**テキストボックスで、以下の操作を行います。
 - 既存のチャプターを追加するには、必要な **.Data** フォルダを含むチャプターフォルダのパス名を入力するか、または **Browse** をクリックしてそのフォルダを指定します。
 - 新しいチャプター（新しいチャプターフォルダ内の新しい **.Data** フォルダ）を作成して追加するには、新しいチャプターフォルダのパス名を入力します。表示される確認ダイアログで **OK** をクリックします。

2. チャプターを読み取り専用として追加するときは、**Attach read-only** チェックボックスを選択します。
3. **Label** テキストボックスで、チャプターを識別するのに使用するラベルを指定します。この名前が、**サーチパス**の下の**オブジェクト・エクスプローラ**の左パネルに現れます。
4. **Position** フィールドの  ボタンをクリックして、チャプターを追加するサーチパスの位置を選択します。そのチャプターを作業データとして使用するには、**Position** を **1** に設定します。現在、**Search Path** にあるデータベースの **Position** が 1 つずつ下がります。

重要な注意

サーチパスの位置 1 に追加されたチャプターは、読書き可能なチャプターでなければなりません。一般には、このデータベースは現在のプロジェクトフォルダの作業データベースです。

現在の仕様では、**S-PLUS** は、後でプログラムを再起動するときに以前のセッションで追加されたデータベースを覚えていません。特定のプロジェクトフォルダの作業データベースと **S-PLUS** システムデータベースだけが、サーチパス内に復元されます。(特定のプロジェクトのためのサーチパスを確立するには、**.First** 関数を作成します。詳細については、624 ページの「起動時と終了時のセッションのカスタマイズ」を参照してください。)

5. **OK** をクリックします。

チャプターを終了する

特定のチャプターによる作業が終了したら、以下のいずれかの操作を行い、そのチャプターを終了します。

- **オブジェクト・エクスプローラ**の右パネルで、終了したいチャプターのアイコンを右クリックし、ショートカットメニューから **Detach Database** を選択します。表示される確認ダイアログでは **はい** をクリックします。

第9章 オブジェクトおよびデータベースによる作業

- メインメニューから、ファイル ▶ チャプター ▶ チャプターのデータタッチを選択します。**Detach Chapter** ダイアログ (図 9.18 を参照) で、分離したいデータベースを選択して **OK** をクリックします。

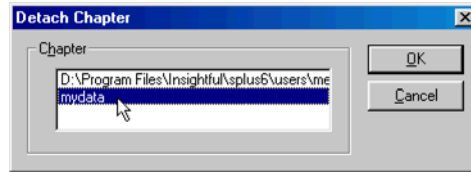


図 9.18 : Detach Chapter ダイアログ

注意

システムデータベースはどれも分離することはできません。

新しい作業 チャプターを 選択する

すでに実行中の S-PLUS セッションで新しい作業チャプターを選択するには、以下のいずれかの操作を行って、**New Working Chapter** ダイアログを開きます。

- メインメニューから、ファイル ▶ チャプター ▶ 新規ワーキングチャプターを選択します。

図 9.19 に示すように、**New Working Chapter** ダイアログが開きます。

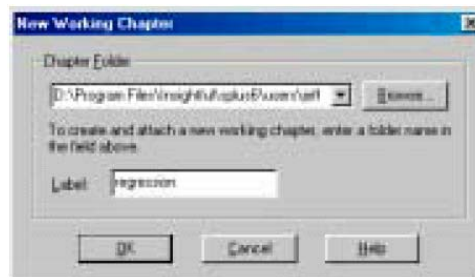


図 9.19 : New Working Chapter ダイアログ

1. **Chapter Folder** テキストボックスで、以下のいずれかの操作を行います。
 - 既存の作業データベースを指定するには、必要な **.Data** フォルダを含むチャプターフォルダのパス名を入力するか、または **Browse** をクリックしてそのフォルダを指定します。

- 新しい作業データベース（新しいチャプターフォルダ内の新しい**Data** フォルダ）を作成して指定するには、新しいチャプターフォルダのパス名を入力します。表示される確認ダイアログで**OK**をクリックしてください。
2. **Label** テキストボックスで、チャプターを識別するのに使用するラベルを指定します。この名前が**オブジェクト・エクスプローラ**に現れます。
 3. **OK** をクリックします。

注意

デフォルトの位置 1 に他のチャプターがすでに追加されている場合は、古いチャプターが位置 2 に移動して終了します。新しい作業チャプターが位置 1 に追加されます。

第 10 章 コマンド・ウィンドウ の使い方

はじめに	491
コマンド・ウィンドウの基礎	492
式を入力する	492
基本シンタックス	493
S-PLUS を終了する	496
コマンドラインの編集	496
S-PLUS のヘルプの使用	497
S-PLUS 言語の基礎	500
データオブジェクト	500
データオブジェクトを管理する	505
関数	507
演算子	508
式	510
関数の省略可能な引数	512
データのインポートと編集	514
データファイルを読み込む	514
キーボードからデータを入力する	514
ASCII ファイルを読み込む	515
データを編集する	516
組込みデータセット	517
データのサブセット抽出	518
ベクトルから抽出する	518
行列から抽出する	519
S-PLUS のグラフィックス	522
プロットを作成する	522
複数プロットのレイアウト	525

第 10 章 コマンド・ウィンドウの使い方

統計	526
要約統計	526
仮説検定	527
統計モデル	529
定義関数	532
バッチモードでの S-PLUS の使用	533

はじめに

S-PLUS は、探索的データ解析と統計のために特別に開発された高機能な言語処理系です。コマンド・ウィンドウは、S-PLUS プログラミング環境のためのウィンドウで、強力な S-PLUS 言語で対話的なプログラミングを直接行うことができます。


注意

この章では、“S-PLUS” という言葉を、言語と評価プログラムの両方における略称として使用します。
--

この章では、S-PLUS 言語を簡単に紹介します。言語とそのプログラムの詳細については、『Programmer's Guide』を参照してください。

コマンド・ウィンドウの基礎

コマンド・ウィンドウを開くときは、以下のいずれかの操作を行ってください。

- メインメニューから、ウィンドウ ▶ コマンド・ウィンドウを選択します。
- 標準ツールバーのコマンド・ウィンドウボタン  をクリックします。

式を入力する

コマンド・ウィンドウを使用するときはプロンプトに式を入力し、RETURN キーを押します。S-PLUS は通常「値 (value)」を返しますが、S-PLUS グラフシートでグラフを作成しているときのように、コマンド・ウィンドウへプロンプトを返すだけのときもあります。

最も単純な S-PLUS 式は、次のような算術式です。

```
> 3+7
[1] 10
> 3*21
[1] 63
```

記号 + および * は、それぞれ加法と乗法の S-PLUS 演算子を表します。S-PLUS には通常の算術演算子と論理演算子の他に、特別な目的の演算子があります。たとえば、コロン演算子 : は次のような数列を得るために使用されます。

```
> 1:7
[1] 1 2 3 4 5 6 7
```

それぞれの出力行の [1] は、S-PLUS 返り値の最初の要素の「インデックス」です。S-PLUS が長い結果ベクトルを返す場合は、行の最初の値に対するインデックスがその行の一番先頭にあります。

```
> 1:30
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
[19] 19 20 21 22 23 24 25 26 27 28 29 30
```

最も一般的な S-PLUS 式は、「関数呼び出し (function call)」です。S-PLUS の「関数 (function)」の例には、カンマで区切った複数のオブジェクトを 1 つのオブジェクトに“結合”するために使用される関数 `c` があります。関数呼び出しは、その後に必ず 1 対の丸かっこがあり、その丸かっこの中には「引数 (argument)」がある場合とない場合があります。

```
> c(3,4,1,6)
[1] 3 4 1 6
```

これまでの例では、S-PLUS は値を返し、コマンド・ウィンドウに表示しているだけでした。S-PLUS 式の値を繰り返し使用するときは、演算子 `<-` で「付値 (assign)」してください。たとえば `newvec` という名前の S-PLUS オブジェクトに上の式を付値するときは、次のように入力します。

```
> newvec <- c(3, 4, 1, 6)
```

オブジェクト `newvec` が作成され、S-PLUS プロンプトが返されます。新しく作成したオブジェクトの内容を見るときは、そのオブジェクトの名前を入力してください。

```
> newvec
[1] 3 4 1 6
```

基本 シンタックス

この節では、S-PLUS の基本的なシンタックスと表記について説明します。

スペース

S-PLUS は、たとえば次のように、ほとんどのスペースを無視します。

```
> 3+    7
[1] 10
```

ただし、数字や名前間にスペースを入れしないでください。また、2 文字の代入演算子 `<-` の両側には必ずスペースを入れるようにしてください。スペースがないと、付値ではなく比較を実行することがあります。

英大小文字

S-PLUS は、Windows や DOS と違って「大文字と小文字を区別」します。S-PLUS のオブジェクト／引数／名前などは、すべて大文字と小文字を区別します。S-PLUS オブジェクトの名前を間違っていると、エラーメッセージが表示されます。次の例のように、すべての大文字と小文字を一致させるように注意してください。

第 10 章 コマンド・ウィンドウの使い方

```
> newvec
[1] 3 4 1 6
> NEWvec
Problem: Object "NEWvec" not found
Use traceback() to see the call stack
Dumped
```

特殊文字

表 10.1 に、キーボードにない文字を打ったり文字列を区切ったりするキャリッジ制御用の特殊文字を示します。

表 10.1 : 特殊文字

文字	説明
¥t	タブ
¥n	改行
¥"	" (二重引用符)
¥'	' (アポストロフィ)
¥¥	¥ (円マーク)
¥###	8進数としての ASCII 文字 (すなわち、#は 0~7 の範囲)

任意の ASCII 文字は、3 桁の 8 進数として表すことができます。S-PLUS では、8 進表現の前に円マーク (¥) を置くことによって文字を指定することができます。たとえばキーボードに縦線がない場合は、"¥174" を使ってそれを指定することができます。8 進表現の ASCII 文字セットは、標準プログラミングテキストにあります。

継続

RETURN キーを押して式が不完全 (たとえば最後の文字が演算子だったり丸かっこが抜けている) ことが明らかなきは、式を完成させるよう促す「継続 (continuation)」プロンプトが表示されます。デフォルトの継続プロンプトは、+ です。

次に、S-PLUS 継続プロンプトで応答を返す 2 つの不完全な式の例を示します。

```
> 3*
+ 21
[1] 63
> c(3,4,1,6
+ )
[1] 3 4 1 6
```

最初の例では、乗算演算子 * の後にデータオブジェクトが続かなければならないため、S-PLUS は式が不完全だと判断しました。次の例では、c(3,4,1,6 に右丸かっこが必要であるために不完全だと判断しました。

上のそれぞれのケースで、ユーザが継続プロンプト + の後に式を完成させ、その結果 S-PLUS は完全に式を評価した結果を返しました。

式の評価を 中断する

時には S-PLUS 式の評価を中断したいことがあるかもしれません。たとえば別のコマンドを使いたくなったり、画面上のデータの出力表示があまりにも長いのでそのすべてを表示させたくないと思うかもしれません。

S-PLUS の処理を中断するときは、ESC キーを押すだけです。

エラー メッセージ

コマンド・ウィンドウでコマンドを入力するときに間違いを恐れないください。入力を間違えても何も壊れることはありません。通常は何らかのエラーメッセージが表示され、その後でもう 1 度試行することができます。

次に、“不適切な” 式の入力によって生じた誤りの例を示します。

```
> .5(2,4)
Problem: Invalid object supplied as function
Use traceback() to see the call stack
Dumped
```

この例では、丸かっこがあるために入力したものを S-PLUS が関数として解釈しようとした。しかし “.5” という名前の関数はありません。

S-PLUS を 終了する

コマンド・ウィンドウから S-PLUS を終了するときは、次のような関数 `q` を使用します。

```
> q()
```

`q` は S-PLUS 関数であり S-PLUS 関数にはすべて丸かっこが必要なので、S-PLUS を終了するためには `q` にも `()` が必要です。


コマンド ラインの編集

コマンド・ウィンドウでは、前に実行した S-PLUS コマンドを呼び出して編集することができます。上矢印と下矢印を使って、セッション中に入力したコマンドのリストを上下にスクロールして表示させることができます。Windows の標準の編集コマンドを使って簡単に入力エラーを訂正し、前に実行したコマンドに基づいて新しいコマンドを作成することができます。たとえば、次の式を入力し、ENTER を押します。

```
> lm(Mileage ~ Weight, data=fuel.frame)
```

次に別の予測変数を追加したい場合には、上矢印を押してコマンドをもう 1 度呼び出し、それを編集することができます。

```
> lm(Mileage ~ Weight + Disp., data=fuel.frame)
```

次に、標準ツールバーの**コマンド履歴**ボタン  をクリックします。図 10.1 に示す **Commands History** ダイアログに、それまでに発行したコマンドのリストが表示され、別の方法でそのコマンドを編集することができます。**Commands History** ダイアログを使って、コマンドを検索し実行することもできます。

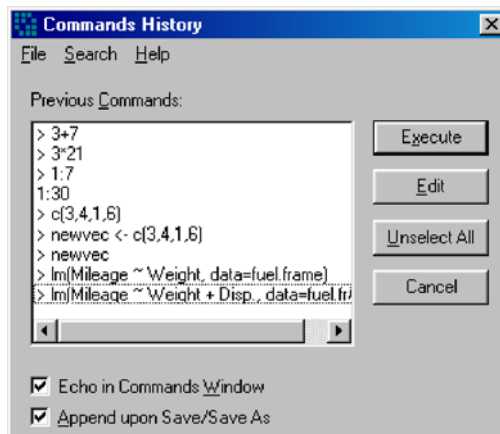


図 10.1 : Commands History ダイアログ

S-PLUS の ヘルプの使用

S-PLUS セッション中にヘルプを見たい場合は、関数 `?` および `help` を使って簡単にヘルプを呼び出すことができます。関数 `?` はシンタックスが単純で、ほとんどの例では丸かっこは不要です。たとえば、

```
> ?lm
```

は、下の図 10.2 に示した `lm` のヘルプファイルを開きます。`?` と `help` は両方とも、ヘルプファイルを HTML 形式で表示します。

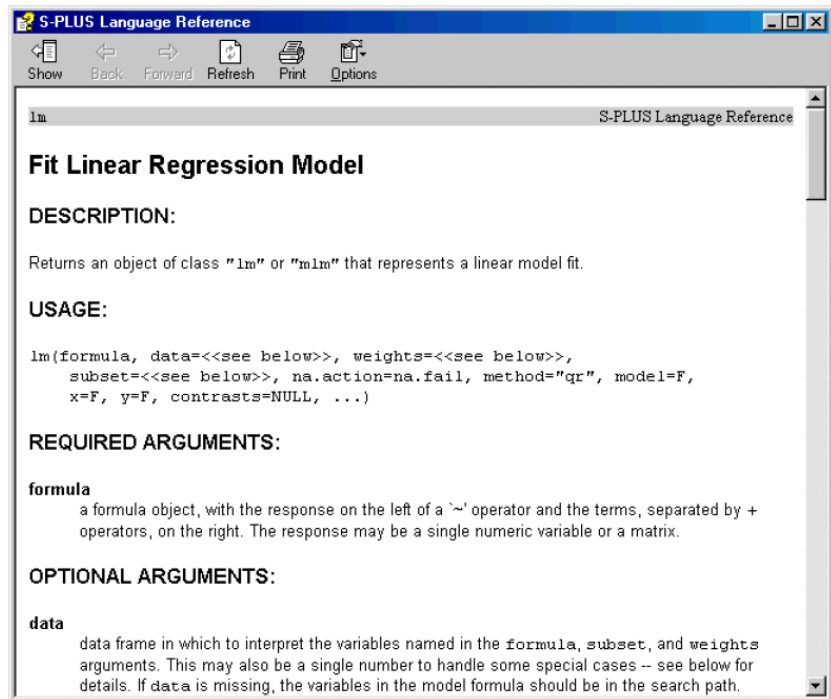


図 10.2 : 関数 `lm` のヘルプファイル

コマンド `?` は、クラスとメソッドに関する情報を得るために特に有効です。関数呼び出しで `?` を使用すると、その関数名自体に関する説明と、評価する場合にその関数が利用できるすべてのメソッドに関する説明が表示されます。特に関数呼び出しが `methods(name)` で `name` が関数名の場合、S-PLUS は現在の参照リストで利用可能な `name` のすべてのメソッドに関する説明を表示します。

```
> ?methods(summary)
The following are possible methods for summary
  Select any for which you want to see documentation:
1:  summary()
2:  summary(<Default>)
3:  summary(object=groupVecVirtual)
4:  summary(object=numericSequence)
5:  summary(object=seriesVirtual)
6:  summary(object=timeDate)
7:  summary(object=timeEvent)
8:  summary(object=timeRelative)
9:  summary(object=timeSequence)
10: summary(object=timeSpan)
11: summary(object=timeZoneC)
12: summary(object=timeZoneS)
Selection:
```

必要なメソッドの番号が入力されると、ヘルプファイルがあれば、S-PLUS はそのメソッドに関連するヘルプファイルを Windows のヘルプシステムに表示します。コマンド ? は、メニューを作成する前にヘルプファイルの有無を確認しません。メニュー選択後、S-PLUS は残りの選択肢を示す更新されたメニューを表示します。

?メニューから S-PLUS プロンプトに戻るときは、0 を入力してください。

引数として S-PLUS 関数/演算子/データセットの名前を指定することで、ヘルプファイルを呼び出すことができます。たとえば以下のコマンドは、関数 c のヘルプファイルを表示します。

```
> help("c")
```

(引用符はほとんどの関数では省略可能ですが、<- などの特殊文字を含む関数と演算子には必須です。)

S-PLUS ヘルプ ファイルを 読み込む

S-PLUS ヘルプシステムから情報を得るためには、以下のようなヘルプファイルの全体の構成に慣れることが重要です (すべての節がすべてのファイルに含まれるわけではありません)。

- **DESCRIPTION** : 関数の主な用途の簡単な説明。
- **USAGE** : すべての引数による関数呼出し。
- **REQUIRED ARGUMENTS** : 必ず指定が必要な引数の説明。
- **OPTIONAL ARGUMENTS** : 省略可能な引数の説明。
- **VALUE** : 関数からの返り値。

- **SIDE EFFECTS** : 関数からの副作用。
- **GRAPHICAL INTERACTION** : ユーザが期待するグラフィカルインタラクションの説明。
- **CLASSES** : デフォルトのメソッドの場合に関数を適用できるクラスの説明。
- **WARNING** : 関数を使用するときにユーザに出される警告について。
- **DETAILS** : アルゴリズム的な詳細と実装についての説明。
- **BACKGROUND** : 関数またはメソッドに関する背景情報。
- **NOTE** : 以上のカテゴリに当てはまらないすべての情報。
- **REFERENCES** : ユーザが追加情報を参照することができる利用可能な文書および論文。
- **BUGS** : 関数の既知のバグの説明。
- **SEE ALSO** : 関連する S-PLUS 関数へのリンク。
- **EXAMPLES** : S-PLUS のコード例。
- **Keywords** : ヘルプファイルをヘルプシステムの目次に入れるキーワードのリスト。

S-PLUS 言語の基礎

この節では S-PLUS 言語を使用するために必要である最も基本的な概念、つまり式／演算子／付値／データオブジェクト／関数呼び出しを説明します。

データ オブジェクト

S-PLUS を使用するとき、データセットをある「クラス (class)」に属する「データオブジェクト (data object)」として考えてください。各クラスには、「スロット (slots)」という名前のリストとして定義される特定の「表現 (representation)」があります。各スロットは、他のクラスのオブジェクトを含んでいます。最も一般的なクラスは、`numeric`、`character`、`factor`、`list`、および `data.frame` です。この章では、最も基本的なデータオブジェクトを紹介します。詳細は、『*Programmer's Guide*』を参照してください。

最も単純なタイプのデータオブジェクトは、すべてが数値、論理値または文字列で、同じ型からなる値の一元配列です。たとえば、数の配列 `-2.0 3.1 5.7 7.3` とすることができます。あるいは、論理値の配列 `T T F T F T F F` とすることができます。ここで、`T` は真を表わし、`F` は偽を表しています。あるいは、たとえば `"sharp claws"` や `"COLD PAWS"` などの一組の順序のある文字列とすることができます。それらの単純な一元配列は、S-PLUS の中では「ベクトル」と呼ばれます。クラス `"vector"` は、オブジェクトを一元配列と見なすことができるすべての基本的なクラスを包含する「仮想のクラス (virtual)」です。ベクトルは、「インデックス (index)」すなわち個数配列内の位置を参照することによって個々の値を抽出したり置換することができます。ベクトルの「長さ (length)」は配列における値の個数であり、ベクトルオブジェクト `x` の有効なインデックスは範囲 `1:length(x)` にあります。ほとんどのベクトルは、`numeric`、`integer`、`logical`、または `character` のクラスのうちの1つに属しています。たとえば、前述のベクトルはそれぞれ、長さが4、8および2、クラスが `numeric`、`logical`、および `character` です。

S-PLUS は、様々な種類の値を含むベクトルのクラスに対しては、情報量が最も多くなるものに合わせます。文字列は最も多くの情報を含み、数値はその次、論理値は最も少ない情報量を含みます。S-PLUS は、情報量の少ない値を情報量の多いタイプの等価な値に強制的に変えます。

```
> c(17, TRUE, FALSE)
[1] 17 1 0
> c(17, TRUE, "hello")
[1] "17" "TRUE" "hello"
```

データ オブジェクト名

オブジェクト名は、英大小文字、数字、およびピリオド (.) で組み合わせることができませんが、文字で始めなければなりません。たとえば、次のものはすべて有効なオブジェクト名です。

```
mydata
data.ozone
RandomNumbers
lottery.ohio.1.28.90
```

次のようにピリオド (.) を使うと、似ているデータセット名を区別しやすくなる場合があります。

```
data.1
data.2
data.3
```

S-PLUS 6 以降で新しく定義されたオブジェクトとメソッドは、区別しやすくするためにピリオドを省略し大文字を加える命名方法をとることがあります。

```
setMethod
signalSeries
```

警告

S-PLUS の組込み関数の名前と同じ名前を使ってはいけません。S-PLUS の組込み関数と同じ名前関数を付値すると、作成したオブジェクトを削除または名前を変更するまで、一時 S-PLUS 関数が使えなくなります。S-PLUS の組込み関数が、新しく作成された関数によってマスクされていることを S-PLUS は警告します。他のオブジェクトをマスクしているオブジェクトのリストを表示させたいときは、関数 `masked` を使用してください。

少なくとも 7 つの S-PLUS 関数の名前は `C/D/c/I/q/s/t` というように 1 文字です。これらの名前はすべて使用を避けるべきですが、特に S-PLUS で頻繁に使用する関数 `c` または `t` は、自作の関数の名前として付けないよう特に注意してください。

ベクトルデータ オブジェクト

これで S-PLUS の最も基本的なオブジェクト、すなわち 1 組の数値／文字値／論理値などであるベクトルについて十分にお分かりになったことでしょう。「ベクトルは、単一の型でなければなりません」。すなわち、値が `T` と `-2.3` から成るベクトルは作成できません。そのようなベクトルを作成しようとすると、S-PLUS はその要素を強制的に共通の型にします。次の例をご覧ください。

```
> c(T, -2.3)
[1] 1.0 -2.3
```

ベクトルには、「長さ (length)」と「型 (mode)」の 2 つの基本的な属性があります。長さは関数 `length` で表示させることができ、型は関数 `mode` で表示させることができます。

行列データ オブジェクト

S-PLUS の重要なデータオブジェクトの型は、たとえば次のような「二次元配列 (two-way array)」すなわち「行列 (matrix)」オブジェクトです。

```
-3.0    2.1    7.6
 2.5    -0.5  -2.6
 7.0    10.0  16.1
 5.3   -21.0  -6.5
```

行列やその高次元である「配列 (arrays)」は、ベクトルと関連付けられますが、特別な構造をもっています。S-PLUS は、行列と配列クラスを別の仮想クラスである `structure` クラスから継承させることによって、これらのオブジェクトを同じように処理します。

行列を作成するには、関数 `matrix` を使用します。関数 `matrix` は、引数として 1 つのベクトルと、行と列の数を指定する 2 つの数を使用します。たとえば、次の通りです。

```
> matrix(1:12, nrow=3, ncol=4)
      [,1] [,2] [,3] [,4]
[1,]    1    4    7   10
[2,]    2    5    8   11
[3,]    3    6    9   12
```

この例において、関数 `matrix` の最初の引数は、1~12 の整数のベクトルです。第 2 と第 3 の引数は、それぞれ行の数と列の数です。各行と各列にラベルが付けられます。行のラベルは [1,] [2,] [3,]、列のラベルは [,1] [,2] [,3] [,4] です。行番号と列番号のこの表記は、行列の数学的表記と同じです。

上の例で、ベクトル 1:12 は最初に第 1 列、次に第 2 列を埋め、その後同じように埋めていきます。このことを、行列を“列単位で”埋めると言います。行列を“行単位で”埋めたい場合は、`matrix` に省略可能な引数 `byrow=T` を指定します。

行列を埋めるために使用されるベクトルの場合、行の数によって列の数が決まり、またその逆のこともあります。したがって、`matrix` の引数として行と列の数を両方与える必要はありません。一方を指定するだけで十分です。次のコマンドは前の例と同じ行列を生成します。

```
> matrix(1:12, 3)
```

また、列の数だけを指定して同じ行列を作成することができます。次のように入力します。

```
> matrix(1:12, ncol=4)
```

デフォルトでは 2 番目の引数が行の数として受け取られるため、省略可能な引数 `ncol=4` を「名前=値」の形で入力しなければなりません。第 2 の引数として「名前=値」の形式 `ncol=4` を使用すると、デフォルト値が置き換わります。関数呼び出しにおける省略可能な引数指定の詳細については、「関数の省略可能な引数」(512 ページ) を参照してください。

配列のクラスには、一般に、実際の値を保持する `.Data` スロット、次元を表すベクトルを保持する `.Dim` スロット、および行と列名を保持する省略可能な `.Dimnames` スロットの 3 つのスロットがあります。行列データオブジェクトの最も重要なスロットは、次元スロット `.Dim` です。関数 `dim` を使用してオブジェクトの次元を表示することができます。

```
> my.mat <- matrix(1:8,4,2)
> dim(my.mat)
[1] 4 2
```

これは、行列 `my.mat` の次元が 4 行×2 列であることを示しています。また、行列オブジェクトには長さや型があり、これらは `.Data` スロットのベクトルの長さや型に対応します。関数 `length` および `mode` を使用して行列のこのような特性を表示させることができます。ベクトルと同じように、行列オブジェクトには、1 つの `mode` だけがあります。したがって、たとえば 1 列の数値データと 1 列の文字データを含む 2 列の行列は作成できません。このような場合にはデータフレームを使います。

データフレーム オブジェクト

S-PLUS には、行列オブジェクトによく似た「データフレーム (data frame)」オブジェクトと呼ばれるオブジェクトがあります。データフレームオブジェクトは行と列のデータから成り、列の型が異なってもよいという点以外は行列オブジェクトと同じです。以下のオブジェクト `baseball.df` は、1988 年からの野球データを含むデータフレームオブジェクトです。最初の 2 列は `factor` オブジェクト (選手名のコード)、次の 2 列は `numeric`、最後の列は `logical` です。

```
> baseball.df
      bat.ID pitch.ID event.typ outs.play err.play
r1 pettg001 clemr001         2         1         F
r2 whitl001 clemr001        14         0         F
r3 evand001 clemr001         3         1         F
r4 trama001 clemr001         2         1         F
r5 andeb001 morrj001         3         1         F
r6 barrm001 morrj001         2         1         F
r7 boggw001 morrj001        21         0         F
r8 ricej001 morrj001         3         1         F
```

リスト オブジェクト

「リスト (list)」オブジェクトは、一般的で柔軟性の高いオブジェクトです。リストは、順序付けした「成分 (component)」の集まりです。リストの各成分は、データオブジェクトであり、成分の型がそれぞれ異なってもかまいません。たとえばリストは、文字列のベクトル、数値の行列および別のリストからなる 3 つの成分を含むことができます。したがってリストは異なるタイプまたは型の成分を含むことができるので、ベクトルや行列よりも一般的であり、四角形 (行×列) になるように制限されないためデータフレームよりも一般的です。

リストは、関数 `list` で作成することができます。たとえば、一方が `numeric` 型のベクトルで他方が文字列のベクトルである 2 つの成分のリストを作成するときは、次のように入力します。

```
> list(101:119,c("char string 1","char string 2"))
[[1]]:
 [1] 101 102 103 104 105 106 107 108 109 110 111 112 113
 [14] 114 115 116 117 118 119

[[2]]:
 [1] "char string 1" "char string 2"
```

