SPSS

SPSS Base 16.0 User's Guide

SPSS® ソフトウェア製品の詳細については、SPSS 社の Web サイト (http://www.spss.co.jp) にアクセスするか、下記にご連絡ください。

エス・ピー・エス・エス株式会社 (SPSS Japan Inc.)
〒150-0012 東京都渋谷区広尾 1-1-39
恵比寿プライムスクェアタワー 10F
電話: (03) 5466-5511
Fax: (03) 5466-5621

SPSS は、SPSS Inc. の登録商標であり、SPSS の他のコンピュータ ソフトウェア製品名は、SPSS Inc. の商標です。商標やソフトウェアのライセンスまたは出版物の著作権の所有者の書面による同意がない限り、これらのソフトウェアについて記載した出版物の印刷や配布は認められません。

ソフトウェアやドキュメントは、限定的な権利の下で提供されます。米国政府による使用、複製、または開示には、52.227-7013 条の「The