リストの成分は二重の角かっこで囲んだ数字 (この場合は `[[1]]` と `[[2]]`) でラベルが付けられます。この表記により、リストの成分のナンバリングがベクトルや行列のナンバリングと区別されます。それぞれの成分ラベルの後に、その成分の内容が表示されます。

成分に名前を付けると、リストの成分を参照しやすくなります。これは、関数 `list` の各引数に名前を付けることによって行います。たとえば上記と同じリストの成分に “a” と “b” と名前を付けて作成し、次にリストデータオブジェクトを名前 `xyz` で保存することができます。

```
> xyz <- list(a = 101:119,
+ b = c("char string 1", "char string 2"))
```

`list` オブジェクトから成分名を利用して抽出を行うには、リストの名前の後に記号 `$` と成分の名前を続けて使用します。たとえば次の 2 つのコマンドはそれぞれ、リスト `xyz` の成分 `a` と成分 `b` を示します。

```
> xyz$a
[1] 101 102 103 104 105 106 107 108 109 110 111 112 113
[14] 114 115 116 117 118 119
> xyz$b
[1] "char string 1" "char string 2"
```

データ オブジェクト を管理する

S-PLUS では、コマンドラインで作成したオブジェクトは削除するまでディスク上に永久的に保存されます。この節では、データオブジェクトの名前付け／保存／リスト／削除方法について説明します。

データ オブジェクトの 付値

S-PLUS のデータに名前を付けて保存するときは、「付値 (assignment)」演算子 `<-` または `_` を使用します (名前に文字 `_` を使用しないでください)。たとえば数値 4、3、2、1 から成るベクトルを作成し、そのベクトルを名前 `x` で保存するときは、関数 `c` を使って次のように入力します。

```
> x <- c(4,3,2,1)
```

`<-` はキーボードの 2 つのキー、すなわち “より小さい” キー (`<`) と次にマイナス記号 (`-`) を、間にスペースを入れずに入力します。

`y` に 1~10 の整数を含むベクトルを保存するときは、次のように入力します。

```
> y <- 1:10
```

次の付値式は演算子 `_` を使用しており、前の 2 つの付値と同じです。

```
> x _ c(4,3,2,1)
> y_1:10.
```

<- 形式の付値演算子はきわめて分かりやすいため、このマニュアルの例では矢印を使用します。

データ オブジェクトを 保存する

作業ディレクトリにあるデータオブジェクトは永続オブジェクトです。S-PLUS を終了して再起動した場合でもそれらは残っています。

また関数 `attach` (または**オブジェクト・エクスプローラ**) を使って、S-PLUS オブジェクトを保存するディレクトリの位置を変更することもできます。詳細は、`attach` ヘルプファイルを参照してください。

データ オブジェクトを リストする

作業ディレクトリにあるデータオブジェクトの名前のリストを表示するときは、次のように関数 `objects` を使用します。

```
> objects()
```

505 ページのようにベクトル `x` および `y` を作成した場合は、作業ディレクトリのデータオブジェクトのリストにそれらが表示されます。

また S-PLUS の関数 `objects` は、名前が引数として与えられた文字列と一致するオブジェクトを探します。引数 `pattern` を使うと、ワイルドカードの文字を含むことができます。たとえば次の式は、英字 `d` で始まるすべてのオブジェクトを表します。

```
> objects("d*")
```

ワイルドカードとその働きについては、`grep` のヘルプファイルを参照してください。

データ オブジェクトを 削除する

S-PLUS のオブジェクトは永続的であるため、必要のないオブジェクトをときどき削除してください。オブジェクトを削除するには関数 `rm` を使用します。関数 `rm` は引数として省略可能な数のオブジェクトを取得し、1 つずつ削除します。たとえば `a` と `b` と名前をついた 2 つのオブジェクトを削除するときは、次の式を使用します。

```
> rm(a,b)
```

データ オブジェクトを 表示する

保存したデータオブジェクトの内容を見るときは、その名前を入力してください。

```
> x
[1] 4 3 2 1
> y
[1] 1 2 3 4 5 6 7 8 9 10
```

関数

「関数 (function)」は、通常 1 つまたは複数の「引数 (argument)」に基づいて演算を実行した後で値を返す S-PLUS 式です。たとえば関数 `c` は、`c` の引数として指定されたベクトルを結合することによって形成されるベクトルを返します。関数名とその後の 1 対の丸かっこから成る式を入力することにより、関数を「呼び出し (call)」ます。この丸かっこは、カンマで区切ったいくつかの引数を囲むことができます。たとえば `runif` は、0~1 で一様に分布した乱数を生成する関数です。S-PLUS でそのような 10 個の数を計算させるときは、次のように `runif(10)` と入力します。

```
> runif(10)
[1] 0.6033770 0.4216952 0.7445955 0.9896273 0.6072029
[6] 0.1293078 0.2624331 0.3428861 0.2866012 0.6368730
```

S-PLUS は、関数で計算した結果の後に新しいプロンプトを付けて表示します。この場合、結果は一様乱数ジェネレータによって生成された 10 の乱数から成る `vector` オブジェクトです。角かっこで囲まれた数 (ここでは [1] と [6]) は、各行に何個の数が表示されているかを追跡したり、特定の数を探したりするのに役立ちます。

たとえばよく使用される S-PLUS 関数の 1 つは、次のようにデータ値をベクトルに結合することができる関数 `c` です。

```
> c(3,7,100,103)
[1] 3 7 100 103
> c(T,F,F,F,T,T)
[1] T F F F T T
> c("sharp teeth", "COLD PAWS")
[1] "sharp teeth" "COLD PAWS"
> c("sharp teeth", 'COLD PAWS')
[1] "sharp teeth" "COLD PAWS"
```

最後の例は、二重引用符文字 " または一重引用符文字 ' を使って文字列を区切ることができることを示しています。

たとえば関数 `c` の結果を、次のように (削除するまで) 永続的に保存される別の名前のオブジェクトに割り当てたいとします。

```
> weather <- c("hot day", "COLD NIGHT")
> weather
[1] "hot day" "COLD NIGHT"
```

一般に、S-PLUS のいくつかの関数は引数なしで使用されます。たとえば、S-PLUS を終了するときに `q()` を入力することを思い出してください。ただし、式が関数であることを S-PLUS が認識できるように丸かっこは必要です。

関数名だけを入力すると、画面に関数定義テキストが表示されます。(オブジェクト名を入力すると、S-PLUS はそのオブジェクトを表示します。関数オブジェクトは単に関数の定義です。) 関数を呼び出すときは、関数名に丸かっこをつけて再入力する必要があります。

たとえば S-PLUS を終了させたいときに、`q()` ではなく間違っ `q` と入力すると、関数 `q` の定義が表示されます。この場合、関数の本体の長さは 2 行だけです。

```
> q
function(...)
.Internal(q(...), "S_dummy", T, 33)
>
```

これによる障害はありません。正しく `q()` と入力して S-PLUS を終了してください。

```
> q()
```

演算子

「演算子 (operator)」は、引数の間にある 1 つまたは複数の特殊記号によって表すことができる最大 2 つの引数を含む関数です。

たとえば加法／減法／乗法／除法の通常の算術演算はそれぞれ、演算子 `+`、`-`、`*`、`/` で表されます。次に、算術演算子を使ったいくつかの簡単な計算例を示します。

```
> 3+71
[1] 74
> 3*121
[1] 363
> (6.5 - 4)/5
[1] .5
```

指数演算子 \wedge は、次のように使用することができます。

```
> 2 ^ 3
[1] 8
```

いくつかの演算子は1つの引数のみではたらくため、「単項 (unary)」演算子と呼ばれます。たとえば、減法演算子は単項演算子としてはたらくことができます。

```
> -3
[1] -3
```

コロン : は、連続する整数を生成する重要な演算子です。

```
> 1:10
[1] 1 2 3 4 5 6 7 8 9 10
```

表 10.2 は、比較と論理のための S-PLUS 演算子のリストです。比較は論理データにとって最も一般的な演算の一つです。

```
> (1:10) > 5
[1] F F F F F T T T T T
```

比較演算と論理演算は、多くの場合データのサブセットを抽出するのに便利です。論理的な比較を指定する条件付き論理演算子は、関数の制御フローにおいて重要な役割を果たします。

表 10.2 : 論理演算子と比較演算子

演算子	説明	演算子	説明
==	～と等しい	!=	～と等しくない
>	より大きい	<	より小さい
>=	～以上	<=	～以下
&	ベクトル論理積		ベクトル論理和
&&	条件付き論理積		条件付き論理和
!	否定		

式

「式 (expression)」は、関数／演算子／データオブジェクトの任意の組み合わせです。たとえば、次の式は、次のように演算子 (付値演算子) と関数 (結合関数) を含む式です。

```
x <- c(4,3,2,1)
```

次に、S-PLUS で使用する様々な式のいくつかの例を示します。

```
> 3 * runif(10)
[1] 1.6006757 2.2312820 0.8554818 2.4478138 2.3561580
[6] 1.1359854 2.4615688 1.0220507 2.8043721 2.5683608
> 3*c(2,11)-1
[1] 5 32
> c(2*runif(5),10,20)
[1] 0.6010921 0.3322045 1.0886723 0.3510106
[5] 0.9838003 10.0000000 20.0000000
> 3*c(2*x,5)-1
[1] 23 17 11 5 14
```

上の最後の 2 つの例は、S-PLUS 関数の一般的な特徴を表しています。すなわち、関数の引数自体が S-PLUS 式でもかまいません。

次の 3 つの式の例は、ベクトルと数の両方を含む式を使用するときに、S-PLUS で演算がどのように行われるかをよく表しているのが重要です。x が数 4、3、2、1 からなる場合、次の演算は x のそれぞれの要素に適用されます。

```
> x-1
[1] 3 2 1 0
> 2*(x-1)
[1] 6 4 2 0
> x ^ 2
[1] 16 9 4 1
```

一方の引数がベクトルで、他方の引数が数値の場合、演算はベクトルの各成分について適用されます。

優先順位

S-PLUS 式の評価は、下の表 10.3 に示す「優先順位 (precedence hierarchy)」に従います。表の上の方にある演算子は下の方にある演算子よりも優先度が高く、同じ行の演算子の優先度は同じです。

表 10.3 : 演算子の優先順位

演算子	用途
\$	リストの成分
[[[部分、要素の抽出
^	べき乗
-	負号 (単項減算)
:	等差数列
%% %/% %*%	剰余、整数商、行列積
* /	乗算、除算
+ -	加算、減算
<> <= >= == !=	比較
!	否定
& &&	かつ、または
~	モデル式
<<- -> <- _	付値

注意

演算子 \wedge を使用するとき、底が負の数の場合、指数は整数でなければなりません。

優先度が同じ演算子の中では、評価は式の左から右に進みます。式の評価の優先順位がはっきりしないときは、丸かっこを使って順位を明らかにしてください。S-PLUS には、一番内側の丸かっこから一番外側の丸かっこに向かって評価が行われるという、多くのコンピュータ言語と共通の特徴があります。たとえば、 x という名前のベクトル (長さ 1) に値 5 を付値してみます。

```
> x <- 5
```

「等差数列 (sequence)」演算子 `:` を使って、丸かっこがある場合とない場合の式の評価の違いを示します。式 $1:(x-1)$ の場合、 $(x-1)$ が先に評価されて結果が 4 になり、S-PLUS は 1~4 までの整数を表示します。

```
> 1:(x-1)
[1] 1 2 3 4
```

丸かっこを外すと、式は $1:x-1$ となります。演算子 `:` は演算子 `-` よりも優先度が高いため、 $1:x-1$ は “1~5 の整数をとり、次に各整数から 1 を引く。” という意味で S-PLUS に解釈されます。したがって、出力は次のように長さが 4 ではなく 5 になり、1 ではなく 0 から始まります。

```
> 1:x-1
[1] 0 1 2 3 4
```

S-PLUS を使用するとき、丸かっことデフォルト演算子の階層の効果に注意してください。

関数の省略可能な引数

S-PLUS 関数の 1 つの優れた特徴は、「省略可能 (optional)」な引数の使用による柔軟性の高さです。また、省略可能な引数にはあらかじめ適当な「デフォルト値 (default)」が組み込まれ、「必須 (required)」の引数の数が最小限に抑えられているため簡潔性が維持されます。

ヘルプファイルの **REQUIRED ARGUMENTS** と **OPTIONAL ARGUMENTS** の節を調べることによって、どの引数が必須でどの引数が省略可能かを判断することができます。

たとえば平均が 0 で標準偏差が 1 の 50 個の正規乱数を生成するとき、以下の式を使用します。

```
> rnorm(50)
```

平均が 3 と標準偏差が 5 の 50 個の正規乱数を生成したい場合は、次のいずれかの式を使用します。

```
> rnorm(50, 3, 5)
> rnorm(50, sd=5, mean=3)
> rnorm(50, m=3, s=5)
> rnorm(m=3, s=5, 50)
```

最初の式では、省略可能な引数を「値単位 (by value)」で入力しています。省略可能な引数を値単位で入力するときは、すべての引数をヘルプファイルの **USAGE** 文で示された順序で入力してください。

上の 2 番目から 4 番目の式では、省略可能な引数を「名前=値 (by name)」で入力しています。省略可能な引数を「名前=値」で入力するときは、順序は重要ではありません。ただしスタイルを統一させるために、必須の引数の後に省略可能な引数を入力することをお勧めします。

3 番目と 4 番目の式は名前が一意に識別される場合に限って、便宜上省略可能な引数名を省略できることを示しています。引数を「名前=値」で入力すると、引数をどの順序でも入力できるため便利であることがお分かりになるでしょう。

もちろん省略可能な引数をすべて指定する必要はありません。たとえば、次に平均が 0 (デフォルト) と標準偏差が 5 の 50 個の正規乱数を生成する同じ 2 つの方法を示します。

```
> rnorm(50, m=0, s=5)
> rnorm(50, s=5)
```

データのインポートと編集

S-PLUS で処理するデータには様々な種類とサイズがあります。最初のステップは、データを S-PLUS の適切なデータオブジェクト形式にすることです。この節では、ファイルとして存在するデータをインポートする方法と、小さいデータをキーボードから入力する方法を説明します。

データ ファイルを 読み込む

対象となるデータは S-PLUS で作成されることもありますが、ASCII ファイルのような他の形式であったり、SAS などの他のソフトウェアパッケージで誰かが作成したものである可能性もあります。S-PLUS 関数 `importData` を使って、様々なソースからデータを読み込むことができます。

たとえば S-PLUS の作業ディレクトリに、`test.sd2` という名前の SAS ファイルがあるとします。このファイルを関数 `importData` を使ってインポートするときは、`file` (読み込むファイルの名前) と `type` (読み取るファイルのタイプ) の 2 つの必須の引数を入力してください。

```
> myData <- importData(file="test.sd2", type="SAS")
```

S-PLUS はデータファイルを読み込み、データフレーム `myData` を作成します。

キーボード からデータを 入力する

S-PLUS に小さいデータを入力するときは、引数なしの関数 `scan()` を使って S-PLUS データオブジェクトを作成します。

```
mydata <- scan()
```

ここで `mydata` は、任意の正当なデータオブジェクト名です。S-PLUS は、以下の例に示すように入力を要求します。14 個のデータ値を入力し、それをオブジェクト `diff.hs` に付値します。S-PLUS プロンプトで名前 `diff.hs` を入力し、それに `scan` コマンドの結果を付値してください。これに応じて S-PLUS は、最初の値を入力しなければならないことを意味するプロンプト `1:` を表示します。

スペースで区切って 1 行に好きなだけいくつも値を入力することができます。RETURN を押すと、S-PLUS は次の値のインデックスを表示し、入力を要求します。次の例では、最初の行に 5 つの値を入力したので S-PLUS は `6:` で応答します。データの入力が終わったら、プロンプト: に応答して RETURN を押すと、S-PLUS は S-PLUS コマンドプロンプト `>` に戻ります。

画面に、完成した例が次のように現われます。

```
> diff.hs <- scan()
1: .06 .13 .14 -.07 -.05
6: -.31 .12 .23 -.05 -.03
11: .62 .29 -.32 -.71
15:
>
```

ASCII ファイルを 読み込む

キーボードからのデータの inputs は、S-PLUS では比較的珍しい作業です。多くの場合、S-PLUS に読み込みたいベクトルデータが ASCII ファイルとして既に保存されています。ASCII ファイルは通常、スペース/タブ/復帰改行/その他の区切り記号で分けられた数値から成ります。

たとえば S-PLUS 作業ディレクトリに、次のようなデータを含む **vec.dat** という名前のファイルがあるとしましょう。

```
62 60 63 59
63 67 71 64 65 66
88 66 71 67 68 68
56 62 60 61 63 64 63 59
```

引数として "vec.dat" を含む scan コマンドを使って、ファイル **vec.dat** を S-PLUS へ読み込みます。

```
> x <- scan("vec.dat")
```

関数 scan の引数 vec.dat の両側には引用符が必要です。次に x を入力して、ファイル **vec.dat** から S-PLUS に読み込んだ x という名前のデータオブジェクトを表示することができます。

読み込みたいファイルが S-PLUS 作業ディレクトリにない場合は、フルパス名を使用してください。したがってファイル **vec.dat** がパス名 **c:\mabel\test\vec.dat** のディレクトリにある場合は、次のように入力します。

```
> vec.data <- scan("c:\mabel\test\vec.dat")
```

(S-PLUS は円マーク ¥ をエスケープ文字として扱うため、円マークを二重にしなければならないことに注意してください。)

ASCII ファイルから、他のデータオブジェクト、特にデータフレーム作成のためにデータテーブルを読み込むこともできます。たとえば、以下のデータファイル **auto.dat** があると仮定します。

Model	Price	Country	Reliab	Mileage	Type
AcuraIntegra4	11950	Japan	5	NA	Small
Audi1005	26900	Germany	NA	NA	Medium
BMW325i6	24650	Germany	94	NA	Compact
ChevLumina4	12140	USA	NA	NA	Medium
FordFestiva4	6319	Korea	4	37	Small
Mazda929V6	23300	Japan	5	21	Medium
MazdaMX-5Miata	13800	Japan	NA	NA	Sporty
Nissan300ZXV6	27900	Japan	NA	NA	Sporty
OldsCalais4	9995	USA	2	23	Compact
ToyotaCressida6	21498	Japan	3	23	Medium

次のように関数 `read.table` を使って、S-PLUS データフレームに読み込むことができます。

```
> auto <- read.table("auto.dat", header=T)
```

省略可能な引数 `header=T` は、変数名にファイルの最初の行を使用するよう S-PLUS に指定します。

またタイプ "ASCII" を使って、関数 `importData` によって ASCII ファイルを読み込むこともできます。

データを編集する

S-PLUS データオブジェクトを作成した後で、入力したデータの一部を変更したいことがあるかもしれません。データオブジェクトの編集には、S-PLUS データ・ウィンドウにデータを開く関数 `Edit.data` を使用します。S-PLUS 関数を編集するとき、修正の最も簡単な方法は、関数 `Edit` を使って編集用の S-PLUS スクリプト・ウィンドウに関数をダンプすることです。より高機能なテキストエディタを使用するときは、S-PLUS セッションのオプションで指定したエディタ（デフォルトではメモ帳）を使用する関数 `fix` を使用してください。

`fix` を使って、元のデータオブジェクトのコピーを作成して編集し、次にその結果をその元の名前で再び付値することができます。既に好みのエディタをお持ちの場合は、関数 `options` で指定することによってそのエディタを使用することができます。たとえばエディタとして **Microsoft Word** を使用したい場合は、以下のように簡単に設定することができます。

```
> options(editor="c:\Program Files\Microsoft Office\
+ Office\winword")
```

組込みデータ セット

S-PLUS には多数の「組込み (built-in)」データセットが付属しています。これらのデータセットは、S-PLUS の使い方を説明する例として用いることができます。S-PLUS を補助教材として使用するとき、この組込みデータセットはデータ解析における問題解決の基礎として役立ちます。

組込みデータセットを表示させるには、> プロンプトでその名前を入力してください。S-PLUS の組込みデータセットには、様々なタイプのデータオブジェクトが含まれています。

これらの組込みデータセットを見つけるときは、関数 `search` を使用します。そのとき参照可能なオブジェクトデータベースのリストが出力されます

```
> search()
[1] "D:\PROGRA~1\INSIGH~1\splus70\users\lenk"
[2] "splus"
[3] "stat"
[4] "data"
[5] "trellis"
[6] "nlme3"
[7] "menu"
[8] "sgui"
[9] "winjava"
[10] "SPXML"
[11] "main"
```

データのサブセット抽出

S-PLUS 言語のもう 1 つの優れた特徴は、表示またはさらなる操作のためにデータのサブセットを抽出する機能です。この章の例ではベクトルと行列のサブセットの抽出について説明しますが、同じ方法を使って他の S-PLUS データオブジェクトからデータのサブセットを抽出することができます。

ベクトルから抽出する

次のように、整数 5、14、8、9、5 から成る長さ 5 のベクトルを作成します。

```
> x <- c(5,14,8,9,5)
> x
[1] 5 14 8 9 5
```

このベクトルの 1 つの要素を表示するときは、ベクトルの名前の後に、[] の間に要素のインデックスをはさんだものを入力してください。たとえば次のように最初の要素を表示するときは `x[1]` と入力し、第 4 の要素を表示するときは `x[4]` と入力します。

```
> x[1]
[1] 5
> x[4]
[1] 9
```

複数の要素を一度に表示させるときは、[] 文字内に関数 `c` を使います。次の式は、`x` の 2 番目と 5 番目の要素を表示します。

```
> x[c(2,5)]
[1] 14 5
```

指定した要素または要素のリストを除くすべての要素を表示するときは、マイナスを使用してください。たとえば `x[-4]` は、4 番目以外のすべての要素を表示します。

```
> x[-4]
[1] 5 14 8 5
```

同様に、`x[-c(1,3)]` は最初と 3 番目以外の要素をすべて表示します。

```
> x[-c(1,3)]
[1] 14 9 5
```

サブセット化のより高度な使い方は、[]の間に論理式をはさんで使用することです。論理式は、ベクトルを真および偽の2つのサブセットに分割します。添え字として使用されるとき、式は条件が真のサブセットのみを返します。

たとえば次の式は、値が8よりも大きい要素をすべて選択します。

```
> x[x>8]
[1] 14 9
```

この場合、値が14および9であるxの2番目と4番目の要素は、論理式 $x > 8$ の要件を満たすので表示されます。

S-PLUS では通常、演算の結果を別のオブジェクトに付値することができます。たとえば上で選択したサブセットを y という名前のオブジェクトに付値して y を表示したり、次の計算に y を使用したりすることができます。

```
> y <- x[x>8]
> y
[1] 14 9
```

次の節では、同じ原理が行列データオブジェクトにも適用されるけれども、選択できる次元が2つあるためシンタックスが少し複雑になることがお分かりになるでしょう。

行列から抽出する

角かっこ内にカンマで区切った行と列のインデックスを入力することによって、行列の1つの要素を選択することができます。組込みデータセット `state.x77` を使ってこれを示します。[]演算子の中の最初のインデックスは行インデックスであり、第2のインデックスは列インデックスです。以下のコマンドは、`state.x77` の第3行第8列の値を表します。

```
> state.x77[3,8]
[1] 113417
```

第 10 章 コマンド・ウィンドウの使い方

行と列の各軸にラベルが定義された場合に、行と列の `dimnames` (軸名) を使って要素を表示することもできます。したがって、`Arizona` という名前の行と `Area` という名前の列にある `state.x77` の値を表示させるときは次のコマンドを使用します。

```
> state.x77["Arizona", "Area"]
[1] 113417
```

行列オブジェクトから連続する行/列を選択するときは、行/列のインデックスに演算子 `:` を使用してください。以下の式は、最初の 4 つの行と第 3 列～第 5 列を選択し、オブジェクト `x` に付値する例です。

```
> x <- state.x77[1:4, 3:5]
> x
      Illiteracy Life Exp Murder
Alabama      2.1   69.05   15.1
Alaska       1.5   69.31   11.3
Arizona      1.8   70.55    7.8
Arkansas     1.9   70.66   10.1
```

関数 `c` を使って、前にベクトルに使用したのと同じように行列の複数の行/列を選択することができます。たとえば次の式は、`state.x77` の第 5 行、第 22 行および第 44 行と第 1 列、第 4 列および第 7 列を選択します。

```
> state.x77[c(5, 22, 44), c(1, 4, 7)]
      Population Life Exp Frost
California  21198   71.71    20
Michigan    9111   70.63   125
Utah       1203   72.90   137
```

前と同じように、行名または列名が定義されている場合はそれをインデックス番号の代わりに使用することができます。

```
> state.x77[c("California", "Michigan", "Utah"),
+ c("Population", "Life Exp", "Frost")]
      Population Life Exp Frost
California  21198   71.71    20
Michigan    9111   70.63   125
Utah       1203   72.90   137
```


すべての行を選択するときは、カンマの前の式を空白のままにします。すべての列を選択するときは、カンマの後の式を空白のままにします。次の式は California、Michigan および Utah の行に対するすべての列を選択します。後ろの角かっこがカンマのすぐ後にあるので、すべての列が選択されていることに注目してください。

```
> state.x77[c("California","Michigan","Utah"),]
      Population Income Illiteracy Life Exp Murder
California    21198    5114         1.1    71.71    10.3
Michigan      9111    4751         0.9    70.63    11.1
Utah          1203    4022         0.6    72.90     4.5

      HS Grad Frost Area
California    62.6    20 156361
Michigan     52.8    125 56817
Utah         67.3    137 82096
```

S-PLUS のグラフィックス

グラフィックスは、データ解析の全体のステップとしてデータを視覚的に調べる S-PLUS の中心的なツールです。組み込まれている様々なグラフィックス関数とそのプログラム機能により、S-PLUS を使ってデータをあらゆる角度から調べることができます。この節では、S-PLUS を使って簡単なプロットを作成する方法を説明します。他の様々な種類のプロットを作成する作業に S-PLUS を使用するときは、『Application Developer's Guide』の第 2 章「Traditional Graphics」と第 3 章「Traditional Trellis Graphics」を参照してください。

プロットを作成する

レーザープリンタの普及に伴い、技術データ／科学データ／金融データ／マーケティングデータのプロットは、最も強力によく利用される S-PLUS の機能の 1 つです。S-PLUS では、様々な作図関数とグラフィックス関数を使用することができます。

最もよく使用される S-PLUS 作図関数は、plot です。次のように作図関数を呼び出すと、S-PLUS グラフィックスのウィンドウにプロットが表示されます。

```
> plot(car.miles)
```

引数 car.miles は、組込み S-PLUS ベクトルデータオブジェクトです。他の引数がないため、データはその固有のインデックス、言い換えれば 1~120 の観測値の数に対してプロットされます。

燃費に関心があるため、car.gals に対する car.miles の関係をプロットしたいことがあります。これも、plot で簡単に処理することができます。

```
> plot(car.gals, car.miles)
```

結果を図 10.3 に示します。

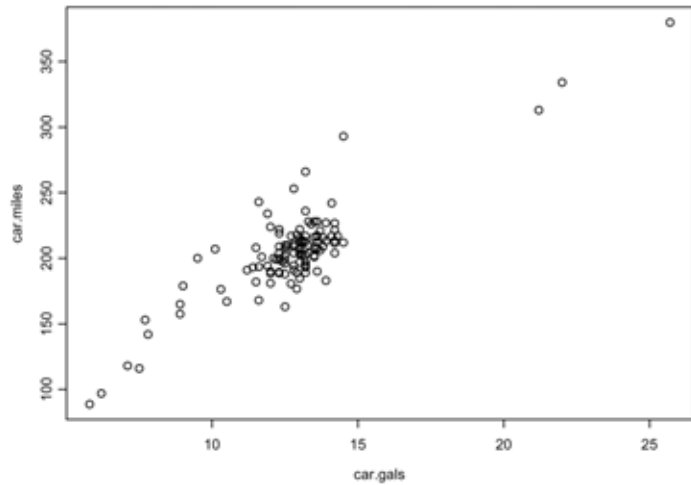


図 10.3 : S-PLUS プロット

S-PLUS グラフィックスのウィンドウにグラフを表示するには `plot` の他に、様々な S-PLUS 関数を使用することができます。これらの関数の多くは、それぞれ高水準と低水準の作図関数を示す表 10.4 と表 10.5 にリストされています。高水準作図関数は軸を含む新規のプロットを作成し、低水準作図関数は一般的に既存のプロットに点などが追加されます。

表 10.4 : 一般的な高水準作図関数

<code>barplot, hist</code>	棒グラフ、ヒストグラム
<code>boxplot</code>	箱型図
<code>brush</code>	ブラッシュと対散布図：スピン 3D 軸
<code>contour, image, persp, symbols</code>	等高線、イメージ、鳥かん図、シンボルグラフ
<code>coplot</code>	条件付きプロット
<code>dotchart</code>	点グラフ
<code>faces, stars</code>	多変量データを表示する（チャーノフの顔、星型図）

表 10.4 : 一般的な高水準作図関数 (続き)

map	米国の全体または一部分をプロットする (地図ライブラリの一部)
pairs	すべての対散布図をプロットする
pie	円グラフ
plot	作図 (総称関数)
qqnorm, qqplot	正規および一般 QQ プロット
scatter.smooth	平滑曲線付き散布図
tsplot	時系列をプロットする
usa	米国の境界線をプロットする

表 10.5 : 一般的な低水準作図関数

abline	線を切片勾配形式で追加する
axis	軸を加える
box	プロットのまわりに枠を加える
contour, image, persp, symbols	プロットに等高線、イメージ、鳥かん図、シンボルを加える
identify	マウスを使ってグラフ上の点を識別する
legend	プロットに凡例を加える
lines, points	プロットに線または点を加える
mtext, text	余白またはプロットにテキストを加える
stamp	プロットに日付と時間情報を加える
title	プロットにタイトル/x 軸ラベル/y 軸ラベル/サブタイトルを加える

複数プロットのレイアウト

1つのウィンドウまたは1枚のハードコピーページに、複数のプロットを表示させたいことがあります。これを行うときは、S-PLUS 関数 `par` を使ってプロットのレイアウトを制御します。次の例では、この目的で `par` を使用する方法を説明します。関数 `par` は、S-PLUS プロットの多くの部分を制御したりカスタマイズするために使用されます。

この例では、`par` を使ってウィンドウまたはページに4つのプロットを2つずつ2段で作成します。以下の関数 `par` に続けて、4つのプロットコマンドを発行します。それぞれメインタイトルを持つ簡単なプロットが作成されます。

```
> par(mfrow=c(2,2))
> plot(1:10,1:10,main="Straight Line")
> hist(rnorm(50),main="Histogram of Normal")
> qqnorm(rt(100,5),main="Samples from t(5)")
> plot(density(rnorm(50)),main="Normal Density", type="l")
```

結果を図 10.4 に示します。

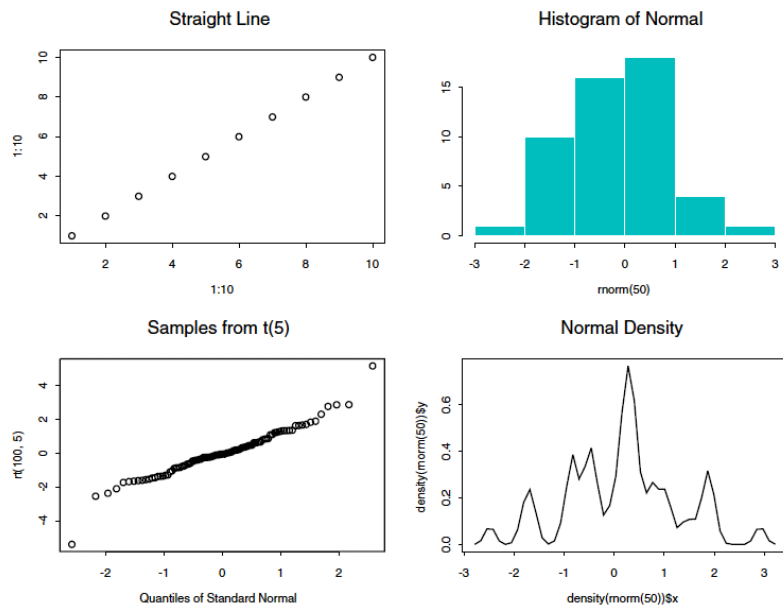


図 10.4 : 複数のプロットのレイアウト

統計

S-PLUS には、仮説検定／線形回帰／分散分析／分割表／因子分析／生存時間解析／時系列解析などのすべての種類の統計解析を行う関数があります。

この節では、要約統計を生成し、仮説検定を実行し、統計モデルをあてはめる関数の概要を説明します。

要約統計

S-PLUS は、分布の中心と尺度（スケール）に対して頑健で抵抗力のある様々な推定量と共に、データセットのすべての標準的な要約統計を計算する関数を含みます。表 10.6 に、要約統計として最も一般的な関数のリストを示します。

表 10.6 : 要約統計用の一般的な関数

cor	相関係数
cummax, cummin, cumprod, cumsum	累積最大、累積最小、累積積、累積和
diff	差分の計算
max, min	最大、最小
pmax, pmin	部分のベクトルの最大と最小
mean	算術平均
median	中央値
prod	ベクトルの要素を積算する
quantile	標本確率点を計算する
range	ベクトルの最小と最大を返す
sample	ベクトルのランダムサンプリングまたは並べかえ
sum	ベクトルの和
summary	オブジェクトを要約する
var	分散と共分散

要約関数は、様々なタイプのデータに適したサマリーを提供する総称関数です。たとえば、線形モデルをあてはめることによって作成されたクラス `lm` のオブジェクトでは、返されるサマリーには推定した係数とその標準誤差および `t` 値の表が他の情報と共に含まれます。標準的なベクトルのサマリーは、次のような最小/最大/平均/中央値および第 1 位と第 3 位の四分位値の 6 つの値です。

```
> summary(stack.loss)
  Min. 1st Qu. Median Mean 3rd Qu. Max.
    7      11      15 17.52      19     42
```

仮説検定

S-PLUS には、表 10.7 に示すように、古典的な仮説検定を行う多くの関数があります。

表 10.7 : 仮説検定用の S-PLUS 関数

検定	説明
<code>t.test</code>	スチューデントの 1 標本 <code>t</code> 検定または 2 標本 <code>t</code> 検定
<code>wilcox.test</code>	ウィルコクソン順位和検定および符号付き順位和検定
<code>chisq.test</code>	2 次元分割表のピアソンのカイ二乗検定
<code>var.test</code>	2 つの分散を比較する <code>F</code> 検定
<code>kruskal.test</code>	クラスカル=ワリス順位和検定
<code>fisher.test</code>	2 次元分割表のフィッシャーの正確確率検定
<code>binom.test</code>	二項検定
<code>friedman.test</code>	フリードマン順位和検定
<code>mcnemar.test</code>	マクネマーのカイ二乗検定
<code>prop.test</code>	比率検定
<code>cor.test</code>	零相関の検定 (ピアソン、ケンドール、スピアマン)
<code>mantelhaen.test</code>	マンテル=ヘンツェルのカイ二乗検定

第 10 章 コマンド・ウィンドウの使い方

次の例は `t.test` を使って 2 標本 t 検定を実行し、平均の差を検出する方法を示します。この例は、分布 $N(0,1)$ と $N(1,1)$ から生成された 2 つのランダムサンプルを使用します。乱数シードを関数 `set.seed` でセットするため、この例は次のように再現することができます。

```
> set.seed(19)
> x <- rnorm(10)
> y <- rnorm(5, mean=1)
> t.test(x,y)
Standard Two-Sample t-Test

data: x and y
t = -1.4312, df = 13, p-value = 0.176
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
-1.7254080 0.3502894
sample estimates:
mean of x mean of y
-0.4269014 0.2606579
```


統計モデル

S-PLUS の統計モデリング関数のほとんどは、入力データはデータフレーム、あてはめるモデルはモデル式として指定されることを想定しています。モデル式を個別の S-PLUS オブジェクトとして保存し、モデリング関数に引数として指定することができます。

表 10.8 に、S-PLUS モデリング関数の一部のリストを示します。

表 10.8 : S-PLUS モデリング関数

関数	説明
aov, manova	分散分析モデル
lm	線形モデル (回帰)
glm	一般化線形モデル (ロジスティック回帰とポアソン回帰を含む)
gam	一般化加法モデル
loess	局所回帰モデル
tree	分類および回帰樹形モデル
nls, ms	非線形モデル
lme, nlme	混合効果モデル
factanal	因子分析
princomp	主成分分析
pam, fanny, diana, agnes, daisy, clara	クラスター分析

モデル式では最初に反応変量を指定し、次にティルド (~) とモデルの予想変量とする項を指定します。モデル式の変量は、数値ベクトル／因子または順序付き因子／行列を評価する任意の式です。表 10.9 は、モデル式のシンタックスをまとめたものです。

表 10.9 : S-PLUS モデル式シンタックスのまとめ

式	意味
A ~ B	A を B でモデル化する
B + C	モデルに B と C の両方を含める
B - C	モデルに B から C を除いて含める
B:C	B と C の間の交互作用
B*C	モデルに B と C およびその交互作用を含む
C %in% B	C が B にネストにされる
B/C	モデルに B および C %in% B を含む

次の S-PLUS セッションの例は、60 台の車に関する 5 つの変量を含む fuel.frame データに回帰モデルをあてはめるステップを示します。出力は示しません。したがって、これらのコマンドをコマンド・ウィンドウに入力して、S-PLUS 言語でデータ解析を行うのに慣れてください。

```

> names(fuel.frame)
> par(mfrow=c(3,2))
> plot(fuel.frame)
> pairs(fuel.frame)
> attach(fuel.frame)
> par(mfrow=c(2,1))
> scatter.smooth(Mileage ~ Weight)
> scatter.smooth(Fuel ~ Weight)
> lm.fit1 <- lm(Fuel ~ Weight)
> lm.fit1
> names(lm.fit1)
> summary(lm.fit1)
> qqnorm(residuals(lm.fit1))
> plot(lm.influence(lm.fit1)$hat, type="h",
+ xlab = "Case Number", ylab = "Hat Matrix Diagonal")
> o.type <- ordered(Type, c("Small", "Sporty", "Compact",
+ "Medium", "Large", "Van"))
> par(mfrow=c(1,1))

```

```
> coplot(Fuel ~ Weight | o.type,  
+ given.values=sort(unique(o.type)))  
> lm.fit2 <- update(lm.fit1, . ~ . + Type)  
> lm.fit3 <- update(lm.fit2, . ~ . + Weight:Type)  
> anova(lm.fit1, lm.fit2, lm.fit3)  
> summary(lm.fit3)
```

定義関数

S-PLUS は、大きくかつ複雑なシステムの設計に使用できる強力なプログラム言語です。他のプログラム言語と同じように、S-PLUS 言語を修得するにつれてその機能を使いこなすことができるようになります。しかしほとんどのプログラム言語と違って、S-PLUS ではその機能の多くをすぐに使用することができます。この章では、データ生成、データ操作、および統計処理のための様々な関数について見てきました。これらの関数をいくつか組み合わせたものを繰り返し使用し、ときには長い引数のリストを何度も入力していることに気づくことでしょう。繰返シタスクを 1 つの関数に取り込むことによって、生産性を高め、入力上の間違いを防ぐことができます。

新しい関数を定義するときは、次の形式の式を入力します。

```
newfunction <- function(arguments){
  body of definition
}
```

ここで `newfunction` は新しい関数のために選択した名前であり、`arguments` は引数の名前（ある場合）、`body of definition` はセミコロンの復帰改行で分けられた 1 つまたは複数の有効な S-PLUS 式を含みます。

たとえば、ボランティアの報告者から気象情報を得ていると仮定します。各報告者は、ひと月に 1 度自分の町の毎日の最高気温と最低気温のリストを送ってきます。報告者はボランティアなので、そのリストは観測の仕方や観測値の記録が完璧ではありません。したがって、ほとんどのリストは欠損値を含みます。欠損値を含む彼らの観測値を値 `NA` を使ってすべて根気よく入力し、次にデータを解析したいと思います。関数 `mean` は、それぞれの場所の平均気温を計算する前に、`NA` を削除する引数 `na.rm` を持つことを知っていますが、平均を求めたい場所は 20 もあり、これは大変な作業です。

以下の関数は、場所ごとに 1 つの変数を含むデータセットを引数 `location` として入力することを想定しています。実行すると、この関数はそれぞれ場所の平均気温を返します。

```
temp.means <- function(location)
{
  apply(location, 2, mean, na.rm=T)
}
```

ここで関数 `apply` は、引数 `location` として提供されるデータセットの列に関数 `mean` を「適用」します。（呼び出しで“2”ではなく“1”を使用した場合は関数は行に適用されます。）`apply` の最後の引数は、`mean` に与えたい引数、この場合は `na.rm` です。

バッチモードでの S-PLUS の使用

複雑な解析を行う関数を作成してそれがうまく働くことを確認した後で、その関数をきわめて大きいデータセットに使用したい場合があります。しかし、大きいデータセットで複雑な解析を行うには長い時間がかかります。Windows と Unix のバッチコマンドの使用についての詳しい情報は、『Application Developer's Guide』の第 22 章、「Verbose Logging」にあります。

バッチモードを使用すると、かかりきりにならず大量の計算を行うことができます。BATCH コマンドには、コマンドプロンプト (DOS プロンプトと呼ばれることもあります) から以下の形のように実行します。

```
Plus SBATCH [-flags ] inputfile
```

ここで、inputfile は、バッチジョブで実行されるプログラムから成る入力ファイルです。コマンド SBATCH は大文字で指定してください。

BATCH コマンドを実行するには、次のいずれかの操作を行います。

- Windows のスタートメニューからファイル名を指定して実行を選択し、ダイアログで S-PLUS BATCH コマンドラインを入力し、OK をクリックします。
- DOS プロンプトから、S-PLUS BATCH コマンドラインを入力し、ENTER を押します。

たとえば、時間の経過によりがん治療が白血球数に及ぼす影響を調査するため、毎月 14 の各病院から ASCII ファイルの形のレポートを受け取ると仮定します。ASCII ファイルには、調査している患者ごとに、名前、年齢、性別、治療の種類、白血球数などの情報が含まれています。レポートの 1 つを読み込み、リストにある各患者のデータを更新する関数 (update.data) を作成しました。毎月の各レポートを読み込むとき、毎月ちょうど 14 回その関数を使用します。次のものを含む入力ファイル update.dat を作成すると、長期的に時間の節約になりまた便利になります。

```
update.data("hospital.1")  
update.data("hospital.2")  
...  
update.data("hospital.14")
```

第 10 章 コマンド・ウィンドウの使い方

次に、毎月、各病院のレポートを適切な番号を付けたファイル (**hospital.x**) にコピーします。すべてのレポートが含まれているとき、次のように入力します。

```
Spplus SBATCH -logfile update.slg update.dat
```

コマンドは、ファイル **update.dat** (入力された最後の引数) から実行するプログラムを読み込み、出力とエラーをログファイル **update.slg** に書き込みます。

エラーメッセージを別のファイルに保存したい場合は、エラーファイルを指定することができます。たとえば、ファイル **myerrors.dat** に更新ジョブからのエラーを保存するには、次のコマンドを使用します。

```
Spplus SBATCH -logfile myerrors.dat -output update.txt  
update.dat
```

inputfile として次のように *stdin* を指定すると、キーボードから直接バッチコマンドを入力することができます。

```
Spplus SBATCH outputfile inputfile
```

また、*stdout* または *stderr* を *outputfile* または *errorfile* として指定することによって、バッチ出力またはエラーがファイルに保存されないようにすることができます。たとえば、ファイル **myerrors.dat** にエラーを保存しているときに最新の更新ジョブの出力を無視するときは、次の **BATCH** コマンドを使用します。

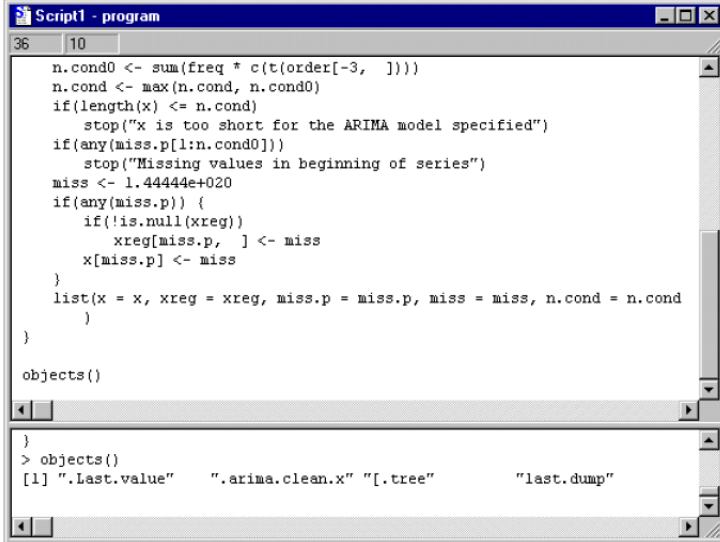
```
Spplus SBATCH update.dat stdout myerrors.dat
```

第 11 章 スクリプト・ウィンドウ とレポート・ウィンドウ の使い方

はじめに	536
スクリプト・ウィンドウ	538
スクリプトを使った作業	538
検索と置換の使い方	543
文脈依存ヘルプ	545
スクリプト・ウィンドウの機能	546
区切り記号の自動整合	546
右波かっこの自動挿入	546
自動インデント	546
スクリプト・ウィンドウの設定の変更	547
スクリプト使用のための時間節約のヒント	549
履歴ログ	549
スクリプト・ウィンドウへのグラフオブジェクトのドラッグ	552
スクリプト・ウィンドウへの関数オブジェクトのドラッグ	553
レポート・ウィンドウ	554
スクリプトまたはレポートの印刷	556

はじめに

スクリプト・ウィンドウでは、データ解析とグラフ作成の繰り返し行う部分を自動化するためにスクリプト（プログラム）を作ることができます。また、スクリプト・ウィンドウを使ってスクリプトを編集します。それぞれのスクリプト・ウィンドウには、実行中のスクリプトからの出力を表示する出力パネルと、スクリプト作成のコマンド入力に使用されるプログラムパネルがあります。スクリプトによって S-PLUS プログラム言語を利用することができます。たとえば、データをインポート／エクスポート／変換し、解析を実行して、グラフを生成／修正／表示するために実行するコマンドを書くことができます。



```


Script1 - program
36 10
n.cond0 <- sum(freq * c(t(order[-3, ])))
n.cond <- max(n.cond, n.cond0)
if(length(x) <= n.cond)
  stop("x is too short for the ARIMA model specified")
if(any(miss.p[1:n.cond0]))
  stop("Missing values in beginning of series")
miss <- 1.44444e+020
if(any(miss.p)) {
  if(!is.null(xreg))
    xreg[miss.p, ] <- miss
  x[miss.p] <- miss
}
list(x = x, xreg = xreg, miss.p = miss.p, miss = miss, n.cond = n.cond
)
}
objects()

}
> objects()
[1] ".Last.value" ".arima.clean.x" "[.tree" "last.dump"

```

図 11.1 : プログラムパネル（上）と出力パネル（下）を示すスクリプト・ウィンドウ

スクリプトは S-PLUS 内から、または別のアプリケーションから（DDE を介して、あるいは S-PLUS を呼び出してコマンドラインでスクリプト名を渡すことによって）実行することができます。

S-PLUS 言語を使用するときは、**コマンド・ウィンドウ**の代わりに**スクリプト・ウィンドウ**を使用することもできます。**コマンド・ウィンドウ**は対話式で、**コマンド・ウィンドウ**に入力されたコマンドは、インタプリタによってすぐに評価され、結果がすぐ下の行に出力されます。一方**スクリプト・ウィンドウ**では 1 組のコマンドと関数を入力し、それを必要なときだけ評価することができます。スクリプトは、**スクリプト・ウィンドウ・ツールバー**の**実行**ボタン  をクリックすることにより実行することができます。スクリプトの一部分が選択されている（ハイライトされている）場合は、その選択部分だけが実行されます。出力は、各コマンドの下ではなく出力パネルに表示されます。**コマンド・ウィンドウ**は対話式に探索的データ解析を行うのに適しており、**スクリプト・ウィンドウ**は長い関数を記述するのに適しています。

スクリプト・ウィンドウ

スクリプト・ウィンドウに実行可能なステートメントまたはコマンドを入力して実行することができます。たとえばスクリプト・ウィンドウに、次の S-PLUS 式を入力することができます。

```
objects( )
```

スクリプト・ウィンドウには必ず出力パネルがあり、この出力パネルには印刷ステートメントによる出力、スクリプトの実行時に生じたワーニング／エラーメッセージに関する情報が表示されます。スクリプト・ウィンドウのプログラムパネルには、編集集中にスクリプト内の行を探すのに便利な行列番号インジケータがウィンドウの左上にあります。




図 11.2 : スクリプト・ウィンドウ・ツールバー

スクリプトを使った作業

スクリプトを作成するかまたは開いて、編集／実行／保存／印刷することができます。

新しいスクリプトを作成するには、以下の操作を行います。

1. メインメニューから**ファイル** ▶ **新規ファイル**を選択するか、**標準ツールバー**の**新規ファイル**ボタン  をクリックします。ウィンドウタイプのリストがポップアップします。
2. **Script File** を選択し、**OK** をクリックします。

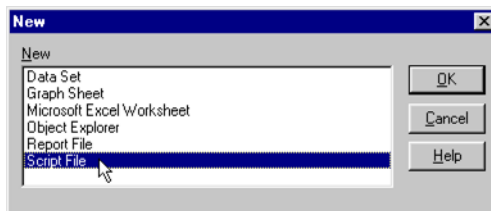


図 11.3 : New ダイアログでスクリプトを含む多くのファイルタイプが作成可能


新しいスクリプトファイルが作成され、ウィンドウ内に表示されます。新しいスクリプトには、一時的なデフォルト名が付けられています。

S-PLUS 言語のコマンドと式を使って、スクリプト・ウィンドウのプログラムパネルにコマンドを直接入力することができます。スクリプト・ウィンドウは、デフォルトでコマンドを S-PLUS インタープリタに送ります。

スクリプト・ウィンドウの上パネル内をマウスでクリックすると、ウィンドウのキャプション（またはタイトル）が、スクリプト名 **program** に変化します。このパネル内で入力すると、行列番号インジケータが変化して、編集中の場所を示します。

下側パネルは、スクリプトの出力に使用されます。通常はスクリプトを実行すると、関数 `print` による出力、スクリプトの関数呼び出し、ワーニングやエラーがこの出力パネルに表示されます（これは**オプション ▶ テキスト出力オプション**によって変更することができます）。このパネルをマウスでクリックすると、スクリプト・ウィンドウのキャプションがスクリプト名 **output** に変化します。出力パネルからクリップボードにテキストをコピーすることができますが、テキストを入力することはできません。

既存のスクリプト・ファイルを開くには、以下のいずれかの操作を行います。

- 標準ツールバーの**開く**ボタン  をクリックします。
- メインメニューから**ファイル ▶ 開く**を選択します。
- **Open** ダイアログで、ファイルの種類として S-PLUS スクリプト・ファイル (***.ssc; *.q**) を選択します。希望のフォルダに移動し、スクリプト・ファイルを選択し、**開く**をクリックして新しいスクリプトウィンドウでスクリプトを開きます。

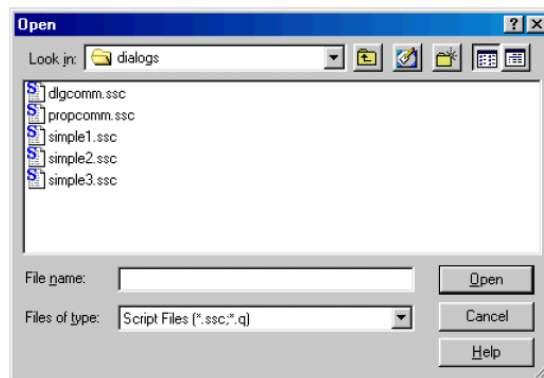




図 11.4 : スクリプトファイルを開く


スクリプト・ウィンドウからスクリプトを実行する

実行ボタン  またはスクリプトの**実行**メニューオプションを使って、スクリプトを実行することができます。

スクリプトを実行するには、以下のいずれかの操作を行います。


- スクリプト・ウィンドウ・ツールバーの**実行**ボタン  をクリックします。
- メインメニューから**スクリプト ▶ 実行**を選択します。

スクリプトの一部分を実行するには、以下の操作を行います。

1. スクリプト内の実行したい行を選択します。
2. メインメニューから**スクリプト ▶ 実行**を選択するか、**スクリプト・ウィンドウ・ツールバーの**実行**ボタン ** をクリックします。

スクリプトファイルを実行しているとき、**スクリプト・ウィンドウのタイトル**がスクリプト名**-running**に変化します。スクリプトが止まるか実行が終了すると、タイトルはスクリプト名**-program**に戻ります。

スクリプトをスクリプトファイルに保存するには、以下の操作を行います。

1. **標準**ツールバーの**保存**ボタン  をクリックするか、または**ファイル**メニューから**保存**または**名前を付けて保存**のメニュー項目を選択するか、**CTRL-S**を選択してください。
2. **名前を付けて保存**を選択した場合、または**スクリプト・ウィンドウ**がまだ保存されていない場合には、ブラウザのウィンドウが現れます。
3. 希望するフォルダに移動し、**ファイル名**フィールドをたとえば**savetrees.ssc**という希望のファイル名に変更します。
4. **保存**をクリックして、**savetrees.ssc**と名前をつけた新しいスクリプトファイルを作成します。

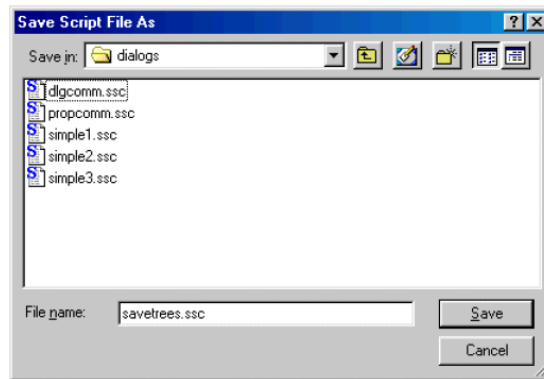



図 11.5 : スクリプトファイルを保存する

スクリプト ファイルの印刷

スクリプト・ウィンドウの内容を印刷するには、以下の操作を行います。

1. 印刷するスクリプト・ウィンドウを選択します。
2. ファイルメニューからスクリプト印刷を選択します。
3. 通常の印刷ダイアログを使って印刷オプションを指定し、**OK** をクリックします（このとき表示される実際の印刷ダイアログは、使用しているプラットフォームとデフォルトのプリンタにより異なります）。

また、以下のようにしてスクリプトを印刷することもできます。

1. 標準ツールバーの印刷ボタン  をクリックします。
2. 印刷したいスクリプト・ウィンドウを確認するダイアログがポップアップします。



3. （デフォルトのプリンタとデフォルトの設定で）印刷するときには**はい (Y)** をクリックし、印刷を中止するときには**いいえ (N)** をクリックします。

スクリプトの 中止

スクリプトまたはスクリプトの選択部分を実行中に、ESC キーを使ってスクリプトの実行を中止することができます。これで、スクリプトがそれ以上実行されるのを防ぐことができます。

エラーと ワーニングの 解釈

インタープリタがスクリプトファイルに入力された式やコマンドの問題点に遭遇すると、スクリプト・ウィンドウの出力パネルにエラーまたはワーニングが表示されます。そのワーニングまたはエラーメッセージは、問題点と、場合によってはその問題に対して考えられる原因を提示します。次にその行に移動してスクリプトを編集し、その問題を修正した後でスクリプトを再び実行することができます。

ワーニングは、エラーほど重大なものではありません。一般にワーニングでスクリプトの実行は中止されませんが、エラーでは中止されます。

スクリプト・ ウィンドウ内の テキストの選択

編集メニューから**すべて選択**を選択するか CTRL-A を押すことによって、スクリプト・ウィンドウ内のすべてのテキストを選択することができます。

スクリプト・ ウィンドウ内の テキストの クリア/ 切り取り/ コピー/ 貼り付け




編集メニューの切り取り、コピーおよび貼り付けコマンドを使用するか、標準ツールバーの切り取りボタン、コピーボタン、貼り付けボタンを使用するか、CTRL-X、CTRL-C および CTRL-V を使用することによって、スクリプト・ウィンドウ内のテキストを移動させることができます。これらのコマンドを使用すると、同じスクリプト・ウィンドウ内、開いている別のスクリプト・ウィンドウ内、または S-PLUS やその他のアプリケーション間で移動/コピーすることができます。

切り取りまたはコピーを行ったテキストは、クリップボードに入れられます。クリップボードに入れられたものは、**切り取り**または**コピー**コマンドが選択されるまでそこに置かれます。そのテキストは、クリップボードからスクリプト・ウィンドウに必要な回数だけ貼り付けることができます。

テキストを移動/コピーする方法と同じ方法で、項目または文字を移動/コピーすることができます。


編集メニューの**クリア**コマンド（または DELETE キー）を使用すると、クリップボードにテキストのコピーを保存することなく、スクリプト・ウィンドウからテキストを削除することができます。

スクリプト内のテキストを移動またはコピーするには、以下の操作を行います。

1. テキストを選択します。
2. 標準ツールバーの切り取りボタン  またはコピーボタン  をクリックするか、編集メニューから切り取り (CTRL-X) またはコピー (CTRL-C) を選択します。これで、テキストがクリップボードに入ります。
3. スクリプト・ウィンドウ内の新しい場所に挿入点を位置決めします。標準ツールバーの貼り付けボタン  をクリックするか、編集メニューから貼り付け (CTRL-V) を選択します。

スクリプト・ウィンドウで元に戻すを使用

スクリプト・ウィンドウには、グラフシートやデータ・ウィンドウで作業しているときに使用される元に戻すとは別に、それ自体の元に戻す機能があります。スクリプトの編集中に、グラフシートやデータオブジェクトの操作を元に戻したりやり直したりすることはできません。スクリプトの編集中はメニューの編集 ▶ 元に戻すを選択することによって、入力中の変更を元に戻すことができます。スクリプト・ウィンドウを終了するとすぐに、グラフシートまたはデータ・ウィンドウの元に戻すのキューが回復します。

スクリプト・ウィンドウで最後に行った変更を取り消すには、標準ツールバーの元に戻すボタン  をクリックするか、編集メニューから元に戻すを選択するか、CTRL-Z を押します。


最後に行った変更が元に戻されます。スクリプト・ウィンドウを元に戻す前の状態に戻す必要がある場合は、もう 1 度元に戻すを実行して元に戻した変更を復元することができます。

検索と置換の使い方

スクリプト・ウィンドウ内のテキストを検索したり変更したりするときは、検索または置換オプションを使用します。検索を使用すると、スクリプト内にある特定のテキストを探することができます。置換を使用すると、スクリプト内で特定のテキストを探して置き換えることができます。検索と置換は、ある単語や句、またはコマンド全体のような一連の文字に使用することができます。

S-PLUS はスクリプトの一部を選択しない限り、スクリプト全体にわたって指定したテキストを置換します。その結果が気に入らない場合には、変更を保存せずにスクリプト・ウィンドウを閉じることができるように、置換を使用する前にスクリプトを保存しておくことをお勧めします。また元に戻すを使用して、スクリプトに最後に行った置換を取り消すことができます。

テキストを検索するには、以下の操作を行います。

1. スクリプト・ウィンドウ・ツールバーの**検索**ボタン  をクリックするか、**編集**メニューから**検索**を選択するか、CTRL-F を押します。**検索**ダイアログがポップアップします。
2. **検索する文字列**ボックスに、検索するテキストを入力します。

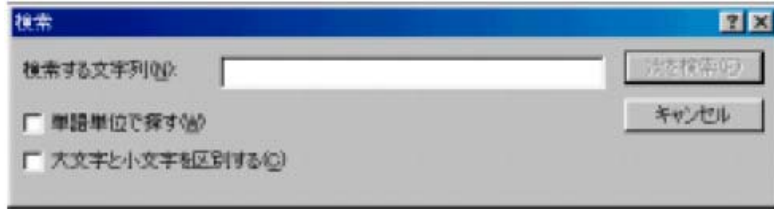


図 11.6 : 検索ダイアログは最大 122 文字までの文字列を検索する ; 入力するときテキストが横にスクロールする

現在の作業セッション中に**検索**または**置換**を使用した場合は、**検索する文字列**ボックスに最後に検索したテキストが選択されています。違うテキストを検索するときは、そのテキストに上書きしてください。

3. **次を検索**を選択すると検索を始めます。

検索ダイアログには、表 11.1 のオプションがあります。

表 11.1 : 検索および置換ダイアログのチェックボックスオプション

オプション	目的
単語単位で探す (W)	このオプションを選択すると、部分的な文字列ではなく単語全体を検索。
大文字と小文字を区別する (C)	このオプションを選択すると、指定したパターンの英大小文字を含む単語だけを検索。

テキストを検索し置換するには、以下の操作を行います。

1. **編集**メニューから**置換**を選択するか、CTRL-H を押します。**置換**ダイアログがポップアップします。
2. **検索する文字列**ボックスに、検索するテキストを入力します。
3. 現在の作業セッション中に**検索**または**置換**を使用した場合は、**検索する文字列**ボックスには最後に検索したテキストが選択されています。違うテキストを検索するときは、そのテキストに上書きしてください。

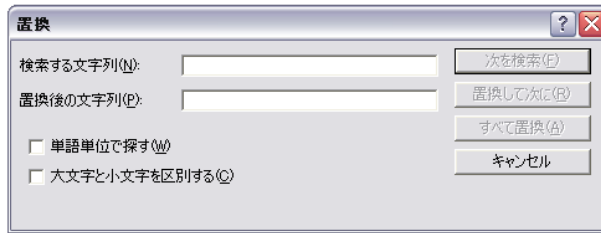


図 11.7 : 置換ダイアログのテキストの長さの制限は検索ダイアログと同じ

4. **置換後の文字列**ボックスに、置き換えるテキストを入力してください。

検索する文字列ボックスと同じように、現在の作業セッション中に**置換後の文字列**を使用する場合は、最後に指定した置換文字が**置換後の文字列**ボックスに選択されています。違う置換文字を指定するときは、そのテキストに上書きしてください。

- **次を検索**を選択すると、**検索する文字列**内の単語が次にある場所にカーソルが移動します。
- **置換して次に**を選択すると、**検索する文字列**で指定して見つかった単語を**置換後の文字列**の単語と置き換えます。
- **すべて置換**を選択すると、**検索する文字列**で指定して見つかったすべての単語を**置換後の文字列**の単語と置き換えます。確認ダイアログはありません。

また、**置換**オプションでテキストを削除することもできます。その場合は、上記のステップにしたがい、**置換後の文字列**ボックスを空白のままにしてください。

文脈依存ヘルプ

カーソルが、スクリプト・ウィンドウ内の単語の最初、中間または最後にある場合に、F1 キーを押すとその単語のヘルプがポップアップします。具体的には、単語が S-PLUS 関数の名前の場合には、その関数のヘルプが表示されます。

スクリプト・ウィンドウの機能

スクリプト・ウィンドウには、S-PLUS 関数の入力をもっと簡単にするように設計されたいくつかの機能があります。各機能は、他とは独立して利用可能にしたり利用不能にしたりすることができます。

区切り記号の自動整合

スクリプト・ウィンドウは、波かっこ { }、丸かっこ ()、角かっこ [] および一重引用符 ' ' と二重引用符 " " を自動的に整合させます。たとえば右の丸かっこを入力すると、エディタが自動的に対になる左の丸かっこをハイライトします。この動作は、波かっこ、角かっこおよび引用符でも同じです。区切り記号の整合機能によって、思い通りに整合させることができます。

たとえば右丸かっこを入力した後、カーソルは対になる左丸かっこまで自動的に移動し、それを所定の長さの時間（デフォルトでは 0.5 秒すなわち 500 ミリ秒）だけハイライトします。次にカーソルは、右丸かっこの後のスペースに移動します。片方の丸かっこがハイライトされている間に入力を続けた場合、キーストロークが失われないようにその間のキーストロークはバッファに保存されます。ハイライトする時間の長さは変更することができます。

デフォルトでは、スクリプト・ウィンドウはスクリプト全体を探して自動整合を見つけます。大きいスクリプトでは時間がとてもかかることがあるため、探索を指定した文字数に制限することができます。

右波かっこの自動挿入

右波かっこの自動挿入を利用可能にすると、左波かっこを入力した後で ENTER を押すことにより、2 行下にもう片方の右波かっこが自動的に挿入され、カーソルがその間の行に移動します。

自動インデント

自動インデントを利用可能にすると、エディタが関数定義 / if 文 / for 文 / while 文の本文を自動的にインデントします。デフォルトではインデントの量はスペース 4 つですが、この値は変更することができます。

次の関数の例は、サポートされるインデントスタイルを示します。

```
"test1"<-
function(x)
{
  if(x > 0) {
    for(i in 1:x) {
      cat(i, "¥n")
    }
  }
  else {
    i <- x
    while(i > 0) {
      cat(i, "¥n")
      i <- i - 1
    }
  }
}
```

スクリプト・ウィンドウの設定の変更

スクリプト・ウィンドウのデフォルト設定値は、スクリプト・ウィンドウを右クリックし、ポップアップメニューから**プロパティ**を選択して表示させた**Script** ダイアログで変更することができます。

次のプロパティのいずれかを利用不能にするときは、対応するチェックボックスのチェックを外してください。

- **Output Pane Word Wrap**
- **Auto Match : {, (), [], " and ''**
- **Auto Indent**
- **Auto Insert Right Brace**

Tab Size を変更するときは、対応するテキストボックスに希望のスペースの数を入力してください。

対応する区切り記号をハイライトする時間の長さを変更するときは、**Match Time (msec)** テキストボックスの値 (ミリ秒で示した) を変更してください。

自動整合で探す文字の数を変更するときは、**Match CharLimit** テキストボックスに値を入力してください。デフォルト値**1** では、カーソルからファイルの先頭まで探します。

第 11 章 スクリプト・ウィンドウとレポート・ウィンドウの使い方

またスクリプト・ウィンドウのプロパティは、**オブジェクト・エクスプローラ**で変更することもできます。これを行うときは、右パネル内の適切なスクリプトを右クリックし、ポップアップメニューから **Properties** を選択します。

次のスクリプト・ウィンドウのセッションのために希望の設定値をデフォルトとして保存するときは、**オプション ▶ ウィンドウサイズ・プロパティ**の**今の状態をデフォルト値にする**を選択します。

スクリプト使用のための時間節約のヒント

S-PLUS は、スクリプトを記述するいくつかの方法を提供します。最も簡単な方法は、新しいスクリプト・ウィンドウを開き、コマンドを入力して実行することです。スクリプトを生成する他の方法には、履歴ログとメニューを使用したり、スクリプト・ウィンドウにオブジェクトをドラッグしてオブジェクトを作成または修正するコマンドを記録する方法があります。この節では、履歴ログを表示する方法や S-PLUS コマンドを使って所定のプロットを生成する方法、およびスクリプト・ウィンドウを使って S-PLUS 関数定義を編集する方法について説明します。ツールバー/メニュー項目/ClassInfo オブジェクト/FunctionInfo オブジェクトを含む他のオブジェクトタイプを、オブジェクト・エクスプローラからスクリプト・ウィンドウにドラッグ アンド ドロップすることもできます。

履歴ログ

S-PLUS は、メニュー/ツールバー/ダイアログ操作の連続的なレコードすなわち履歴を保持しています。また、データ・ウィンドウ内のセルを変更したり、グラフシート上のオブジェクトの位置を変更したりするような表示上の編集は、コマンド・ウィンドウ内で発行されたコマンドと同じように記録されます。これらの操作に対応する S-PLUS プログラム言語は、履歴ログに記録されます。


履歴ログは、スクリプト・ウィンドウに表示させることができます。

ダイアログ操作を履歴ログに記録するときは、ダイアログで **OK** または **Apply** ボタンを使って変更を確定してください。Cancel を選択するかダイアログから **ESC** を押すと、そのダイアログに対応するコマンドは、履歴ログには記録されません。

他のスクリプトとまったく同じように履歴ログの行を編集することができます。そのような編集により履歴ログ自体は修正されず、このスクリプト・ウィンドウ内のコピーが修正されるだけです。スクリプトの一部を切り取って他のスクリプトに貼り付け、スクリプトの一部を実行したり、スクリプト全体を実行したり、スクリプトをファイルに保存することができます。

履歴ログの最大サイズ（記録された操作の合計数）は、オプションメニューの元に戻す+履歴から使用できる **Undo & History** ダイアログの **History Entries** フィールドで指定することができます。

スクリプト・ウィンドウに現在の履歴ログを表示させるには、以下の操作を行います。

1. デフォルト設定で履歴ログを表示するときは、標準ツールバーの履歴ログボタン  をクリックします。設定を変更して表示させるには、ウィンドウメニューから履歴次に表示を選択します。Display History Log ダイアログが現れます。

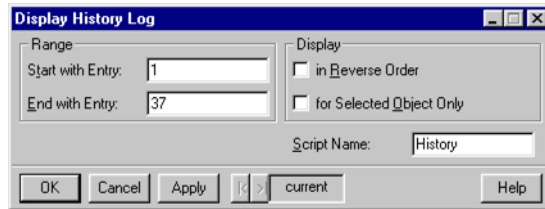



図 11.8 : Display History Log ダイアログ

2. 希望の表示オプションを指定します。
3. OK をクリックすると、履歴ログを表示します。

表 11.2 : Display History Log ダイアログのオプション

フィールド	説明
Start with Entry End with Entry	履歴ログに表示する最初と最後のエントリ番号を指定します。これによりエントリの数と、履歴ログに入れられるエントリを管理することができます。
Display in Reverse Order	生成されたのと逆の順序でエントリを表示するように選択します。スクリプトの一番上に最も最近実行されたコマンドが表示されます。
Display for Selected Object Only	スクリプトが選択されたオブジェクト専用にエントリを含むように選択します。これは、特定のオブジェクトのコマンドに注目したいときに便利です。たとえば記号を選択すると、スクリプトは記号の作成と修正に関連するすべてのエントリを含みます。
Script Name	履歴ログを含むスクリプトの名前を指定します。これは省略可能です。デフォルトのスクリプト名は、History です。

履歴ログ内の記録されたコマンドを実行するには、以下の操作を行います。

1. マウスを使って実行したいエントリをハイライトします。
2. スクリプト・ウィンドウ・ツールバーの**実行**ボタン  をクリックするか、**スクリプトメニュー**から**実行**を選択します。

またこれらのコマンドを切り取り、別のスクリプトファイルに貼り付け／保存することができます。

記録ステップを始める前に**履歴ログ**をクリアしてください。これによりその後の編集時間を節約でき、メニューとダイアログで行った操作によってどのコマンドが生成されたかがはっきりします。

履歴ログをクリアするには、**ウィンドウメニュー**から**履歴**を選択し、次にサブメニューから**クリア**を選択してください。

Condensed と Full

デフォルトとして**履歴ログ**は、必須入力とデフォルト値と異なるオプションだけを含む主コマンドによる圧縮形式で記述されます。デフォルトを含め、それぞれのコマンドとそのすべてのパラメータからなる完全な履歴を表示させることができます。たとえば完全な履歴は、特定のプロットの作成に使用した引数の正確な値が欲しい場合に有効です。**履歴ログ**を使用して、スクリプトからまたは**コマンド・ウィンドウ**で**S-PLUS** グラフィックスの作成方法を学習する場合は、圧縮形式の方が便利です。



履歴ログのタイプを選択するには、以下の操作を行います。

1. **メインメニュー**から**オプション** ▶ **元に戻す** + **履歴**を選択します。
2. **History Type** ドロップダウンリストから **Condensed** または **Full** を選択します。

スクリプト・ウィンドウへのグラフオブジェクトのドラッグ


スクリプトを作成するもう 1 つの方法は、プロットや特別な記号などのオブジェクトをグラフからスクリプト・ウィンドウにドラッグすることです。特定の編集可能なプロットを作成または変更するためにどの S-PLUS コマンドを使用するかを知りたい場合は、それをスクリプト・ウィンドウにドラッグすると、それを作成する S-PLUS コマンドが自動的に記述されます。次に、生成されたスクリプトを実行してプロットを作成したり変更したりすることができます。これは、作成や変更メニューのオプションやダイアログを使用する代わりにする方法です。

グラフオブジェクトをスクリプト・ウィンドウにドラッグするには、以下の操作を行います。

1. 標準ツールバーの**新規ファイル**ボタン  をクリックし、リストから **Graph Sheet** を選択します。
2. **図形描画**パレット (**Annotation**) を開き、塗りつぶした四角形ボタンをグラフシート上にドラッグします。
3. 標準ツールバーの**新規ファイル**ボタン  をクリックして、リストから **Script File** を選択し新しいスクリプト・ウィンドウを作成します。
4. スクリプト・ウィンドウとグラフシートを縦に並べた方が便利でしょう。これで、グラフシートとスクリプトの間でオブジェクトの選択やドラッグが容易になります。ウィンドウを並べるときは、**メインメニュー**から **ウィンドウ ▶ 縦に並べる** を選択します。
5. グラフから四角形を選択し、それをスクリプト・ウィンドウの上パネル (プログラムパネル) にドラッグします。


注意

マウスをドラッグしている間、カーソルは“ドロップ”カーソルに変化します。マウスカーソルがプログラムパネルの内側にあるとき、カーソルがある行の左端に灰色の縦のマーカ線が表示されます。これは、マウスをはなすとスクリプトにコマンドが挿入される場所を示しています。マウスをはなすと、コマンドはスクリプト・ウィンドウのオブジェクトをドロップした行に書き込まれます。

6. グラフシート上の四角形を削除します。
7. スクリプトで、FillColor を "Blue" に変更します。
8. スクリプトツールバーの**実行**ボタン  を押します。グラフシートに四角形が青色になって現れます。

編集可能なグラフィックスのプログラムに関する詳細は、『Application Developer's Guide』の第 1 章「Editable Graphics Commands」を参照してください。

スクリプト・ ウィンドウ への関数 オブジェクト のドラッグ

S-PLUS 関数オブジェクトをオブジェクト・エクスプローラ・ウィンドウからスクリプト・ウィンドウ上にドラッグする場合は、関数定義がスクリプト・ウィンドウに展開されます。これは、関数の本文を編集するのにとても便利なショートカットです。変更を試したい場合はスクリプトメニューの**実行**を選択するか、**実行ボタン**  をクリックすると、新しい関数定義が自動的にS-PLUS インタープリタに送られ、再定義されます。

レポート・ウィンドウ

レポート・ウィンドウは、スクリプト・ウィンドウと似ています。これらのウィンドウは両方とも基本的にテキストウィンドウで、ファイルメニューから開いたり保存することができるし、また編集も可能です。スクリプト・ウィンドウと違って、レポート・ウィンドウはプログラムやスクリプトを処理することができません。レポート・ウィンドウは、S-PLUS の操作によるテキスト出力のためのプレースホルダーです。(テキスト出力の基本設定としてレポート・ウィンドウを選択しなければなりません。このオプションを設定する方法のためのテキスト出力転送に関しては、下の節を参照してください。)

レポート・ウィンドウ内のテキストは編集可能です。図 11.9 に示すレポート・ウィンドウのツールバーにより、S-PLUS レポートに表示されているテキストのサイズ、フォント、およびスタイルを変更することができます。



図 11.9 : レポート・ウィンドウのツールバー

さらにレポート・ウィンドウは、切り取り／コピー／貼り付けなどの基本的な編集機能の他に、以下の操作をサポートしています。

- キーボードからの入力、ポイント アンド クリック、ハイライト ドラッグ アンド ドロップなど。
- 元に戻す／切り取り／コピー／貼り付け／検索／置換などは、編集メニューや文脈依存メニュー（右クリック）によってサポートされています。
- クリップボードからの貼り付け（グラフィックスやその他の OLE オブジェクトを含む）は、RTF ファイルだけに使用することができます。特に、レポート・ウィンドウをテキストファイルとして保存してそれにグラフィックイメージを貼り付けようとする、グラフィックスは保存されず、エラーメッセージも表示されません。
- フォントは、文脈依存メニュー（右クリック）と書式メニューによりサポートされます。フォントは、RTF モードのみで使用できます。
- レポート・ウィンドウでもキーボードによるユーザ入力を行うことができます（トリックリング入力）。

レポート・ウィンドウは、デフォルトではリッチテキスト形式 (**.rtf**) で保存されます。またレポートは、プレーンテキスト形式 (**.txt**) で保存したり、拡張子 **.srp** (S-PLUS の以前のバージョンで使用されていた拡張子) で保存したりすることもできます。


プレーンテキストは様々なプログラムで使用することができ、比較的高速です。RTF は多数の機能を持っていますが、RTF ファイルはプレーンテキストのものよりもサイズが大きく低速です。


レポート・ウィンドウを保存するときは、メインメニューから**ファイル ▶ 保存**または**名前を付けて保存**を選択します。

新しいレポート・ウィンドウを作成するときは、メインメニューから**ファイル ▶ 新規ファイル**を選択します。スクロールボックスが開きます。**Report File**を選択し **OK** をクリックします。新しいレポート・ウィンドウが現れます。

レポート・ウィンドウにファイルを開くときは、メインメニューから**ファイル ▶ 開く**を選択します。希望のファイルを選択して **OK** をクリックします。

スクリプトまたはレポートの印刷

S-PLUS のスクリプトまたはレポートを印刷するときは、Windows 標準の印刷ボタン  を使うか、ファイルメニューのスクリプト印刷またはレポートファイル印刷を選択してください。

印刷ボタンを使って印刷するときは、標準ツールバーの印刷ボタン  をクリックします。印刷オプションの確認を求めるダイアログが現われます。

印刷ダイアログを使って印刷する

1. メインメニューから **ファイル ▶ スクリプト印刷またはファイル ▶ レポートファイル印刷** を選択します。印刷ダイアログが現れます。
2. 印刷ダイアログで希望するオプションを選択します。オプションの説明は、オンラインヘルプを参照してください。
3. **OK** をクリックすると、印刷が始まります。

第 12 章 他のアプリケーション での S-PLUS の使用

Microsoft Excel での S-PLUS の使用	558
Excel と S-PLUS 間のデータのリンク	558
Excel アドインアプリケーションの使用	569
SPSS での S-PLUS の使用	577
MathSoft Mathcad での S-PLUS の使用	583
Microsoft PowerPoint での S-PLUS の使用	590

Microsoft Excel での S-PLUS の使用

Microsoft Excel で S-PLUS を使用するには次の 2 つの方法があります。

- Excel Link Wizards を使って、Excel ワークシートと S-PLUS データフレーム間で「リンク」を確立することにより、S-PLUS 内から Excel ワークシートに保存されたデータをプロットしたり解析したりすることができます。
- Excel アドインアプリケーションを使って、Excel 内から S-PLUS のグラフを作成したり修正したりすることができます。


Excel と S-PLUS 間のデータのリンク

ご使用のコンピュータに Microsoft Excel 97 以上がインストールされている場合は、S-PLUS 内から Excel ワークシートを作成したり、開いたり、保存したりすることができます。


また、Excel 内の特定のセルの範囲と S-PLUS 内のデータフレームの間にリンクを作成することによって、Excel ワークシートに保存されたデータを使って、S-PLUS でグラフを作成したり統計解析を行ったりすることができます。

Excel ワークシートを作成するかまたは開く

S-PLUS 内から Excel ワークシートを作成するときは、以下の操作を行います。

1. 標準ツールバーの**新規ファイル**ボタン  をクリックするか、メインメニューから**ファイル ▶ 新規ファイル**を選択します。
2. **New** ダイアログで、**Microsoft Excel** ワークシートを選択し、**OK** をクリックします。

S-PLUS 内から Excel ワークシートを開くときは、以下の操作を行います。

1. 標準ツールバーの**開く**ボタン  をクリックするか、メインメニューから**ファイル ▶ 開く**を選択します。
2. **Open** ダイアログで**ファイルの種類**ドロップダウンリストから **Excel WorkSheets (*.xls)** を選択し、希望する Excel ファイルに移動し、**開く**をクリックします。

Excel ワークシートを作成したり開いたりするときは、図 12.1 に示すように、S-PLUS 内のウィンドウにワークシートが表示されます。このウィンドウは、S-PLUS に埋め込まれた Excel であるため、Excel で使用可能なすべての機能を使用することができます。

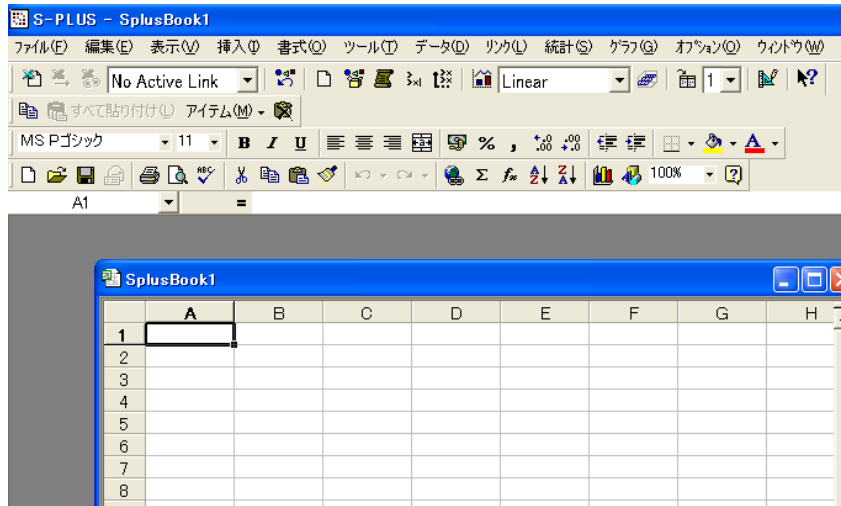


図 12.1 : S-PLUS 内に開かれた Excel ワークシートの例

Excel ワークシートが表示されているときは、S-PLUS の標準ツールバーが **Excel Sheet** ツールバーに変化します。このツールバーには、図 12.2 に示すように、標準ツールバーで使用できるほとんどのボタンと Excel 専用の新しいセクションがあります。新しい **Excel Sheet** ツールバーの下に Excel 自身の一般的なツールバーが現れます。Excel によって表示、挿入、書式、ツール、データメニューが表示されることにも注意してください。

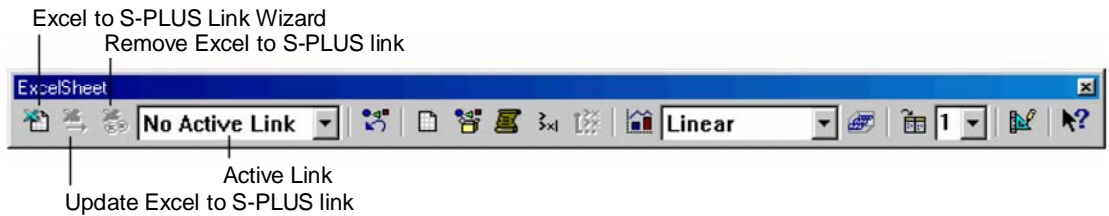



図 12.2 : Excel Sheet ツールバー

Excel to S-PLUS Link Wizard の 使い方

S-PLUS を使って Excel ワークシートに保存されたデータをプロットしたり解析したりする前に、まずデータを S-PLUS データフレームにしなければなりません。これは、**Excel to S-PLUS Link Wizard** を使って、Excel ワークシート内の領域から S-PLUS 内のデータフレームにリンクを確立することにより行います。

Excel から S-PLUS へのリンクを作成するには、以下の操作を行います。

1. S-PLUS でプロットまたは分析したい Excel ワークシート内の領域を選択します。この領域は、データだけを含んだり、列ラベル/行ラベルを含むことができます。
2. **Excel Sheet** ツールバーの **Excel to S-PLUS Link Wizard** ボタン  をクリックするか、メインメニューからリンク ► リンクウィザードを選択します。図 12.3 に示すような **Excel to S-PLUS Link Wizard** が現れます。

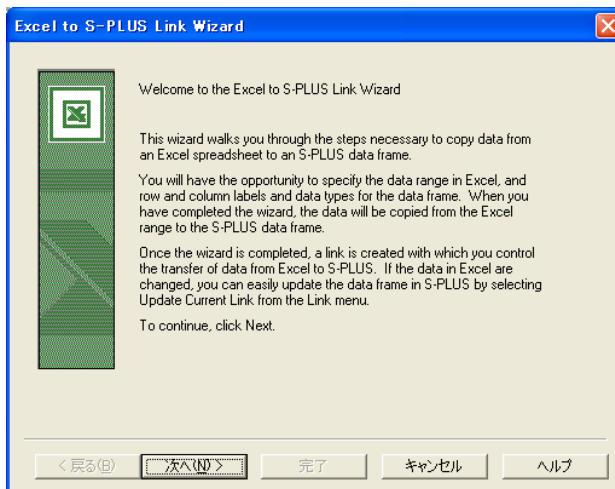


図 12.3 : Excel to S-PLUS Link Wizard

3. **次へ** をクリックします。
4. 図 12.4 で分かるように、ウィザードを起動する前に Excel ワークシート内で選択した範囲が **Data Range** フィールドに自動的に表示されます。

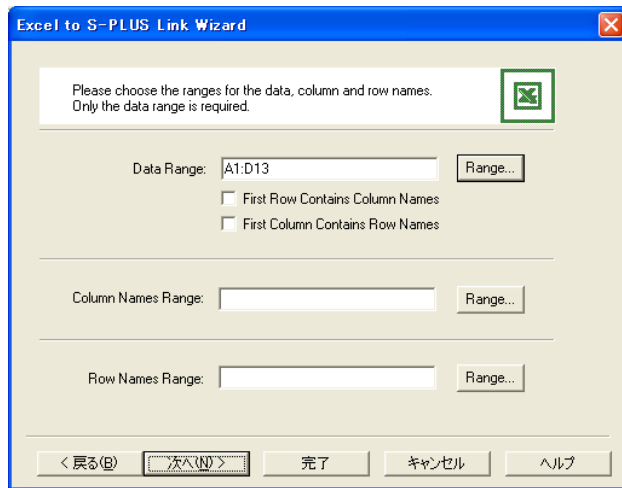


図 12.4 : ウィザードでデータの範囲を指定する

ヒント

データだけを選択したとき、S-PLUS は、ワークシート内で選択したデータ領域の近くで適切な列ラベルと行ラベルを見つけようとします。必要に応じてこれを修正したり削除することができます。あるいは、ウィザードを起動する前に CTRL キーを使ってワークシート内の連続しない複数の領域を選択した場合は、列名/行名にその追加の領域が使用されます。最終的に、これをデフォルトの動作として指定することによって、選択項目がすべて列名/行名を含むように指定することができます。詳細は、第 13 章「S-PLUS セッションのカスタマイズ」を参照してください。

- 選択を変更したい場合は、**Data Range** フィールドの右側の **Range** ボタンをクリックして別の範囲を選択し、ダイアログプロンプトの **OK** をクリックします。
- 選択した範囲に列名/行名が含まれている場合は、適宜 **First Row Contains Column Names** チェックボックス/**First Column Contains Row Names** チェックボックスを指定します。
- 選択した範囲にデータだけが含まれ、列（または行）名だけを指定したい場合は、**Column Names Range**（または **Row Names Range**）フィールドの右側の **Range** ボタンをクリックして別の範囲を選択し、ダイアログプロンプトの **OK** をクリックします。

注意

ウィザードは、列ラベルを正規の S-PLUS 変量名に変換します。行ラベルは、S-PLUS プロット内の注釈に使用されません。

5. 次へをクリックします。
6. ウィザードの 2 ページ目には、S-PLUS データフレームの外観が表示されます (図 12.5 を参照)。

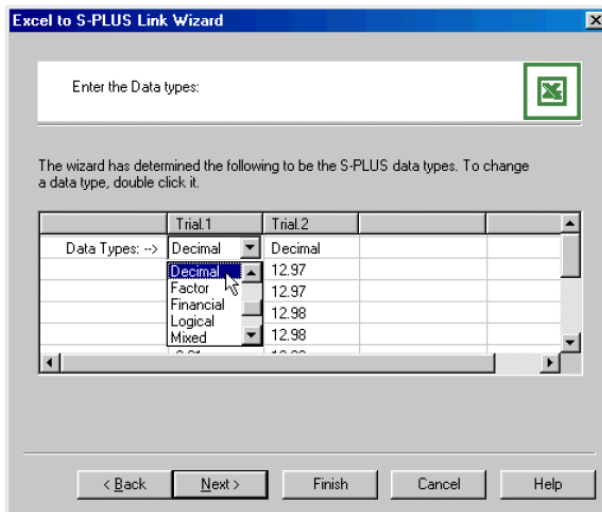


図 12.5 : 異なる列タイプの選択

ウィザードは、データフレームの列タイプが何であるか“推測”しようとします。推測が間違っている場合は、間違っている列タイプを含むセルをクリックし、表示されるドロップダウンリストから違うタイプを選択して列タイプを変更することができます。

注意

ウィザードは、列の最初のセルを調べてその列のデータ形式を決定します。列の最初のセルが欠損値の#N/A かまたは空白の場合、ウィザードは、その列の残りを調べて実際の値を探します。そのような値が見つかった場合は、その形式が列のデータ形式になります。欠損値の#N/A か空白のセルしか見つからなかった場合は、ウィザードは列タイプを **character** に設定します。

下の表 12.1 は、Excel のデータ形式とそれに対応する S-PLUS のデータ形式を横に並べたリストです。

表 12.1 : Excel データ形式とそれに対応する S-PLUS のデータ形式

Excel データ形式	対応する S-PLUS のデータ形式
General	Character
Number	Complex
Currency	Currency
Accounting	Date
Date	Date & Time
Time	Decimal
Percent	Factor
Fraction	Financial
Scientific	Logical
Text	Mixed
Special	Number
Custom	Scientific
	Time

7. **次へ**をクリックします。
8. 図 12.6 に示すように、ウィザードの最後のページで、新しい S-PLUS データフレームの名前を指定するように要求されます。テキストボックスに好きな名前を入力し、**完了**をクリックしてウィザードを閉じ、Excel から S-PLUS へのリンクを作成します。

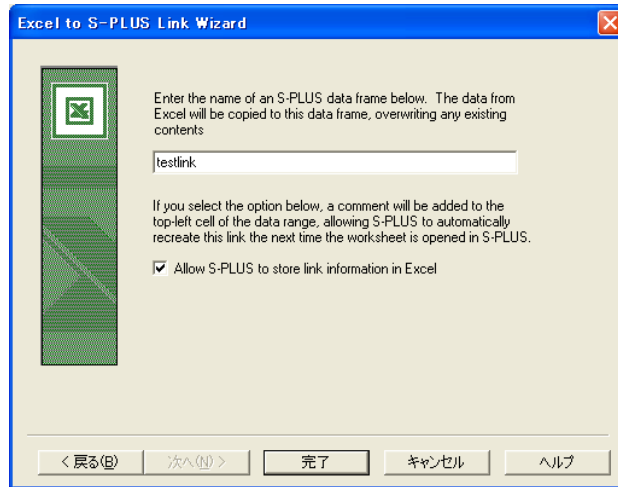


図 12.6 : S-PLUS の新しいデータフレームに名前を付ける

ヒント

同じ Excel ワークシートでたびたび作業していて、その度にウィザードを使ってリンクを作成するプロセスを実行したくない場合は、S-PLUS で Excel ワークシートを開くたびにリンクが自動的に再現される情報を S-PLUS から Excel に保存することができます。これは、(セルの右上角に小さい赤い三角形として示された) 領域の左上のセルにコメントを追加することによって有効になり、データを損なうことはありません。S-PLUS が Excel ファイルを修正しないようにしたい場合は、このページの **Allow S-PLUS to store link information in Excel** チェックボックスのチェックを外してください。このデフォルトの動作をオフにするには、メインメニューから **オプション ▶ 設定** を選択し、**Data** タブをクリックし、**Excel Link** グループの **Save link information** チェックボックスのチェックを外します。

これで、次の 2 つのことが行われました。

- Excel 内で選択したデータ領域が、S-PLUS データフレームに「コピー」されました。
- 「リンク」が作成され、Excel 内のデータが変化したときに S-PLUS 内のデータフレームを容易に更新することができます。


ヒント

Excel to S-PLUS Link Wizard を使うと、リンクを明示的に確立することなく「簡単リンク」を作成することができます。これは、Excel ワークシートがアクティブなドキュメントで、範囲を選択してプロットパレット上のボタンをクリックするか統計メニューからオプションを選択するだけで作成することができます。デフォルトの名前とデフォルトの列ラベル/行ラベルでリンクが自動的に作成されます。

S-PLUS to Excel Link Wizard の 使い方

S-PLUS データフレームに保存されていて S-PLUS 上でプロットまたは解析できるデータを、Excel ワークシートに保存する場合は、**S-PLUS to Excel Link Wizard** を使って S-PLUS 内のデータフレームから Excel ワークシートへのリンクを確立することができます。

S-PLUS から Excel へのリンクを作成するには、以下の操作を行います。

1. S-PLUS 内から新しい Excel ワークシートを作成します。
2. S-PLUS データフレームが表示されているデータ・ウィンドウで、データ・ウィンドウ・ツールバーの **S-PLUS to Excel Link Wizard** ボタン  をクリックするか、**編集** またはショートカットメニューから **S-PLUS ウィザード** または **Link Wizard** を選択します。図 12.7 に示すような **S-PLUS to Excel Link Wizard** が現れます。

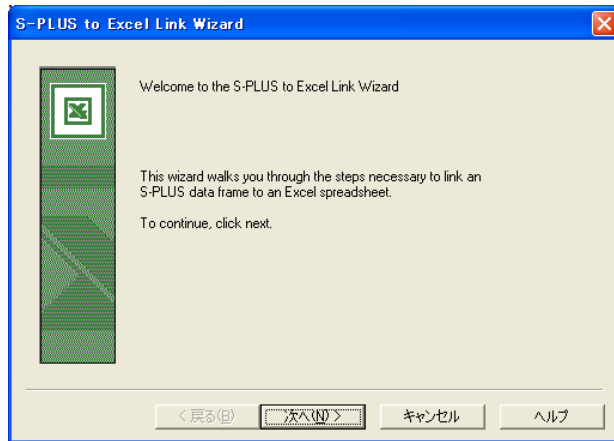


図 12.7 : S-PLUS to Excel Link Wizard

3. **次へ** をクリックします。

4. 図 12.8 から分かるように、**Source data.frame** フィールドにアクティブなデータフレームの名前が自動的に入ります。**Target Excel Workbook** ドロップダウンリストから新しく作成した Excel ワークブックを選択してください。**Target Excel worksheet** フィールドと **Target Excel range** フィールドに値が自動的に入力されるのが分かります。

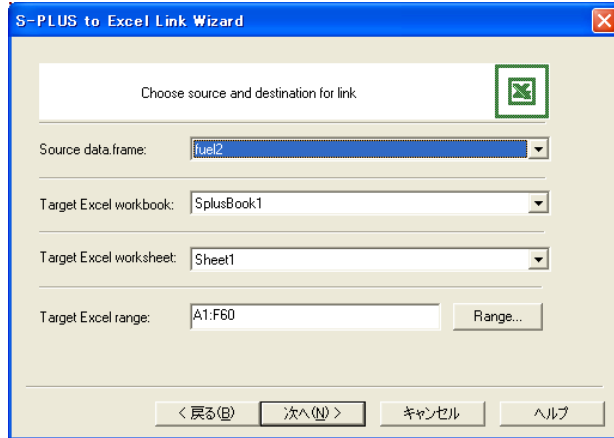


図 12.8 : ウィザードで Excel ターゲットを指定する

- 選択を変更したい場合は、**Target Excel range** フィールドの右側の **Range** ボタンをクリックして別の範囲を選択し、ダイアログプロンプトの **OK** をクリックします。
5. **次へ** をクリックします。
 6. 図 12.9 は、ウィザードの最後のページを示しています。S-PLUS は、リンクに **SplusBook1.Sheet1.A1.D111** の形のデフォルト名を生成します。必要により、テキストボックスに新しいリンクの別の名前を入力することができます。

- 完了をクリックしてウィザードを閉じ、S-PLUS から Excel へのリンクを作成します。

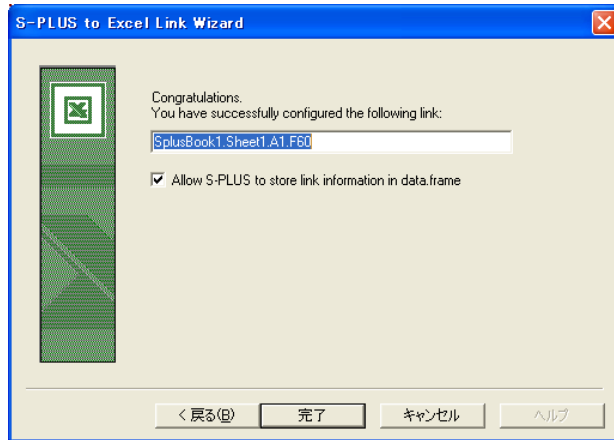


図 12.9 : ウィザードの最終ページ

ヒント

同じ S-PLUS データフレームでたびたび作業していて、その度にウィザードを使ってリンクを作成するプロセスを実行したくない場合は、そのデータフレームを開くたびにリンクが自動的に再現されるようにする情報を保存することができます。これは、データフレームに属性を追加することにより有効になり、データを損なうことはありません。S-PLUS がデータフレームを修正しないようにしたい場合は、このページの **Allow S-PLUS to store link information in data.frame** チェックボックスのチェックを外してください。このデフォルトの動作をオフにするには、メインメニューから **オプション ▶ 設定** を選択し、**Data** タブをクリックし、**Excel Link** グループの **Save link information** チェックボックスのチェックを外してください。

これで、次の 2 つのことが行われました。

- アクティブな S-PLUS データフレームが、新しく作成された Excel ワークシートに「コピー」されました。
- 「リンク」が作成され、S-PLUS 内のデータが変化したときに Excel ワークシートを容易に更新できるようになります。

S-PLUS 内の Excel データの 解析

S-PLUS 内のデータフレームにリンクされた Excel データの領域を解析するときは、以下の操作を行います。

1. **Excel Sheet** ツールバーの **Active Link** ドロップダウンリストから S-PLUS データフレームの名前を選択して、作業したい Excel 領域を選択します。




2. プロットパレットのボタンをクリックしてデータのグラフを作成するか、**統計**メニューからそれぞれの解析手法を選択して統計解析を実行します。

リンクされた データの更新

Excel ワークシートにリンクされた S-PLUS データフレームの更新

Excel 内のデータが変化した場合、対応する S-PLUS データフレームが自動的に更新されることはありません。以下の操作を行って、データを Excel から S-PLUS に強制的にコピーしなおすことができます。


1. **Excel Sheet** ツールバーの **Active Link** ドロップダウンリストから、S-PLUS データフレームの名前を選択します。
2. **Excel Sheet** ツールバーの **Update Excel to S-PLUS link** ボタン  をクリックします。

注意

Update Excel to S-PLUS link ボタンをクリックすると、Excel データは、ウィザードで指定したのと同じデータの範囲、列見出しおよびデータ形式で、指定した S-PLUS データフレームにコピーされます。Excel 領域の次元を変更する場合（たとえば、列/行を追加する）は、リンクを更新しなければならないことに注意してください。

S-PLUS データフレームにリンクされた Excel ワークシートの更新

S-PLUS 内のデータが変化した場合、対応する Excel ワークシートが自動的に更新されることはありません。以下の操作を行って、データを S-PLUS から Excel に強制的にコピーしなおすことができます。

1. **データ・ウィンドウ・ツールバー**の **Active Link** ドロップダウンリストから、Excel リンクの名前を選択します。
2. **データ・ウィンドウ・ツールバー**の **Update current link** ボタン  をクリックします。


データを保存する

Excel ワークシートを閉じるとき、Excel データをファイルに保存するように要求されます。S-PLUS データフレームにリンクされた Excel データの場合、データフレームはデフォルトで自動的に削除されますが、S-PLUS で次に Excel ワークシートを開いたときに再生されます。必要に応じて、**オプション**の **General Settings** ダイアログの **Data** タブでこのデフォルトの動作を変更することができます。詳細は、第 13 章「S-PLUS セッションのカスタマイズ」を参照してください。

リンクの削除

Excel から S-PLUS へのリンクを削除する


Excel 領域からそれに対応する S-PLUS データフレームへのリンクを削除するには、以下の操作を行います。

1. **Excel Sheet** ツールバーの **Active Link** ドロップダウンリストから、S-PLUS データフレームの名前を選択します。
2. **Excel Sheet** ツールバーの **Remove Excel to S-PLUS link** ボタン  をクリックします。

リンクを削除すると、Excel 内の対応するコメントも削除されることに注意してください。

S-PLUS から Excel へのリンクを削除する

S-PLUS データフレームからそれに対応する Excel ワークシートへのリンクを削除するには、以下の操作を行います。

1. **データ・ウィンドウ・ツールバー**の **Active Link** ドロップダウンリストから、Excel リンクの名前を選択します。
2. **データ・ウィンドウ・ツールバー**の **Remove current link** ボタン  をクリックします。

リンクを削除すると、S-PLUS データフレーム内の対応する属性も削除されることに注意してください。

Excel アドインアプリケーションの使用

Microsoft Excel アドインアプリケーションを使用すると、Microsoft Excel 内部で簡単に S-PLUS グラフを作成したり修正できるようになります。このアドインは、選択したデータから S-PLUS グラフを作成したり、Excel に埋め込まれた S-PLUS グラフのレイアウトを修正したり、Excel に埋め込まれた S-PLUS グラフのプロットプロパティを修正することができます。Excel の Chart Wizard のような分かりやすいウィザードにしたがって Excel でデータを選択し、S-PLUS グラフと図表示の種類を選択してグラフを作成することができます。

Excel アドイン をインストール する

S-PLUS を通常インストールする際、セットアップでシステムに Microsoft Excel の適切なバージョンがあるかを調べます。適切なバージョンが検出された場合は、Excel アドインが自動的にインストールされます。

注意

セットアップ時に Excel アドインを自動的にインストールしないようにするには、**変更**オプションを選択し、インストールするコンポーネントのリストの **Excel Add-in** のチェックを外してください。

S-PLUS のインストールの際に Excel アドインをインストールしないように選択した場合は、後でセットアップを実行し、**変更**オプションを選択し、インストールするコンポーネントのリストから **Excel Add-in** を選択することによって、後からインストールすることができます。

Excel アドインをコンピュータにインストールしても **S-PLUS** メニューとツールバーが Excel に現れない場合は、Excel 内からアドインを使用できるようにすることができます。これは、以下のステップで行います。

1. Excel を起動します。
2. まだワークシートがない場合は、新しいワークシートを作成します。
3. ツールメニューから、**アドイン**を選択します。
4. **アドイン**ダイアログで、図 12.10 に示すように、**S-PLUS Add-In** チェックボックスを選択し、**OK** をクリックします。

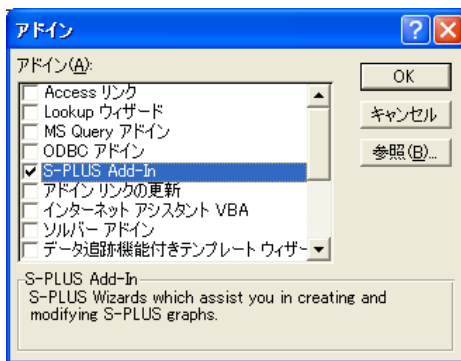


図 12.10 : S-PLUS Add-In がチェックされた Excel の Add-Ins ダイアログ

Excel Add-In を削除する

Excel アドインを削除するように選択した場合は、S-PLUS セットアップを実行するか、Excel 内からアドインを削除することによってそれを行うことができます。

S-PLUS セットアップを使って Excel アドインを削除するには、以下の操作を行います。

1. S-PLUS セットアップを実行し、**変更**オプションを選択します。
2. コンポーネントのリストから、**Excel Add-in** を選択します。
3. 画面上の指示にしたがいます。

注意

S-PLUS をインストールする際に Excel アドインをインストールし、その後で S-PLUS 自身を削除した場合は、Excel アドインは自動的に削除されます。

Excel 内から Excel アドインを使用できないようにするには、以下の操作を行います。

1. Excel を起動します。
2. まだワークシートがない場合は、新しいワークシートを作成します。
3. ツールメニューからアドインを選択します。
4. **Add-Ins** ダイアログで、**S-PLUS Add-In** チェックボックスのチェックを外し（図 12.10 を参照）、**OK** をクリックします。

Excel アドイン を使用する

Excel アドインがインストールされていて Excel が表示されているときは、図 12.11 に示すような **S-PLUS** メニューとツールバーが現れます。

メニューとツールバーには次のオプションがあります。

- **Create Graph**
- **Modify Graph Layout**
- **Modify Plots**

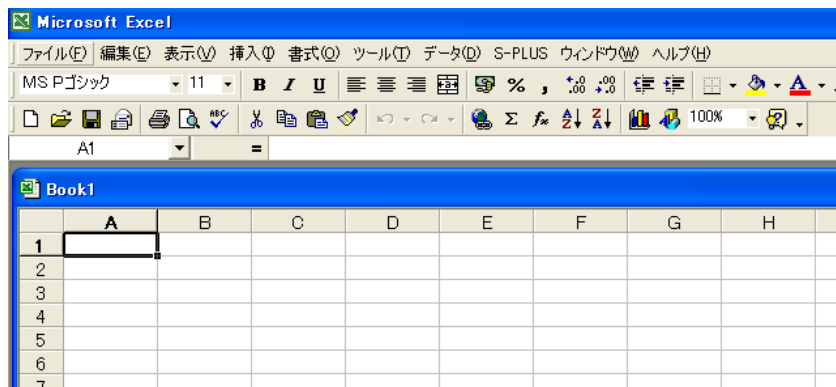


図 12.11 : Excel アドインの **S-PLUS** メニューとツールバー

グラフを作成する

現在のワークシートで選択されているデータで新しい **S-PLUS** グラフを作成するには、以下の操作を行います。

1. グラフにしたい現在のワークシート内のデータブロックを選択します。
2. **S-PLUS** ツールバーの **Create Graph** ボタンをクリックするか、Excel のメインメニューから **S-PLUS ▶ Create Graph** を選択します。
3. **Create S-PLUS Graph** ウィザードの指示にしたがって、グラフを作成します。

グラフのレイアウトを修正する

現在選択されている S-PLUS グラフのレイアウトプロパティを修正するには、以下の操作を行います。

1. ワークシート内の S-PLUS グラフを 1 回クリックして、そのグラフを選択します。(グラフをダブルクリックすると、そのグラフがアクティブになり、編集を行うことができます。)
2. **S-PLUS** ツールバーの **Modify Graph Layout** ボタンをクリックするか、Excel のメインメニューから **S-PLUS ▶ Modify Graph Layout** を選択します。
3. Excel に S-PLUS の **Graph Sheet** ダイアログが開き、このグラフのレイアウトプロパティを修正することができます。

プロットを修正する

現在選択されている S-PLUS グラフ内のプロットのプロパティを修正するには、以下の操作を行います。

1. ワークシート内の S-PLUS グラフを 1 回クリックして、そのグラフを選択します。
2. **S-PLUS** ツールバーの **Modify Plots** ボタンをクリックするか、Excel のメインメニューから **S-PLUS ▶ Modify Plots** を選択します。
3. ダイアログが開き、そのグラフ内のグラフ領域のリストが表示され(グラフ内に複数のグラフ領域があってもかまいません。すなわち同じグラフ内のあるグラフ領域が 2 次元で、別のグラフ領域が 3 次元の場合もあります)、各グラフ領域ごとに、そのグラフ領域内のすべてのプロットを示すリストが表示されます。
4. 編集したいグラフ領域とその領域内のプロットを選択し、次に **Next** をクリックします。
5. Excel に S-PLUS プロットプロパティ・ダイアログが現れ、そのプロットのプロパティを修正することができます。

グラフのデータを選択する

グラフを作成する前に、まず現在のワークシートのデータを選択しなければなりません。幅または長さに 2 つ以上のセルがあるデータブロックを選択した後で、**Create S-Plus Graph** ウィザードで処理を続けることができます。

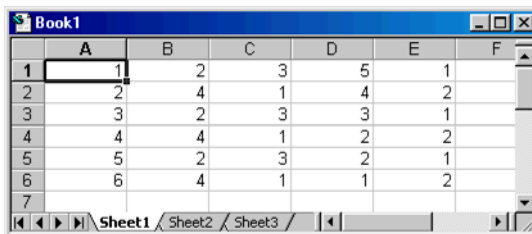
第 12 章 他のアプリケーションでの S-PLUS の使用

S-PLUS プロットでは、様々なフォーマットのデータを受け入れることができます。あるプロットには、3つの列が x、y および z のデータ値として解釈される少なくとも 3 列のデータが必要です。他のプロットには、4つの列が x、y、z および w のデータ値として解釈される少なくとも 4 列のデータが必要です。**Create S-Plus Graph** ウィザードの最後のページの図表示のリストは、必要なデータ指定の種類を示します。図表示のリストに x、y、z または w が指定されていない場合は、x の 1 つまたは複数の列、あるいは 1 つまたは複数の列を含む x および y データを受け入れるということです。

警告

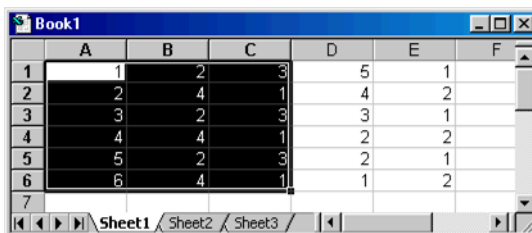
Excel は行が空白であるかどうかに関係なく、グラフ作成のために S-PLUS に列のすべての行を送るため、通常は S-PLUS グラフのデータ指定のとき Excel の列全体を選択しないでください。選択すると、S-PLUS にエラーが発生するか、グラフが作成できないことがあります。

Excel アドインは、S-PLUS グラフのデータを指定するために、Excel における複数列／行の選択と連続していないブロックの選択を完全にサポートしています。たとえば、Excel の次のようなデータを考えてみましょう。



	A	B	C	D	E	F
1	1	2	3	5	1	
2	2	4	1	4	2	
3	3	2	3	3	1	
4	4	4	1	2	2	
5	5	2	3	2	1	
6	6	4	1	1	2	
7						

グラフに 2 つの線グラフを作成したい場合は、次のようにデータを選択することができます。



	A	B	C	D	E	F
1	1	2	3	5	1	
2	2	4	1	4	2	
3	3	2	3	3	1	
4	4	4	1	2	2	
5	5	2	3	2	1	
6	6	4	1	1	2	
7						

Create S-Plus Graph ウィザードは、この選択部分の **A** 列を x データとして扱い、**B** および **C** 列を y データとして扱います。このケースでは、x データが **A** 列で y データが **B** 列の線グラフと、x データが **A** 列で y データが **C** 列の線グラフの 2 つが作成されます。

また、連続していない列を選択して、同じデータを選択することができます。

	A	B	C	D	E	F
1	1	2	3	5	1	
2	2	4	1	4	2	
3	3	2	3	3	1	
4	4	4	1	2	2	
5	5	2	3	2	1	
6	6	4	1	1	2	
7						

この例では最初に A 列の第 1 行～第 6 行を選択し、次に CTRL キーを押したまま **B1** から **C6** のブロックを選択します。この選択により最初の例と同じように、2 つのプロットを含むグラフが生成されます。

プロットが、線グラフの x データのように所定の次に 1 列のデータだけがあると予想している場合、複数の列が選択された場合は選択範囲のうちの第 1 列だけが S-PLUS に送られ、グラフが作成されます。たとえば上記のデータを使用して、ブロック **A1:B6** と **C1:D6** を選択します。

	A	B	C	D	E	F
1	1	2	3	5	1	
2	2	4	1	4	2	
3	3	2	3	3	1	
4	4	4	1	2	2	
5	5	2	3	2	1	
6	6	4	1	1	2	
7						

Create S-Plus Graph ウィザードは、x データとして **A1:A6** を送り y データとして **C1:D6** を送って、2 つの線グラフを作成します。

グラフの条件付けを行うデータを選択する

Create S-Plus Graph ウィザードを使用しているとき、ステップ 2 で Excel ワークシートとデータの範囲を指定して、作成中のグラフの条件付けを行うことができます。条件付けされたグラフを使用すると、元データのサブセットをそれぞれ含む一連のパネルにデータを表示させることができます。各パネル内のサブセットは、選択する条件付けデータ範囲のレベルによって決定されます。このダイアログの **Conditioning Range** フィールドを空白のままにすると、条件付けを省略することができます。

第 12 章 他のアプリケーションでの S-PLUS の使用

条件付けデータ範囲を指定するときは、通常の Excel の範囲シンタックスで有効なデータ範囲を指定することができます。たとえば S-PLUS グラフにプロットするデータに、ワークシートの **Sheet1** からデータ範囲 **A1:B6** を指定します。また条件付けデータに、**Sheet1** からデータ範囲 **C1:C6** を指定することができます。2次元線グラフを作成する場合、プロットは **C1:C6** のデータに基づいて条件付けが行われます。

グラフ作成中のエラーを処理する

Excel のグラフ作成中に S-PLUS に問題が生じると、Excel のモードレスダイアログボックスにエラーメッセージが現われます。エラーが発生した場合、作成されたプロットに無効なデータが指定されていることがあります。また、指定されたデータの範囲またはデータ形式に関連した別の問題を示していることもあります。エラーが発生すると、グラフが作成されないことがあります。

SPSS での S-PLUS の使用

SPSS アドインアプリケーションは、SPSS と共に動作して SPSS 内から S-PLUS グラフを簡単に作成したり修正することができます。このアドインは、SPSS データエディタで選択された変量から S-PLUS グラフを作成し、SPSS 出力ドキュメントに埋め込まれた S-PLUS グラフのレイアウトを修正し、SPSS に埋め込まれた S-PLUS グラフのプロットプロパティを修正することができます。分かりやすいウィザードにしたがって変量を選択したり、S-PLUS グラフと図表示を選択したり、SPSS でグラフを作成することができます。

S-PLUS の SPSS アドインを使用して、100,000 行を超えるデータに基づいたプロットを作成しようとするエラーが生じます。SPSS アドインでは、使用する行数の上限は 100,000 です。

SPSS アドインをインストールする

S-PLUS を通常インストールする際、セットアップでシステムに SPSS の適切なバージョン（バージョン 8.0 以上）があるかを調べます。適切なバージョンが検出された場合は、SPSS アドインが自動的にインストールされます。

注意

セットアップ時に SPSS のアドインを自動的にインストールしないようにするには、**変更**オプションを選択し、インストールするコンポーネントのリストの **SPSS Add-in** のチェックを外してください。

S-PLUS をインストールする際に SPSS アドインをインストールしないように選択した場合は、後でセットアップを実行し、**変更**オプションを選択し、インストールするコンポーネントのリストから **SPSS Add-in** を選択することによって、いつでもインストールすることができます。

SPSS アドインを削除する

SPSS アドインを削除するには、以下の操作を行います。

1. S-PLUS セットアップを実行し、**変更**オプションを選択します。
2. コンポーネントのリストから、**SPSS Add-in** を選択します。
3. 画面上の指示にしたがいます。

注意

S-PLUS をインストールする際に SPSS アドインをインストールし、その後で S-PLUS 自身を削除した場合は、SPSS アドインは自動的に削除されます。

SPSS Add-In の
使用

SPSS アドインをインストールして SPSS データエディタを開いたときは、図 12.12 に示すような **S-PLUS** メニューとツールバーが現れます。メニューとツールバーは、出力ドキュメントを開いたときにいつでも使用することができます。

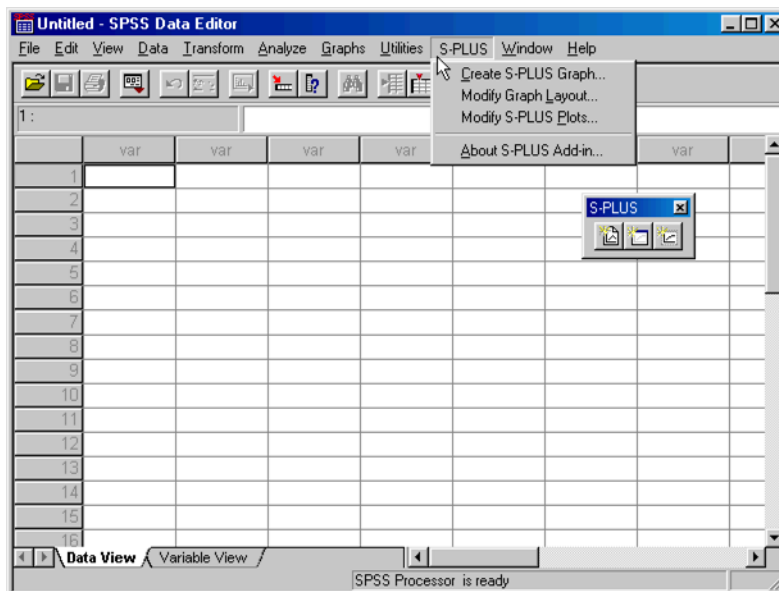


図 12.12 : SPSS アドインの S-PLUS メニューとツールバー

メニューとツールバーには、以下のオプションがあります。

- **Create S-PLUS Graph**
- **Modify Graph Layout**
- **Modify S-PLUS Plots**

グラフを作成する

SPSS アドインで作成した S-PLUS グラフは、出力ドキュメントに入れられます。新しい出力ドキュメントを作成するか、既存の出力ドキュメントを使用するかを選択することができます。

データエディタで現在選択されている変数で新しい S-PLUS グラフを作成するには、以下の操作を行います。

1. データエディタ内でグラフを作成したい変数を選択します。
2. **S-PLUS** ツールバーの **Create Graph** ボタンをクリックするか、SPSS メインメニューから **S-PLUS ▶ Create Graph** を選択します。
3. **Create S-PLUS Graph** ウィザードの指示にしたがって、グラフを作成します。

グラフのレイアウトを修正する

現在選択されている S-PLUS グラフのレイアウトプロパティを修正するには、以下の操作を行います。

1. 出力ドキュメント内の S-PLUS グラフを 1 回クリックして、そのグラフを選択します。(グラフをダブルクリックすると、そのグラフがアクティブになり編集を始めることができます。)
2. **S-PLUS** ツールバーの **Modify Graph Layout** ボタンをクリックするか、SPSS メインメニューから **S-PLUS ▶ Modify Graph Layout** を選択します。
3. SPSS に **S-PLUS Graph Sheet** ダイアログが現れ、そのグラフのレイアウトプロパティを修正することができます。

プロットを修正する

現在選択されている S-PLUS グラフのプロットのプロパティを修正するには、以下の操作を行います。

1. 出力ドキュメント内の S-PLUS グラフを 1 回クリックして、そのグラフを選択します。
2. **S-PLUS** ツールバーの **Modify Plots** ボタンをクリックするか、SPSS メインメニューから **S-PLUS ▶ Modify Plots** を選択します。
3. ダイアログが開き、そのグラフ内のグラフ領域のリストが表示され(グラフ内に複数のグラフ領域があってもかまいません。すなわち同じグラフ内のあるグラフ領域が 2 次元で、別のグラフ領域が 3 次元の場合もあります)、各グラフ領域ごとに、グラフ領域内のすべてのプロットを示すリストが表示されます。
4. 編集したいグラフ領域とその領域内のプロットを選択し、次に **Next** をクリックします。
5. SPSS に **S-PLUS プロットプロパティ・ダイアログ** が現れ、このプロットのプロパティを修正することができます。

グラフのデータを選択する

グラフを作成する前に、まずデータエディタでデータを選択してください。グラフに含めたい各変数の変数名が表示された列ヘッダをクリックすることによって、SPSS データエディタ内の変数を選択することができます。

S-PLUS プロットは、様々なフォーマットのデータを受け入れます。いくつかの S-PLUS プロットには、x、y および z データ値として解釈される少なくとも 3 列のデータが必要です。他のプロットには、x、y、z および w データ値として解釈される少なくとも 4 列のデータが必要です。

たとえば、SPSS で以下の変数を確認してください。

	xdata	ydata1	ydata2	ydata3
1	1.00	2.00	5.00	5.00
2	2.00	4.00	3.00	2.00
3	3.00	2.00	2.00	2.00
4	4.00	4.00	3.00	3.00
5	5.00	2.00	4.00	1.00
6	6.00	4.00	2.00	1.00

グラフ内に 2 つの線グラフを作成したい場合は、次のように変数 **xdata**、**ydata1** および **ydata2** を選択することができます。

	xdata	ydata1	ydata2	ydata3
1	1.00	2.00	5.00	5.00
2	2.00	4.00	3.00	2.00
3	3.00	2.00	2.00	2.00
4	4.00	4.00	3.00	3.00
5	5.00	2.00	4.00	1.00
6	6.00	4.00	2.00	1.00

Create S-PLUS Graph ウィザードは、この選択範囲の変数 **xdata** を x データとして扱い、変数 **ydata1** と **ydata2** を y データとして扱います。これにより、x データが変数 **xdata** で y データが変数 **ydata1** の線グラフと、x データが変数 **xdata** で y データが変数 **ydata2** の線グラフの 2 つが作成されます。

Create S-PLUS Graph ウィザードのステップ 1 と 2 によって、選択されたリストの変数を追加／削除／並べ替えして S-PLUS グラフを作成することができます。

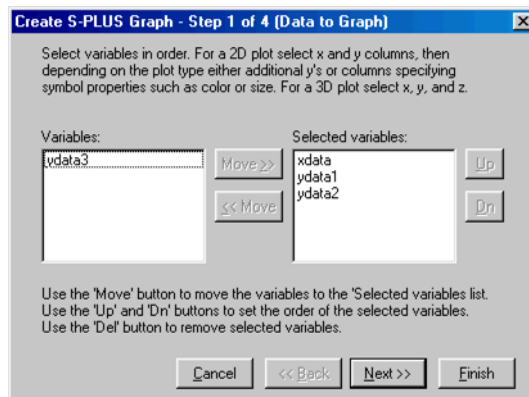


図 12.13 : Create S-PLUS Graph (Data to Graph) ダイアログ

Variables という名前のリストは、データエディタで使用可能なすべての変数のリストです。**Selected variables** という名前のリストは、S-PLUS グラフに含めるように選択した使用可能な変数からの変数のリストです。変数名がいずれかのリストで選択されているとき、**Move** ボタンを使ってその変数をリスト間で移動させることができます。**Selected variables** リストで変数名が選択されているときは、**Up** および **Dn** ボタンを使って変数の順序を変更することができます。選択した変数の順序は、その順序によって S-PLUS がデータをどのようにグラフ化するかが決まるため重要です。類似のダイアログを使用して、作成するグラフの条件付けを行う変数を選択することができます。

グラフの条件付けを行うデータを選択する

Create S-PLUS Graph ウィザードを使用しているとき、ステップ 2 で次のようにして作成中のグラフの条件付けに使用する変数を指定できます。

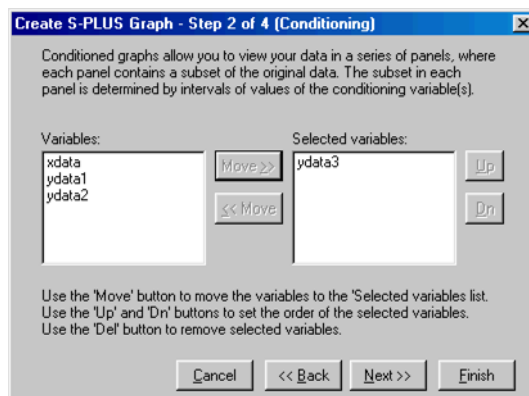


図 12.14 : Create S-PLUS Graph (Conditioning) ダイアログ

グラフの条件付けを行うと、一連のパネルに元データのサブセットをそれぞれ含む状態でデータを表示させることができます。各パネル内のサブセットは、選択する条件付けデータ範囲のレベルによって決定されます。このダイアログ内の **Selected variables** リストを空白のままにすることによって、条件付けを省略することができます。

グラフ作成中のエラーを処理する

SPSS でグラフを作成中に S-PLUS に問題が生じると、SPSS のモードレスダイアログボックスにエラーメッセージが現れます。エラーが生じた場合は、作成されたプロットに無効なデータが指定されていることがあります。また、指定したデータの範囲またはデータ形式に関連した別の問題があることもあります。エラーが発生すると、グラフが作成されないことがあります。

MathSoft Mathcad での S-PLUS の使用

Mathcad 用の S-PLUS コンポーネントを使用すると、MathSoft Mathcad アプリケーションから簡単に S-PLUS グラフを作成したり修正することができます。コンポーネントは選択した Mathcad の変量から S-PLUS グラフを作成し、Mathcad ワークシートに埋め込まれた S-PLUS グラフのレイアウトまたはプロパティを修正することができます。グラフウィザードにしたがって、S-PLUS グラフと図表示の種類を選択して、Mathcad でグラフを作成することができます。

また、S-PLUS 内でスクリプト言語を使ってデータを作成または処理し、その情報を Mathcad に返すことができます。スクリプトウィザードを使用して、S-PLUS 言語コマンドを入力し、Mathcad から渡される入力の数と Mathcad に返す出力の数を指定することができます。

S-PLUS Mathcad コンポーネント をインストール する

S-PLUS を通常インストールする際、セットアップでシステムに適切なバージョンの Mathcad (バージョン 8.02 以上) があるかを調べます。適切なバージョンが検出された場合は、S-PLUS Mathcad コンポーネントが自動的にインストールされます。

注意

Mathcad がそれ以前のバージョンの場合は、セットアップで、Mathcad 8.02 が必要であるという警告が出され、S-PLUS Mathcad コンポーネントをインストールできないことを知らせます。

S-PLUS Mathcad コンポーネント を使用する

S-PLUS Mathcad コンポーネントがインストールされている場合、Mathcad メインメニューから **挿入 ▶ コンポーネント** を選択することによって、任意の Mathcad ワークシートからそのコンポーネントを使用することができます。図 12.15 に示すような **Component Wizard** が現われます。

Component Wizard には、次の 2 つのオプションがあります。

- S-PLUS グラフの作成
- S-PLUS 言語スクリプトの作成と実行

1 つのワークシートで両方のオプションを使用したい場合があります。まず、S-PLUS でスクリプトを使用していくつかのデータを作成または修正し、その後でそのデータをグラフ化することができます。Mathcad ワークシートに、データとグラフの両方と、目的とする処理の説明が表示されます。

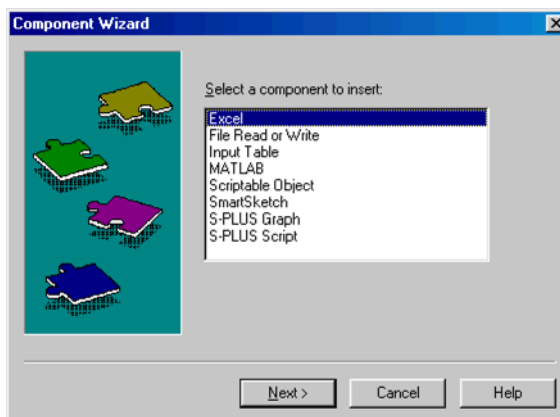


図 12.15 : Component Wizard

グラフを作成する

S-PLUS Mathcad コンポーネントでグラフを作成するには、以下の操作を行います。

1. Mathcad ワークシートの空白部分をクリックします。挿入点が赤い十字線で示されます。
2. Mathcad メインメニューから、**挿入 ▶ コンポーネント**を選択して、**Component Wizard** ダイアログを開きます。
3. **S-PLUS Graph** を選択し、次に **Next** をクリックして、図 12.16 に示すように、**Graph Setup Wizard** を起動します。

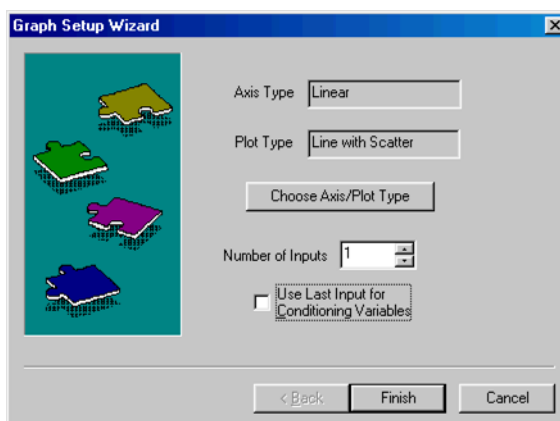


図 12.16 : Graph Setup Wizard

4. デフォルトの図表示は、入力が 1 つの線付き散布図です。これは、1 組のインデックスに対して 1 つの列ベクトルをプロットします。デフォルトの図表示にしたい場合は、**Finish** をクリックします。図表示の種類を変更したい場合は、**Choose Axis/Plot Type** をクリックします。図 12.17 に示すような **Choose Graph and Plot Type** ダイアログが現れます。

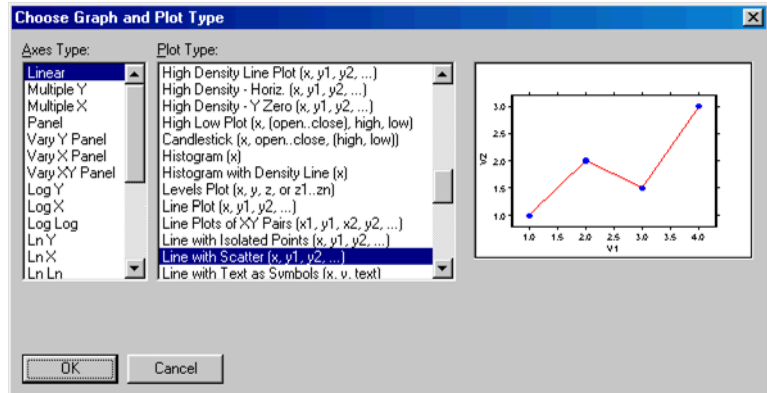


図 12.17 : Choose Graph and Plot Type ダイアログ

5. **Axis Type** リストから軸のタイプを選択し、**Plot Type** リストからの図表示の種類を選択して、次に **OK** をクリックします。

注意

図表示の後の引数リストは、特定のタイプに関してデータをどのように入力しなければならないかを指定します。多くの種類の図は、可変数の引数の入力が可能です。たとえば **Line Plot (x, y1, y2, ...)** は可変数の引数の入力を取り、ここでは最初のものが x 値として解釈され、残りは y 値として解釈されます。

6. **Graph Setup Wizard** の **Number of Inputs** フィールドで図表示に必要な入力数を指定します。
7. **Trellis** グラフを作成するときは、**Use Last Input for Conditioning Variables** ボックスを選択します。最後の入力は、複数の列を含むこともでき、それらはすべて条件付き変量として使用されます。
8. **Finish** をクリックして、コンポーネントを挿入します。Mathcad ワークシートに、指定した入力に対して、1 つまたは複数のプレースホルダーがある空白の四角形が現われます。

第 12 章 他のアプリケーションでの S-PLUS の使用

9. 適切な入力を指定すると、プロットが現われます。

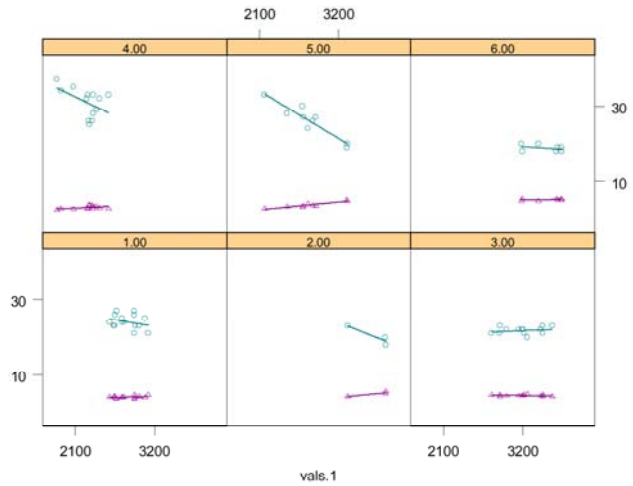
1 つの例として、1 つの条件付き変数を含む 2 つの回帰直線散布図を作成したいと思います。

1. Mathcad ワークシートに変数 x 、 y_1 、 y_2 および $cond$ のデータ点を定義します。
2. S-PLUS グラフのコンポーネントを挿入します。
3. **Graph Setup Wizard** から、**Choose Axis/Plot Type** ボタンをクリックします。
4. **Choose Graph and Plot Type** ダイアログで、**Axes Type** として **Linear**、**Plot Type** として **Fit-Linear Least Squares** を選択して、**OK** をクリックします。
5. **Graph Setup Wizard** で、**Number of Inputs** フィールドに **4** を入力し、**Use Last Input for Conditioning Variables** チェックボックスを選択します。**OK** をクリックします。
6. 入力変数プレースホルダーをクリックし、変数名 (x 、 y_1 、 y_2 および $cond$) を入力します。コンポーネントの外をクリックして、プロットをアクティブにします。
7. コンポーネントのグラフ領域をダブルクリックし、**S-PLUS** ツールバーを利用してプロットを好きなようにフォーマットします。

図 12.18 は、そのようなグラフを示します。この例では、データセット `fuel.frame` を使用し、ここで x は変数 `Weight`、 y_1 は `Mileage`、 y_2 は `Fuel`、 $cond$ は因子変数 `Type` の「コード」です。値 1 はタイプ `Compact`、2 は `Large`、3 は `Medium`、4 は `Small`、5 は `Sporty`、6 は `Van` に対応します。

注意

グラフのコンポーネントへの変更をいくつか入力すると、既存のプロットが削除され新しいプロットが作成されますが、他の設定値はすべて保存されます。しかし、グラフタイプを変更すると、要求したグラフタイプは、すべてのプロットが削除されてから作成されます。このとき、他の修正はすべて失われます。



(x y1 y2 cond)

図 12.18 : Mathcad ワークシート内の Trellis グラフ

S-PLUS スクリプトを作成する

Mathcad で S-PLUS 言語スクリプトを作成するには、以下の操作を行います。

1. Mathcad ワークシート内の空白部分をクリックします。挿入点が赤い十字線で示されます。
2. Mathcad メインメニューから、**挿入 ▶ コンポーネント**を選択します。**Component Wizard** ダイアログが現れます。

3. **S-PLUS Script** を選択し、次に **Next** をクリックすると、図 12.19 に示すような **S-PLUS Script Setup Wizard** が現れます。

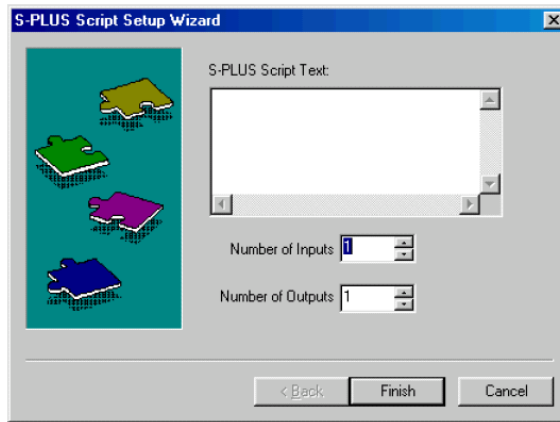


図 12.19 : S-PLUS Script Setup Wizard

4. **S-PLUS Script Text** フィールドに S-PLUS 言語コマンドを入力します。たとえば、以下の一連のコマンドを入力します。

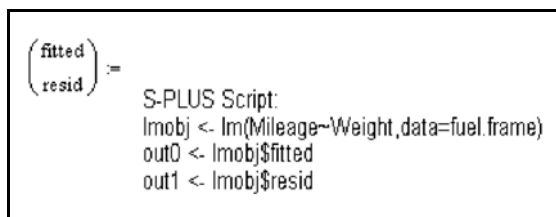
```
lmobj <- lm(Mileage ~ Weight, data = fuel.frame)
out0 <- lmobj$fitted
out1 <- lmobj$resid
```

注意

デフォルトでは、コンポーネントには out0、out1 などの名前がついています。デフォルトの出力変数名を変更するときは、589 ページの「コンポーネントを修正する」を参照してください。

5. **Number of Inputs** フィールドで、Mathcad から S-PLUS に渡される値の数を指定します。(ステップ 4 に示したスクリプトの例では、入力の数 は 0 です。)
6. **Number of Outputs** フィールドで、S-PLUS から Mathcad に渡される値を指定します。(ステップ 4 に示したスクリプトの例では、出力の数 は 2 です。)
7. **Finish** をクリックしてコンポーネントを挿入します。入力用のプレースホルダーがコンポーネントの下に現われ、出力用のプレースホルダーが左上に現われます。

8. 入力／出力の Mathcad 変数名を入力します(ある場合)。この例では、出力変数名として `fitted` と `resid` を使用することができます。ワークシートに、次のようなスクリプトコンポーネントが現われます。



The screenshot shows a rectangular box containing the following text:

```
(fitted)
(resid) := S-PLUS Script:
lmobj <- lm(Mileage~Weight,data=fuel.frame)
out0 <- lmobj$fitted
out1 <- lmobj$resid
```

9. 入力変数と出力変数が指定されているときは、これらの変数を Mathcad に使用することができます。

コンポーネントの外をクリックして、定義した変数を入力または使用します。たとえば、S-PLUS スクリプトの実行結果を見るときは `fitted=` を入力します。

コンポーネントを修正する

Mathcad ワークシートのコンポーネントを右クリックし、次にショートカットメニューから **Properties** を選択することによって、任意の S-PLUS コンポーネントを修正することができます。Script コンポーネントの場合は、スクリプトテキストを右クリックし、次に **Properties** を選択します。スクリプトテキストの外側をクリックすると、別のショートカットメニューがポップアップします。プロパティダイアログが現れ、元のスクリプトまたはグラフウィザードのオプションと、さらにスクリプトコンポーネントに関して入力と出力の変数名を変更するオプションがすべて表示されます。デフォルトでは、これらの名前は、`in0/in1/in2/in3` および `out0/out1/out2/out3` です。入力名と出力名を修正することは、意味のある変数名を使用する S-PLUS スクリプトが既に存在し、それをコンポーネントとして含める前にスクリプトを編集したくない場合に有効です。

Microsoft PowerPoint での S-PLUS の使用

コンピュータに Microsoft PowerPoint7.0 以上がインストールされている場合は、S-PLUS **Graph Sheets** を使用して PowerPoint プレゼンテーションを自動的に作成することができます。

PowerPoint Presentation Wizard をインストールする

S-PLUS を通常インストールする際、セットアップでシステムに適切なバージョンの PowerPoint (バージョン 7.0 以上) があるか調べます。適切なバージョンを検出した場合は、PowerPoint Presentation Wizard が自動的にインストールされます。

注意

セットアップ時に PowerPoint Presentation Wizard を自動的にインストールしないようにするには、**変更オプション**を選択し、インストールするコンポーネントのリストの **PowerPoint Presentation Wizard** のチェックを外してください。

S-PLUS をインストールする際に PowerPoint Presentation Wizard をインストールしないように選択した場合は、後でセットアップを実行し、**変更オプション**を選択し、インストールするコンポーネントのリストから **PowerPoint Presentation Wizard** を選択することによって、後からインストールすることができます。

PowerPoint Presentation Wizard を削除する

PowerPoint Presentation Wizard を削除するには、以下の操作を行います。


1. S-PLUS セットアップを実行し、**変更オプション**を選択します。
2. コンポーネントのリストから、**PowerPoint Presentation Wizard** を選択します。
3. 画面上の指示にしたがいます。

注意

S-PLUS をインストールする際に PowerPoint Presentation Wizard をインストールし、その後で S-PLUS 自身を削除した場合は、PowerPoint Presentation Wizard は自動的に削除されます。

PowerPoint プレゼン テーションを 作成する

S-PLUS グラフシートを使用して PowerPoint プレゼンテーションを作成するには、以下の操作を行います。

1. 標準ツールバーの **PowerPoint** プレゼンテーションボタン  をクリックするか、ファイルメニューから **PowerPoint** プレゼンテーションの作成を選択します。
2. **PowerPoint Presentation Wizard** の **Welcome** 画面が現れます。

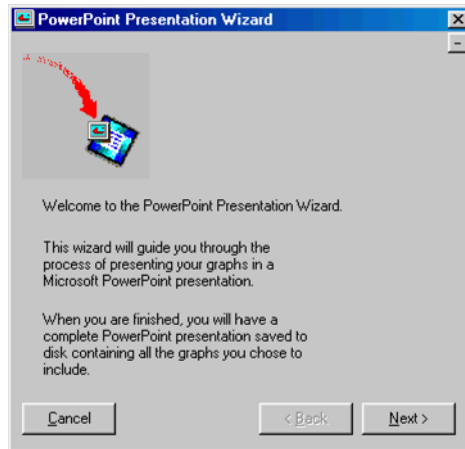


図 12.20 : PowerPoint Presentation Wizard

3. **Next** をクリックします。

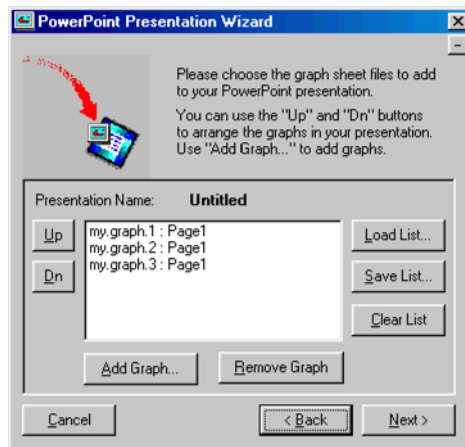


図 12.21 : グラフシートの選択

第 12 章 他のアプリケーションでの S-PLUS の使用

4. 図 12.21 に示すように、ウィザードの最初のページにいくつかのオプションが表示されます。デフォルトでは、プレゼンテーション用にセッションで現在開いている**グラフシート**が選択されます。しかし、以下の操作で選択を変更することができます。
- **Add Graph** ボタンをクリックし、プレゼンテーションに追加する保存されている**グラフシート**を探します。
 - ウィンドウ内の**グラフシート**を選択し、**Remove Graph** ボタンをクリックして、プレゼンテーションからその**グラフシート**を削除します。
 - **Load List** ボタンをクリックして、前に保存したプレゼンテーションリストをロードします。
 - **Save List** ボタンをクリックして、現在のプレゼンテーションリストを保存します。
 - **Clear List** ボタンをクリックして、プレゼンテーションリストの内容をリセットします。

グラフシートの順序を並べ替えるには、**Up** ボタンと **Dn** ボタンを使用します。

注意

グラフシートが複数のページを含む場合は、PowerPoint プレゼンテーションに含めるために、すべてまたは任意のページを選択することができます。

5. **Next** をクリックします。



図 12.22 : プレゼンテーションを作成する準備

6. **Finish** をクリックします。

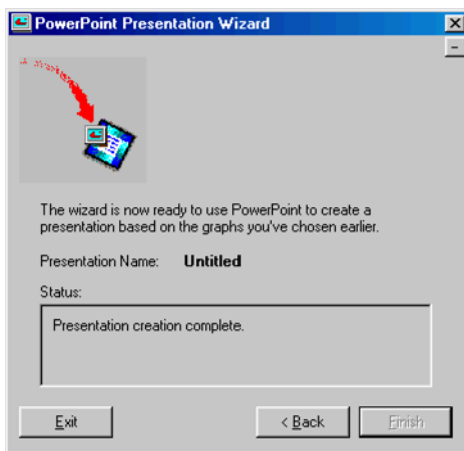


図 12.23 : プレゼンテーションの完成

PowerPoint が起動し、選択したグラフが新しい PowerPoint プレゼンテーションに指定した順序でスライドとして挿入されます。グラフが挿入される時、ステータス情報がウィザード内のボックスに表示されます。

7. プレゼンテーションが完成したら、**Exit** をクリックします。

第 12 章 他のアプリケーションでの S-PLUS の使用

ウィザードでプレゼンテーションリストを保存した場合、PowerPoint は自動的に新しいプレゼンテーションを同じ名前を使って保存します。ウィザードでプレゼンテーションリストを保存せず、**Presentation Name** が **Untitled** になっている場合は、PowerPoint でそのプレゼンテーションリストを明示的に保存しなければなりません。

第 13 章 S-PLUS セッションの カスタマイズ

はじめに	596
デフォルトと設定の変更	597
オブジェクトデフォルト	597
全般の設定	598
コマンドライン	609
「元に戻す」と「履歴」	610
テキスト出力	610
文書の背景色	611
グラフオプション	611
グラフスタイル	615
カラースキーム	620
自動再描画	623
起動時と終了時のセッションのカスタマイズ	624
起動オプションの設定	624
終了オプションの設定	627

はじめに

S-PLUS には、作業環境をカスタマイズする多くのオプションがあります。この章では、セッション全般の基本設定を設定する方法ならびに以下の操作について説明します。

- S-PLUS オブジェクトのユーザ専用のデフォルト設定。
- コマンド・ウィンドウで使用されるフォントの選択。
- 「元に戻す」と「履歴」ログの基本設定の指定。
- テキスト出力表示の制御。
- オブジェクト・エクスプローラや他のウィンドウの背景色の変更。
- グラフの様々なオプション、スタイルおよび配色の詳細なカスタマイズ。
- セッションを開始または終了するたびに、S-PLUS に自動的に一定のオプションを設定させたり、ある一定のタスクを実行させたりする。

また、カスタマイズ可能なメニューとツールバーを使用して、S-PLUS インタフェース自体を調整することもできます。メニューとツールバーの詳細は、『Application Developer's Guide』の第 11 章「Extending the User Interface」を参照してください。

デフォルトと設定の変更

オブジェクト デフォルト

S-PLUS では、記号、プロット、タイトル、グラフシートおよびデータオブジェクトを含む任意のタイプのオブジェクトのユーザ専用のデフォルト設定を定義することができます。以下の操作で行います。

1. デフォルトを定義したいタイプのオブジェクトを選択します。
2. オブジェクトのプロパティを修正して、そのタイプのオブジェクトのデフォルト設定として保存したい内容と一致させます。
3. 変更を保存します（たとえば、ダイアログの **OK** をクリックする）。
4. オブジェクトをまだ選択していない場合は、そのオブジェクトを選択します。
5. 以下のいずれかの操作を行います。
 - メインメニューから **オプション ▶ 選択したオブジェクトをデフォルトにする** を選択します。
 - オブジェクトを右クリックし、ショートカットメニューから **Save [Object] as default** を選択します。

選択したオブジェクトの実際の名前が、メニューオプションの **[Object]** と置き換わります。たとえば、グラフ上の **x** 軸タイトルを選択してフォントを変更すると、メニューオプションには **Save X Axis Title as Default** と表示されます。複数のオブジェクトを選択すると、メニューオプションには **Save Selected Objects as Default** と表示されます。

これで、選択したオブジェクトのプロパティがデフォルト値として保存されます。次にこのタイプのオブジェクトを作成するときに、これらの新しいデフォルト値が使用されます。

全般の設定

セッション全般の設定を指定するには、メインメニューから**オプション ▶ 設定**を選択します。**General Settings** ダイアログは、後で説明するような 4 つのタブ付きページから成ります。

General

図 13.1 に、**General Settings** ダイアログの **General** ページを示します。

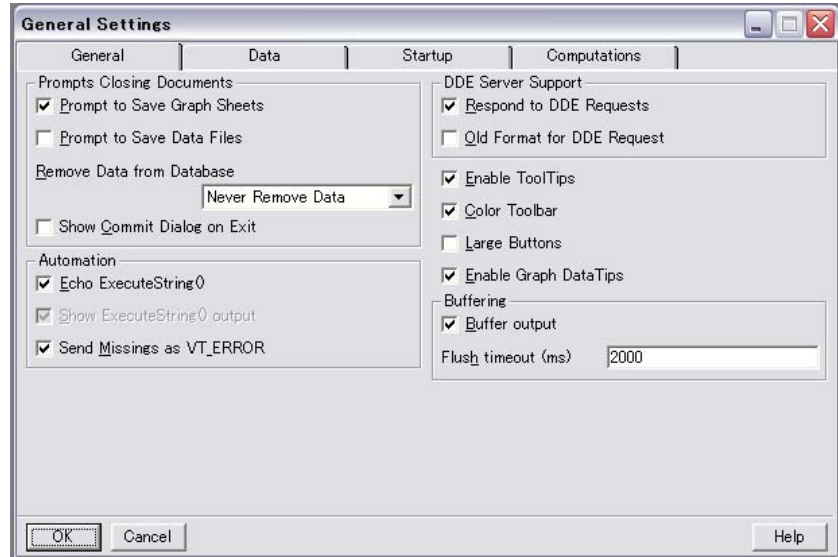


図 13.1 : General Settings ダイアログの General ページ

Prompts Closing Documents グループ

Prompt to Save Graph Sheets このオプションを選択すると、新規または修正したグラフシートを表示しているグラフシート・ウィンドウを閉じるときに、ダイアログプロンプトが出されます。

注意

グラフシートは、スクリプトやレポートと同じようにそのセッションでのみ存在する一時的な「ドキュメントオブジェクト」です。プロンプトダイアログは、グラフシート保存を忘れないよう表示されます。グラフシートを永久的に保存するときは、拡張子 (.sgr) のファイルで保存しなければなりません。このチェックボックスのチェックを外しても、グラフシート・ウィンドウを閉じる前にメインメニューから**ファイル ▶ 保存**を選択することによって、グラフシートを保存することができます。

Prompt to Save Data Files 新規または修正したデータセットを表示しているデータ・ウィンドウを閉じるときにダイアログプロンプトが出されます。

注意

グラフシートと違って、データセットは、「作業データベース」と呼ばれる特別な内部データベースに自動的に永久に保存されます。データを管理する最も簡単な方法は、S-PLUS にデータを自動保存させることです。ただし、必要に応じてデータセットを拡張子 (.sdd) のファイルとして保存することができます。その場合、プロンプトダイアログが、データを保存するように確認してきます。このチェックボックスのチェックを外しても、データ・ウィンドウを閉じる前にメインメニューからファイル ▶ 保存を選択することによって、データセットをファイルとして保存することもできます。

Remove Data from Database セッションを終了するときそのセッションで作成または修正したデータオブジェクトが、作業データ内でどのように処理されるかを制御するオプションを選択します。

注意

このフィールドのデフォルトは **Never Remove Data** です。しかし、常に外部ファイルにデータセットを保存するように決定した場合は、再び外部ファイルを開くときにデータベースオブジェクトを上書きしないように、**Always Remove Data** を選択した方がよいでしょう。

Show Commit Dialog on Exit S-PLUS を終了するとき、図 13.2 に示すようなオブジェクト変更の保存ダイアログが表示されます。

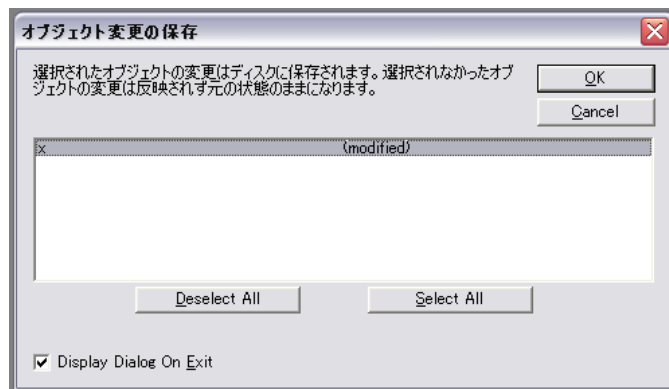



図 13.2 : オブジェクト変更の保存ダイアログ

オブジェクト変更の保存ダイアログで、プログラムを閉じる前に任意の新規や変更したデータオブジェクトを保存または破棄することができます。また、ダイアログの下部にある **Display Dialog On Exit** チェックボックスのチェックを外すと、この機能をオフにすることができます。

ヒント

このダイアログをいつでも表示させておくためには、標準ツールバーの**データオブジェクトを元に戻す**ボタン  をクリックします。

Automation グループ

Echo ExecuteString() このオプションを選択すると、コマンド・ウィンドウを開いたとき、オートメーションメソッド `ExecuteStrings()` で実行されるコマンド文字列がコマンド・ウィンドウのコマンドラインに表示されます。このオプションを選択すると、**Show ExecuteString() output** オプションが自動的に利用できなくなります。

Show ExecuteString() output `ExecuteString()` に渡された文字列を実行した結果の出力が、**Text Output Routing** (610 ページを参照) で指定された出力ウィンドウに表示されます。

Send Missings as VT_ERROR オートメーションクライアントサポートが、データオブジェクト内の欠損値を変量のタイプ `VT_ERROR` として送ります。このオプションは、デフォルトで選択されており、**Visual Basic** などのプログラムが欠損値をエラー値として解釈するため重要です。

DDE Server Support グループ

Respond to DDE Requests このオプションを選択すると、S-PLUS が DDE (Dynamic Data Exchange) クエリに応答します。詳細は、『**Application Developer's Guide**』の第 6 章「**Calling S-PLUS Using DDE**」を参照してください。

Old Format for DDE Request S-PLUS は、DDE に応答するときに S-PLUS の旧バージョンからのテキストフォーマットスタイルを使用します。

その他

Enable ToolTips マウスをツールバーとパレットのボタンの上においたときに、小さなポップアップウィンドウ内にボタンラベルが表示されます。

Color Toolbar ツールバーがカラーで表示されます。選択しない場合は、ツールバーは黒白で表示されます。

Large Buttons 大きいツールバーボタンが表示されます。デフォルトでは、小さいツールバーボタンが表示されます。

Enable Graph DataTips グラフシート上のデータ点にマウスをおいたときに、小さいポップアップウィンドウにデータの情報が表示されます。

Buffering グループ

Buffer output 出力時に、出力をバッファして最後に出力します。Flush timeout で指定された時間より出力間隔が長くなると適用されます。

Flush timeout 出力時に出力をバッファする間隔を指定します。マイナスの場合は出力は最後にまとめて行われ、ゼロにすると出力は逐次行われます。

Data

図 13.3 に **General Settings** ダイアログの **Data** ページを示します。

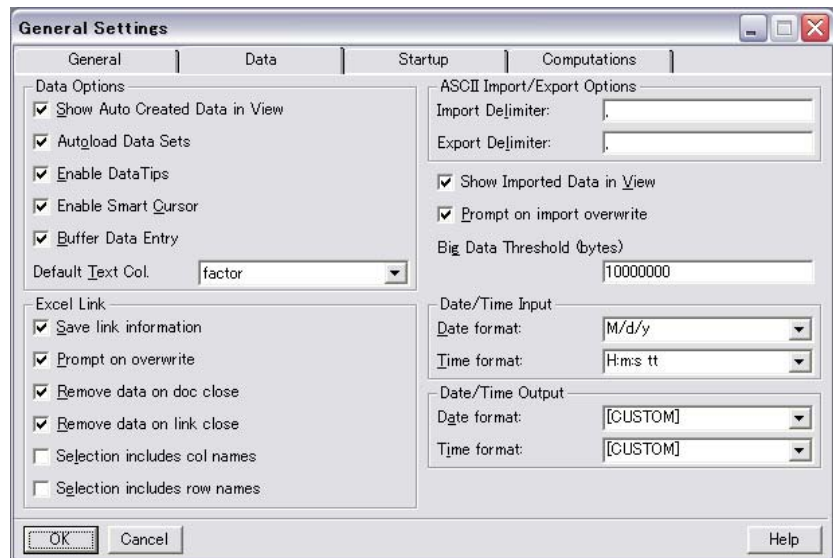


図 13.3 : General Settings ダイアログの Data ページ

Data Options グループ

Show Auto Created Data in View このオプションを選択すると、**Data** ダイアログか **Statistics** ダイアログによってデータが作成されると、そのデータがデータ・ウィンドウに自動的に表示されます。このオプションをオフにすると、データオブジェクトはオブジェクト・エクスプローラに表示されます。

Autoload Data Sets グラフシートを作成するときに対応するデータがデータ・ウィンドウに自動的にロードされます。

Enable DataTips マウスをデータ・ウィンドウの列名の上に置いたときに小さいポップアップウィンドウ内に列の説明（ある場合）が表示されます。

Enable Smart Cursor セルにデータを入力している間に ENTER キーを押すと、カーソルがその前に動いていた方向に移動します。

Buffer Data Entry 特別なバッファを使ってデータ編集を高速化します。このオプションは選択したままにすることをお勧めします。

Default Text Col. デフォルトでは、データ・ウィンドウ内の空白の列に文字データを入力するときに、factor 形式の列が作成されます。デフォルトで character 形式の列を作成するには、このフィールドで **character** を選択してください。

Ragged data.frame デフォルト設定では、データセットの列はすべて等しい長さになるよう短い列が NA で埋められ、長さが均一にされます。この操作を中止したい場合にこのオプションを選択します。

Excel Link グループ

Save link information このオプションを選択すると、リンク情報がソースオブジェクトに保存されます。ソースが Excel ドキュメントの場合、リンク情報は、データ範囲の左上のセルにコメントとして書き込まれます。ソースが S-PLUS データフレームの場合、リンク情報は、データフレームの属性として保存されます。

Prompt on overwrite Excel ドキュメントか S-PLUS データフレームにデータを上書きするときにどちらが特定リンクのターゲットであってもプロンプトを出します。

Remove data on doc close 現在のドキュメントを閉じるときにリンクされているデータが削除されます。

Remove data on link close 現在のリンクを閉じるときにリンクされているデータが削除されます。

Selection includes col names 自動的に選択データ範囲の最初の行が、列名の行として解釈されます。

Selection includes row names 自動的に選択データ範囲の最初の列が、行名の列として選択されます。

ASCII Import/Export Options グループ

Import Delimiter ASCII 文字をインポートするための区切り記号を指定します。デフォルトの区切り記号はカンマです。

Export Delimiter ASCII 文字にエクスポートするための区切り記号を指定します。デフォルトの区切り記号はカンマです。

その他

Show Imported Data in View 自動的にデータ・ウィンドウが開いて、インポートしたデータを表示します。

Prompt on import overwrite データベースに既に保存されているデータセットと同じ名前を付けたデータファイルをインポートしようとしたときに、上書き確認が求められます。

Date/Time Input グループ

Date format 日付データを作成し表示するときに使用したいフォーマットを選択します。使用可能な選択肢は、Windows のものと同じです。このフィールドのデフォルト値は、そのときの Windows のデフォルト値です。

Time format 時間データを作成し表示するときに使用したいフォーマットを選択します。使用可能な選択肢は、Windows のものと同じです。このフィールドのデフォルト値は、そのときの Windows のデフォルト値です。

注意
Date format フィールドと Time format フィールドで行った選択は、そのときのセッションでのみ維持されます。

Date/Time Output グループ

Date format 日付データを出力するときに使用したいフォーマットを選択します。使用可能な選択肢は、Windows のものと同じです。このフィールドのデフォルト値は、そのときの Windows のデフォルト値です。

Time format 時間データを出力するときに使用したいフォーマットを選択します。使用可能な選択肢は、Windows のものと同じです。このフィールドのデフォルト値は、そのときの Windows のデフォルト値です。

注意
Date format フィールドと Time format フィールドで行った選択は、そのときのセッションでのみ維持されます。

Startup

図 13.4 に **General Settings** ダイアログの **Startup** ページを示します。

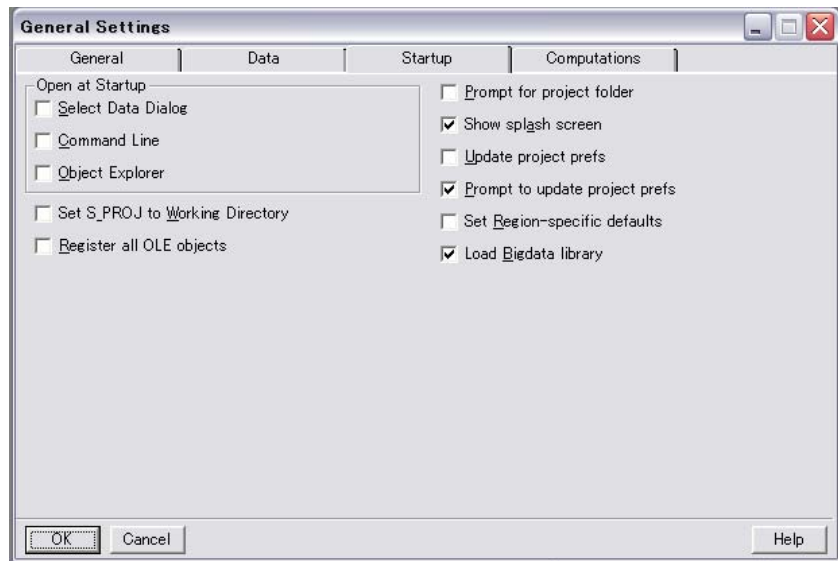


図 13.4 : General Settings ダイアログの Startup ページ

Open at Startup グループ

起動したときに、**Select Data** ダイアログ/コマンド・ウィンドウ/デフォルトのオブジェクト・エクスプローラを S-PLUS が自動的に開くようにするかどうかを選択します。

その他

Set S_PROJ to Working Directory このオプションは、プロジェクトフォルダの位置を決定するときに **S-PLUS Properties** ダイアログの **Shortcut** ページと関連して動作します。(Shortcut ページを開くときは、デスクトップ上の S-PLUS 起動アイコンを右クリックし、プロパティを選択し、ショートカットタブをクリックしてください。)

注意

デフォルトでは、このオプションは選択されておらず、ショートカットページの**作業フォルダ**フィールドは無視されます。プロジェクトフォルダは、以下のいずれかによって決定されます。

- デフォルトの動作。
- ショートカットページの**リンク先**フィールドの **S_PROJ** の値。
- ショートカットページの**リンク先**フィールドの **S_DATA** の値。

Set S_PROJ to Working Directory を選択すると、プロジェクトフォルダは、**Shortcut** ページの**作業フォルダ**フィールドの内容によって決定されます。

混乱を避けるため、方法を混合しないことをお勧めします。すなわち、以下の「いずれか」の方法を実行してください。

- プロジェクト/データフォルダの位置を指定するときは、このオプションを選択せず、**リンク先**フィールドの **S_PROJ** や **S_DATA** を使用します。
- **リンク先**フィールドの **S_PROJ** や **S_DATA** を指定せずにプロジェクトフォルダの位置を指定するときは、このオプションを選択し、**作業フォルダ**フィールドだけを使用します。

Register all OLE objects リンクしたか埋め込んだすべてのオブジェクト (OLE オブジェクト) が登録されます。

Prompt for project folder S-PLUS を起動するたびに図 13.5 に示すダイアログが開き、セッションに使用したいプロジェクトフォルダを指定することができます。この機能をオフにし、S-PLUS を起動するたびにデフォルトのプロジェクトフォルダを使用するためには、このチェックボックスのチェックを外すか、ダイアログの **Always start in this project** チェックボックスを選択します。

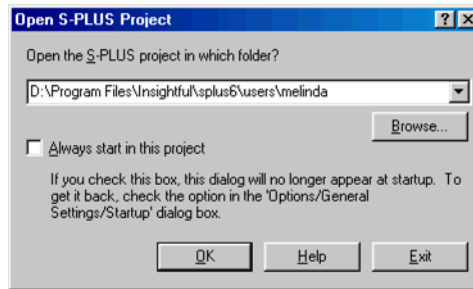


図 13.5 : Open S-PLUS Project ダイアログ

Show splash screen S-PLUS を起動したときに S-PLUS スプラッシュ画面が現われます。この機能をオフにするには、このチェックボックスのチェックを外します。

Update project prefs このチェックボックスは、起動時にプロジェクトの.Prefs フォルダを、S-PLUS プログラムフォルダの **MasterPrefs** フォルダにインストールされている最新の基本設定に更新するかどうかを制御します。このオプションを選択すると、**.Prefs** フォルダの更新が必要かどうかを示すダイアログが表示されます。ファイルを更新するように選択した場合は、バックアップフォルダが作成され、オリジナルのファイルは更新前にそのフォルダにコピーされます。

Prompt to update project prefs Prefs のアップデートを促すダイアログを表示させるときは選択します。

Set Region-specific defaults **time.out.format**、**time.in.format**、**time.zone** のオプション設定を Windows の地域設定されている値にする場合、選択します。

Computations

図 13.6 に、**General Settings** ダイアログの **Computations** ページを示します。

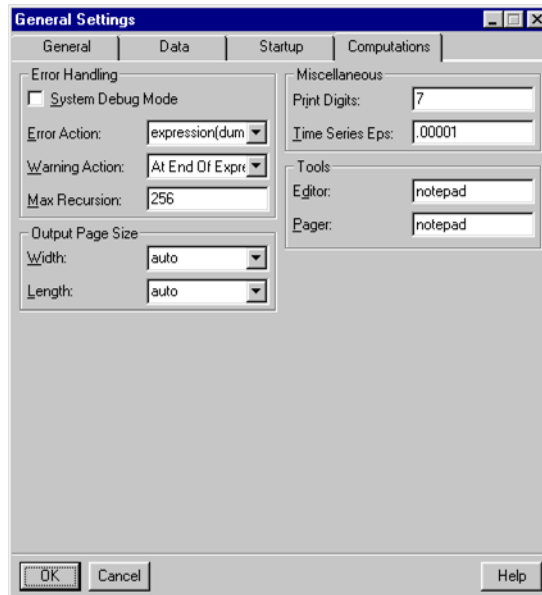


図 13.6 : **General Settings** ダイアログの **Computations** ページ

Error Handling グループ

System Debug Mode このオプションを選択すると、S-PLUS は評価中に様々な内部チェックを行います。このオプションは、警告メッセージとリロードに関する詳しい情報を提供し、S-PLUS が異常終了するような不可解なバグを探すのに役立つことがあります。このオプションを有効にすると評価が実質的に遅くなり、しばしば予想外の動作を引き起こすことがあるため注意してください。

Error Action エラーまたは割込みが生じたときに、呼び出す関数(引数なし)を選択します。S-PLUS では、未処理の関数呼出しや関連フレーム全体をダンプするために `dump.calls` と `dump.frames` を使用することができます。(これらの関数の詳細は、オンラインヘルプを参照してください。) 関数を NULL に設定すると、すべてのエラー動作が表示されなくなります。

Warning Action 警告をいつ表示させるかを選択します。

Max Recursion 式を入れ子にすることができる最大深さを指定します。このオプションは、主に関数のそれ自体への直接的、間接的な再帰呼出しによるプログラムの暴走を避けるためのものです。

Output Page Size グループ

これらの 2 つの設定値はすべての出力ウィンドウに適用され、そのプロジェクトの S-PLUS セッションにまたがって維持されます。デフォルトでは、**Width** と **Length** が、自動的にそのときの出力ウィンドウ（スクリプト・ウィンドウとレポート・ウィンドウならびにコマンド・ウィンドウを含む）の大きさに設定されますが、これらのフィールドに自分の値を入力することによって、このデフォルトの動作を無効にすることができます。

Miscellaneous グループ

Print Digits 出力時に使用する有効数字の数を指定します。この値を 17 に設定すると標準の倍精度数の長さになります。

Time Series Eps 時系列の比較の許容範囲を指定します。この小さい数は、周期の比較のために時系列関数全体に使用されます。絶対値の差がここで指定した数よりも小さい場合は、周期が等しいと見なされます。

Tools グループ

Editor 関数 `edit` で使用されるデフォルトのテキストエディタコマンドを指定します。選択したエディタはすべて `Notepad` 呼び出しのスタイルで、`Notepad` ファイル名のコマンドと、その後の編集コマンドの読みによって呼び出されます。異なる呼び出しまたは異なるユーザ対話形式のエディタは使用しないでください。

Pager 関数 `help` および `page` によって使用されるデフォルトのページャプログラムを指定します。選択したページャは、`pager` ファイル名として呼び出され、ファイル名から読み込まれなければなりません。

コマンド ライン

コマンドラインオプションを指定するには、メインメニューから**オプション ▶ コマンドライン**を選択します。**Command Line Options** ダイアログは、後で説明するような2つのタブ付きページから成ります。

Font

図 13.7 に **Command Line Options** ダイアログの **Font** ページを示します。

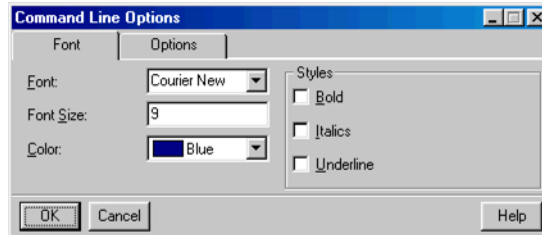


図 13.7 : **Command Line Options** ダイアログの **Font** ページ

コマンド・ウィンドウで使用されるフォント、フォントサイズ、色およびスタイルを指定するときは、このページのオプションを使用します。適正な配置で出力させるには、固定幅フォントをお勧めします。

Options

図 13.8 に **Command Line Options** ダイアログの **Options** ページを示します。

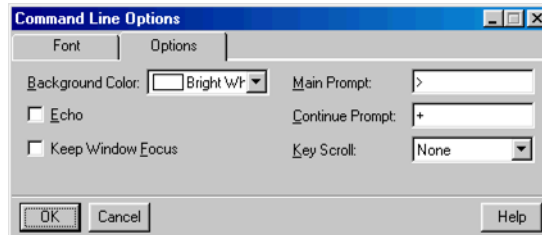


図 13.8 : **Command Line Options** ダイアログの **Options** ページ

Background Color コマンド・ウィンドウで使用される背景色を選択します。

Echo このオプションを選択すると、評価される前にそれぞれの完全な式が繰り返されます。

Keep Window Focus コマンドを実行するときにコマンド・ウィンドウにフォーカスが残ります。

Main Prompt 式を要求するために使用される文字列を指定します。デフォルトは > です。

Continue Prompt 式の続きを要求するときに使用される文字列を指定します。デフォルトは + です。

Key Scroll コマンド・ウィンドウをスクロールする方法を選択します。

「元に戻す」と「履歴」

「元に戻す」と「履歴」のオプションを指定するには、メインメニューから **オプション ▶ 元に戻す+履歴** を選択します。**Undo and History** ダイアログを図 13.9 に示します。

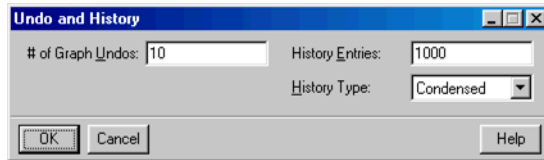


図 13.9 : Undo and History ダイアログ

of Graph Undos グラフオブジェクトの元に戻すキューに保存する元に戻す最大数を指定します。元に戻す数が多いほどメモリの使用量が多くなります。

History Entries 履歴ログに保存するエントリの最大数を指定します。エントリの数が多いほどメモリの使用量が多くなります。

History Type 簡単な履歴を表示する場合は **Condensed** を指定し、詳しい履歴を表示させる場合は **Full** を選択します。

テキスト出力

テキスト出力設定値を指定するには、メインメニューから **オプション ▶ テキスト出力オプション** を選択します。図 13.10 に **Text Output Routing** ダイアログを示します。

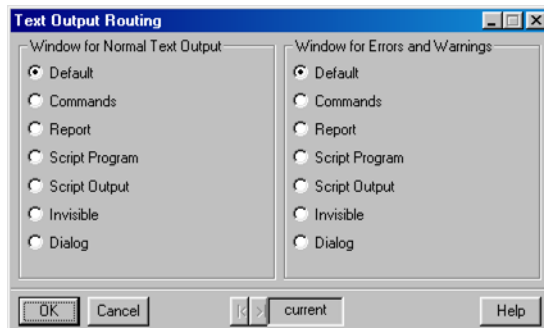


図 13.10 : Text Output Routing ダイアログ

通常テキストおよびエラーと警告の出力ウィンドウの基本設定を指定するときは、このページのオプションを使用します。

文書の背景色

文書の背景色を指定するには、メインメニューから**オプション ▶ バックグラウンドカラー**を選択します。図 13.11 に **Document Background Colors** ダイアログを示します。

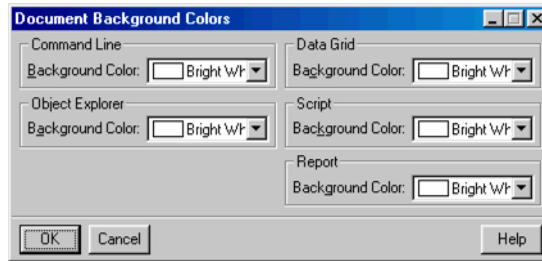


図 13.11 : Document Background Colors ダイアログ

コマンド・ウィンドウ、オブジェクト・エクスプローラ、データ・ウィンドウ、スクリプト・ウィンドウおよびレポート・ウィンドウの背景色の基本設定を指定するときは、このページで選択します。

グラフ オプション

グラフオプションを指定するときは、メインメニューから**オプション ▶ グラフオプション**を選択します。**Graphs** ダイアログは、後で説明する 3 つのタブ付きページから成ります。

Options

図 13.12 に **Graphs** ダイアログの **Options** ページを示します。

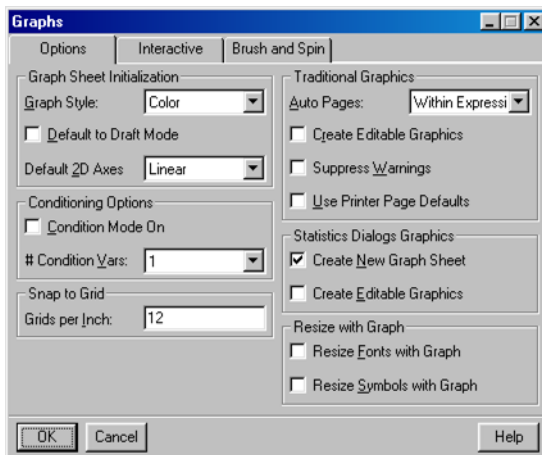



図 13.12 : Graphs ダイアログの Options ページ


Graph Sheet Initialization グループ

Graph Style グラフの **Color** または **Black and White** を選択します。

Default to Draft Mode このオプションを選択すると、グラフが画面にドラフトモードで表示され、再描画時間が大幅に短縮されます。ドラフトモードは画面の解像度だけが変化し、印刷出力は常に印刷品質であることに注意してください。このオプションのオン・オフを切り換えるときは、**標準**ツールバーのドラフトモードのオン・オフボタン  をクリックするか、メインメニューから表示 ▶ **ドラフトモードのオン・オフ** を選択します。

Default 2D Axes デフォルトとして使用する 2 次元軸のタイプを選択します。ここで行った選択は**標準**ツールバーのデフォルトタイプになることに注意してください。特別に指定しない限り、デフォルトのタイプは **Linear** です。

Conditioning Options グループ

Condition Mode On 条件付けモードは、選択したデータをプロットボタンによるプロットの作成にどのように使用するかを決定します。条件付けモードがオンのときは、最後に選択した列がマルチパネルグラフの条件付け変数として使用されます。条件付けに使用される列の数は、下の**#Condition Vars** フィールドで指定されます。(マルチパネルグラフの詳細は、『Application Developer's Guide』の第 3 章「Traditional Trellis Graphics」を参照してください。) このオプションのオン・オフを切り換えるときは、**標準**ツールバーの**条件モードのオン・オフ**ボタン  をクリックしてください。

#Condition Vars 条件付けモードがオンのときに、条件付け変数として使用する列の数を指定します。ここで行った選択が**標準**ツールバーのデフォルト値になることに注意してください。ただし、この数字は、ドロップダウンリ

ストから違う値を選択することによっていつでも変更することができます。

Snap to Grid グループ

Grids per Inch Snap to Grid オプションに使用する隠れたグリッド線の数を指定します。デフォルトのグリッド線は、1 インチ当たり 12 本または 1 センチメートル当たり 5 本です。**Snap to Grid** が与えられているときは、オブジェクトは、そのような隠れた横グリッド線と縦グリッド線の最も近い交点に“スナップ (Snap)”します。

Traditional Graphics グループ

Auto Pages S-PLUS 関数内で、一連のプロットを作成したときにデフォルトでページを自動的に追加するには、**Within Expression** を選択します。
Graph Sheet ダイアログの **Options** ページの **Page Creation** フィールドで、この設定を無効にすることができます。

Create Editable Graphics このオプションを選択すると、S-PLUS 関数内で作成されたプロットが、**グラフシート**上で編集可能なグラフオブジェクトに変換されます。このオプションがオンのとき、グラフの作成がかなり遅くなることに注意してください。このオプションがオフのときは、合成グラフオブジェクトが**グラフシート**に作られ、ショートカットメニューから **Convert to Objects** を右クリックして選択することによって、いつでも編集可能なグラフオブジェクトに変換することができます。

Suppress Warnings デフォルトで作成される新しい `graphsheets()` がすべて `par(err=-1)` に設定され、“境界外の点”メッセージのような警告を出さないようにします。

Use Printer Page Defaults プリンタページのデフォルトが使用されます。

Statistics Dialogs Graphics グループ

Create New Graph Sheet このオプションを選択すると、統計ダイアログを使用してプロットを作成するときに、新しい**グラフシート**が作成されます。

Create Editable Graphics 統計ダイアログを使用して作成したプロットは、編集可能なグラフになります。

Resize with Graph グループ

Resize Fonts with Graph グラフのサイズを変更したときにグラフ内のタイトル、コメント、その他の文字のサイズが変更されます。

Resize Symbols with Graph グラフのサイズを変更したときにグラフ内の記号のサイズが変更されます。白抜き四角形、塗りつぶした四角形、楕円形などの形は、この設定に関係なくともサイズが変更されます。

Interactive

図 13.13 に **Graphs** ダイアログの **Interactive** ページを示します。

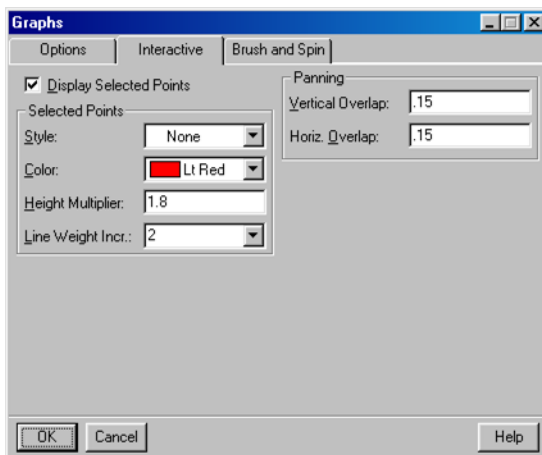


図 13.13 : **Graphs** ダイアログの **Interactive** ページ

Display Selected Points グラフシート内の選択したデータ点がハイライトされます。データ・ウィンドウ内の希望する行を選択するか、**グラフツール**パレットの**矩形内のデータ選択**ツールを使ってデータ点を選択することができます。選択した点の外観を指定するには、次のグループのオプションを使用します。

Selected Points グループ

Style ドロップダウンリストからスタイルを選択します。**None** を選択すると、選択した点はハイライトされますが、別のスタイルに変更できません。

Color ドロップダウンリストから色を選択します。

Height Multiplier 点の高さの倍率を指定します。デフォルト値は1.8です。

Line Weight Incr. 実線でないスタイルの線の太さを指定します。

Panning グループ

Vertical Overlap と **Horiz. Overlap** フィールドを使って、**グラフツール**パレットの縦と横の **Pan** ボタンに使用する 0~0.990 の間の数を指定します。小さい数ほど重なりが少なく、大きい数ほど重なりが大きくなります。デフォルト値は0.15です。

Brush and Spin 図 13.14 に **Graphs** ダイアログの **Brush and Spin** ページを示します。

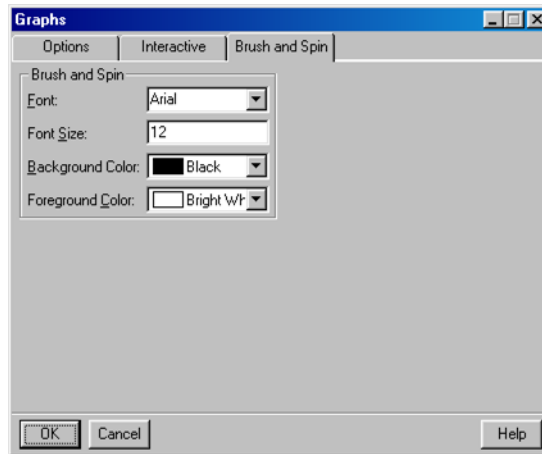


図 13.14 : Graphs ダイアログの Brush and Spin ページ

Font フィールド、**Font Size** フィールド、**Background Color** フィールドおよび **Foreground Color** フィールドを使って、**Brush and Spin** ウィンドウに使用する設定を指定します。

グラフ スタイル

グラフスタイルは、新しいグラフシートのプロパティを初期化するために使用されます。**Color** と **Black and White** の2つのグラフスタイルを定義することができます。グラフスタイルを指定するには、メインメニューから**オプション ▶ グラフスタイル**を選択し、次に**色**か**白黒**を選択します。

Color Style ダイアログと **Black and White Style** ダイアログで選択したオプションは、**Graph Sheet** ダイアログで変更することができます。また、メインメニューから**書式 ▶ スタイルの適用**を選択して、特定のスタイルの仕様に合わせるように**グラフシート**を修正することができます。

Color Style ダイアログと **Black and White Style** ダイアログは、使用する“色”の他は同じであるため（後者の場合は、黒、白、および中間調）、この節では、**Color Style** ダイアログで使用する様々なオプションについて説明します。

Color Style ダイアログは、後で説明する5つのタブ付きページから成ります。

Options

図 13.15 に **Color Style** ダイアログの **Options** ページを示します。

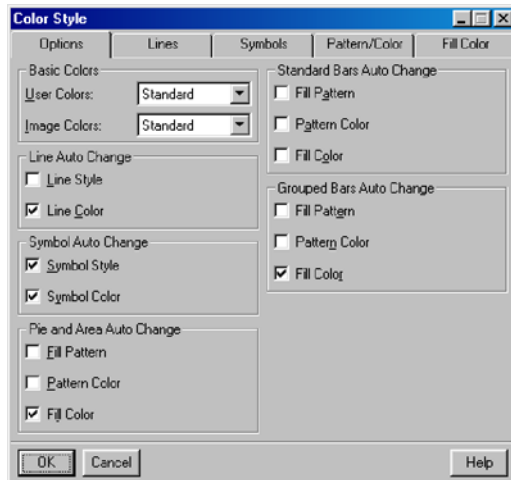


図 13.15 : Color Style ダイアログの Options ページ

Basic Colors グループ

スタイルに使用する **User Colors** と **Image Colors** の配色を選択します。これらの配色に使用されるカラーパレットの編集については、620 ページを参照してください。

Line Auto Change グループ

同じグラフに線グラフを追加するたびに線のプロパティを変化させるには、**Line Style/Line Color** を選択します。線種と色は、このダイアログの **Lines** ページで指定した順序で繰り返し用いられます。

Symbol Auto Change グループ

記号を含むプロットを同じグラフに追加するたびに記号のプロパティを変化させるには、**Symbol Style/Symbol Color** を選択します。記号のスタイルと色は、このダイアログの **Symbols** ページで指定した順序で繰り返し用いられます。いずれも指定されない場合、デフォルトでは最初の記号のスタイルと色が使用されます。

Pie and Area Auto Change グループ

新しく作成した円グラフやエリアチャート内の、円グラフの各扇形または各エリアごとにプロパティを変化させるには、**Fill Pattern/Pattern Color/ Fill Color** を選択します。塗りつぶしパターン、パターン色、塗りつぶし色がこのダイアログの **Pattern/Color** ページと **Fill Color** ページで指定した順序で繰り返し用いられます。いずれも指定されない場合、デフォルトでは最初の塗りつぶしパターン、パターン色および塗りつぶし色が使用されます。

Standard Bars Auto Change グループ

新しく作成した棒グラフの各バーのプロパティを変化させるには、**Fill Pattern/Pattern Color/ Fill Color** を選択します。塗りつぶしパターン、パターン色、塗りつぶし色が、このダイアログの **Pattern/Color** ページと **Fill Color** ページで指定した順序で繰り返し用いられます。いずれも指定されない場合は、デフォルトで最初の塗りつぶしパターン、パターン色および塗りつぶし色が使用されます。

Grouped Bars Auto Change グループ

新しく作成した類別棒グラフで各グループのそれぞれバーのプロパティを変化させるには、**Fill Pattern/Pattern Color/ Fill Color** を選択します。塗りつぶしパターン、パターン色、塗りつぶし色が、このダイアログの **Pattern/Color** ページと **Fill Color** ページで指定した順序で繰り返し用いられます。いずれも指定されない場合、デフォルトでは最初の塗りつぶしパターン、パターン色および塗りつぶし色が使用されます。

Lines

図 13.16 に **Color Style** ダイアログの **Lines** ページを示します。

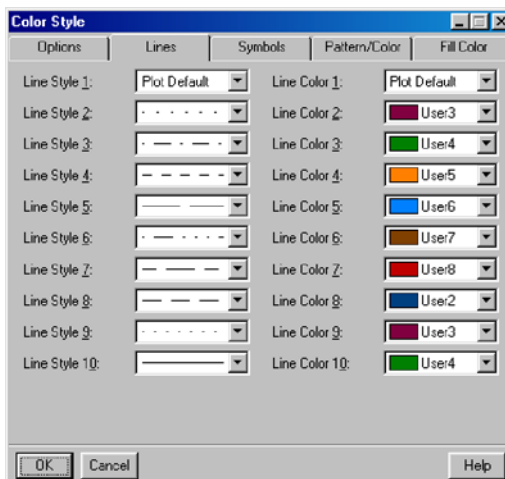


図 13.16 : Color Style ダイアログの Lines ページ

Lines ページには、10 個のスタイルフィールドと 10 個の色フィールドがあります。これらのフィールドを使って、繰り返し用いたい線の種類と色を順番に選択します。

Symbols

図 13.17 に **Color Style** ダイアログの **Symbols** ページを示します。

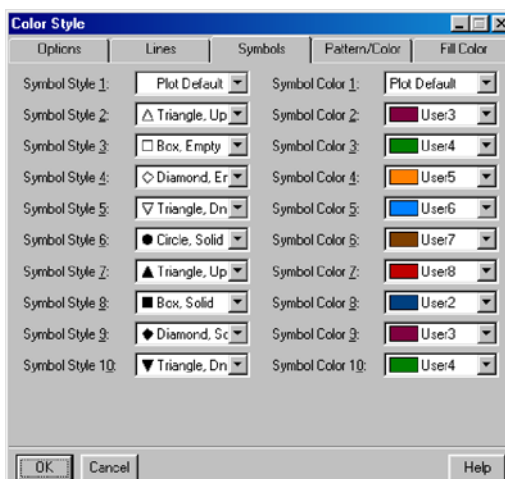


図 13.17 : Color Style ダイアログの Symbols ページ

Symbols ページには、10 個のスタイルフィールドと 10 個の色フィールドがあります。これらのフィールドを使って、繰り返し用いたい記号の種類と色を順番に選択します。

Pattern/Color

図 13.18 に **Color Style** ダイアログの **Pattern/Color** ページを示します。

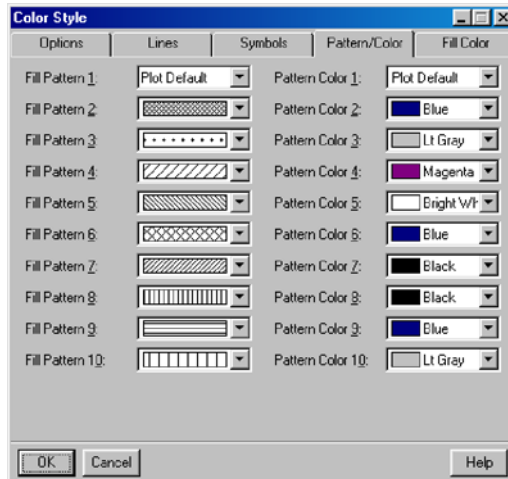


図 13.18 : Color Style ダイアログの Pattern/Color ページ

Pattern/Color ページには、10 個の塗りつぶしパターンフィールドと 10 個のパターン色フィールドがあります。これらのフィールドを使って、繰り返し用いたい塗りつぶしのパターンとパターン色を選択します。

Fill Color

図 13.19 に Color Style ダイアログの Fill Color ページを示します。

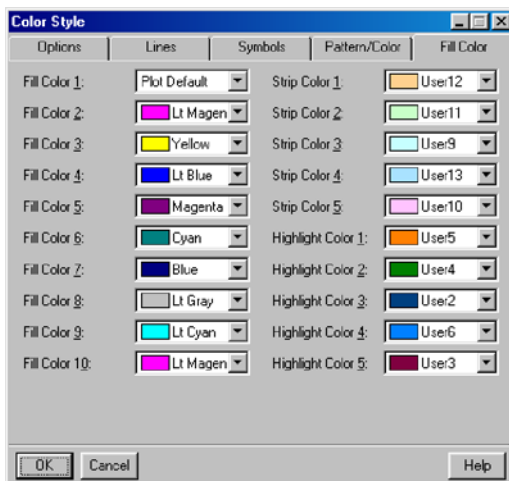


図 13.19 : Color Style ダイアログの Fill Color ページ

Fill Color ページには、10 個の塗りつぶし色フィールド、5 つのストリップ色フィールド (マルチパネルプロットのストリップラベル用)、および 5 つのハイライト色フィールドがあります。これらのフィールドを使って、繰り返し用いたい塗りつぶし色、ストリップ色およびハイライト色を選択します。

カラー スキーム

グラフスタイルを定義するとき、使用されるユーザ色と画像色に使用できる配色は 8 つあります。配色を指定するには、メインメニューから **オプション ▶ カラースキーム** を選択します。**Color Schemes** ダイアログは、後で説明する 2 つのタブ付きページから成ります。

User Colors

図 13.20 に **Color Schemes** ダイアログの **User Colors** ページを示します。

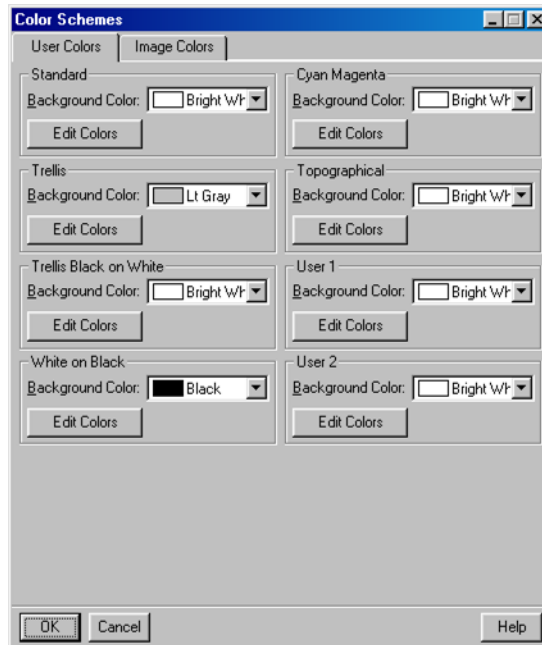


図 13.20 : **Color Schemes** ダイアログの **User Colors** ページ

Color Style ダイアログまたは **Black and White Style** ダイアログの **Options** ページで選択する **User Colors** の配色は、新しく作成するグラフシートのユーザ色を設定するために使用されます。これらの色は、グラフシート内のすべての図形オブジェクトのカラーリストに **User1**、**User2** などのように表示されます。

User Colors ページのフィールドを使って、8つの各配色の背景色を設定します。配色の **User Colors** を修正するときは、**Edit Colors** ボタンをクリックし、**Color** ダイアログを使ってカラーパレットを編集します。

Image Colors

図 13.21 に **Color Schemes** ダイアログの **Image Colors** ページを示します。

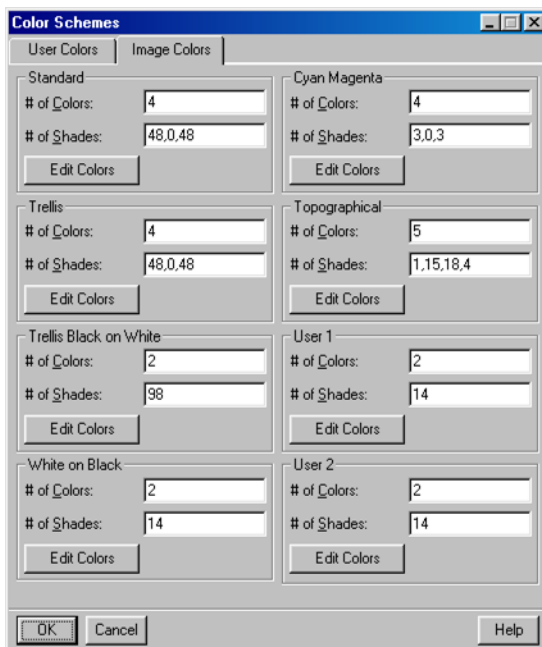


図 13.21 : **Color Schemes** ダイアログの **Image Colors** ページ

画像色は、イメージ鳥瞰図、カラー等高線、およびイメージプロットに使用できる一連の塗りつぶし色です。カラーの指定は、最大 16 のコア色と各コア色の階調または色グラデーションの数を定義するリストからなります。

Image Colors ページのフィールドを使って、8 つの各配色の色と階調の数を設定します。配色の **Image Colors** を修正するには、**Edit Colors** ボタンをクリックし、**Color** ダイアログを使ってコア画像色のカラーパレットを編集します。

of Colors 画像色の定義に使用するコア色の数を指定します。画像の配色に使用されるのは、このフィールドで指定した色の数だけです。

of Shades このフィールドは、各コア色の間で使用される階調の数を示すカンマで区切られた数字のリストで定義されます。たとえば、黒、赤および白の 3 つのコア色と、階調の数に “5,15” を指定すると、黒、赤および白と、黒と赤の間の 5 つの階調と、赤と白の間の 15 の階調の合計 23 の色が画像の配色に使用されます。

自動再描画

変更を行うたびにプロットを自動的に再描画するかどうかを選択することができます。この機能は、デフォルトでオンにされていますが、計算に時間のかかるプロットや複雑なプロットの場合は、再描画時間を節約するためにこの機能をオフにしたほうがよい場合があります。

Auto Plot Redraw 機能は、メインメニューから表示 ▶ **自動再描画**を選択することによってオン・オフを切り換えることができます。この機能をオフにし、プロットを再描画したいときは、メインメニューから表示 ▶ **今すぐ再描画**を選択します。

起動時と終了時のセッションのカスタマイズ

起動オプションの設定

S-PLUS を起動するときにいくつかのオプションを設定する場合や、ライブラリまたは S-PLUS チャプターを自動的に追加したい場合は、それらの選択を保存し、プログラムが起動するたびに S-PLUS にそれらの選択を自動的に設定させることができます。

S-PLUS を起動するとき、以下のような初期設定ステップが行われます。

1. 基本初期設定によって、評価プログラムが式を評価することができるようにします。
2. 標準初期設定ファイル **%SHOME%/S.init** を探します。これは、S-PLUS 式を含むテキストファイルです。デフォルトの初期設定ファイルが、この後のステップを実行します。
3. システム管理者がファイル **%SHOME%/local/S.init** にその場でカスタマイズを行った場合は、そのファイル内のアクションが評価されます。これを編集して、すべてのプロジェクトの起動時に実行される動作を設定することができます。
4. ファイル **%SHOME%/S.chapters** を探します。このファイルは、すべてのユーザに追加されるライブラリまたは S-PLUS チャプターのパスを含むテキストファイルです。デフォルトでは、基本初期設定の際に標準の S-PLUS ライブラリだけが追加されるため、このファイルは存在しません。これは、すべてのプロジェクトに影響を及ぼします。
5. 現行フォルダ内のユーザ個人の **S.chapters** ファイルを探します。このファイルに、そのときのプロジェクトの起動時に追加したいライブラリ、または S-PLUS チャプターをリストします。
6. 作業データベースを決定します。
7. 現行フォルダ内にカスタム化ファイル **S.init** があった場合は、そのファイルが評価されます。**S.init** ファイルは、セッションの開始時に実行される S-PLUS 式を含むテキストファイルです。このファイルは、すべてのユーザのセッションに影響を及ぼす **%SHOME%/S.init** とは異なることに注意してください。
8. ローカルシステム初期設定関数 **.First.local** がある場合は、その評価を含む関数 **.First.Sys** が評価されます。

9. S-PLUS は、環境変数 `S_FIRST` が設定されている場合にはそれを評価し、あるいはステップ 3~5 で設定した探索パスで見つかった最初の関数 `.First` を評価します。

ほとんどの場合、初期設定プロセスには、上記のステップ 7 と 9 の一方だけが含まれています。したがって、起動オプションを設定するときには、以下の方法のいずれかを使用することになるでしょう。

- 希望するオプションを含む `.First` という名前の S-PLUS 関数を作成します。
- 現行フォルダ内に `S.init` という名前の S-PLUS タスクのテキストファイルを作成します。
- 後で説明するような S-PLUS 環境変数 `S_FIRST` を設定します。

関数 `.First` は、従来からある S-PLUS 初期設定ツールです。`.S.init` ファイルは、テキストファイルで S-PLUS の外部で容易に編集できるという利点があります。`S_FIRST` 変数は、特定の S-PLUS セッションの `.First` を無効にするのに便利な方法です。

S.chapters ファイルを作成する

S-PLUS セッションで特定の S-PLUS チャプターまたはライブラリを追加したい場合は、`S.chapters` ファイルを使ってそのフォルダを指定することができます。特定のユーザのユーティリティ関数と `map` ライブラリを追加する、次のようなサンプル `S.chapters` ファイルがあります。

```
D:¥Programs¥Insightful¥splus70¥users¥lenk
maps
```

ドライブ名と“¥¥”で始まるパス（ドライブ名と“¥¥”で始まるパスに評価する環境変数を使用するものを入れる）は、絶対パスとして解釈され、他の文字で始まるものは、`$SHOME¥library` に対するパスとして解釈されます。

S-PLUS を起動したいフォルダ内に `S.chapters` ファイルを作成することができます。S-PLUS は、現在のフォルダを調べてこの初期設定ファイルが存在するかどうかを確認し、そのファイルを見つけた場合はそのファイルの評価します。

関数.First を作成する

デフォルトのグラフィックデバイスを起動する、次のようなサンプル関数 `.First` があるとします。

```
> .First <- function( ) graphsheet( )
```

関数 `.First` を作成した後、ただちにその関数をテストしてそれがうまく動作するか必ず確認してください。うまく動作しないと、S-PLUS は次のセッションでその関数を実行しません。

S.init ファイルを作成する

セッションの出力幅とデフォルトの表示桁数を設定する、次のようなサンプル **S.init** ファイルがあるとします。

```
options(width=55, digits=4)
```

S-PLUS を起動したい任意のフォルダ内に **S.init** ファイルを作成することができます。S-PLUS は、現行フォルダを調べてこの初期設定ファイルが存在するかどうかを確認し、そのファイルを見つけた場合はそれを評価します。

S_FIRST を設定する

S_FIRST 変数に一連のコマンドを保存するときは、S-PLUS コマンドラインに次のようなシンタックスを使用します。

```
SPLUS.EXE S_FIRST=S-PLUS expression
```

たとえば、以下のコマンドは、S-PLUS にデフォルトのグラフィックデバイスを起動させます。

```
SPLUS.EXE S_FIRST=graphsheet( )
```

コマンドラインパーサが間違っず解釈しないように、複雑な S-PLUS 式は一重引用符または二重引用符で囲んだ方が安全です (S-PLUS 式に使用しない場合でも)。たとえば、以下のコマンドは S-PLUS を起動し、いくつかのオプションを修正します。

```
SPLUS.EXE S_FIRST='options(digits=4);options(expressions=128)'
```

オペレーティングシステムに固有の行の長さの制限のため、あるいは使い易くするために、次のようにファイルにコマンドを入れ、コマンドラインに @ 記号を前に付けたファイル名を入れることができます。

```
Echo options(digits=4);options(expressions=128) >  
c:¥myInitialization.txt  
SPLUS.EXE S_FIRST=@c:¥myInitialization.txt
```

また、たとえば次のように、いくつかのコマンドを 1 つの S-PLUS 関数にまとめ、S_FIRST をこの関数に設定することもできます。

```
> startup <- function( ) { options(digits=4)  
+ options(expressions=128) }
```

次のように S_FIRST を設定することによって S-PLUS を起動するたびにこの関数を呼び出すことができます。

```
SPLUS.EXE S_FIRST=startup()
```

S-PLUS が実行されている間は、初期設定だけで変数を設定することはできません。S_FIRST への変更は、S-PLUS を再起動したときだけ有効になります。

終了オプションの設定

S-PLUS は終了するとき、データフォルダ内の `.Last` と呼ばれる関数を探します。`.Last` が存在する場合は、S-PLUS はそれを実行します。たとえば次のように関数 `.Last` を使用して、一時的オブジェクトまたはファイルを削除してフォルダ内を整理することができます。

```
> .Last <- function ( ) dos(paste("del", getenv("S_Tmp"),  
+ "/*.Tmp, dep+"), Trans=T)
```

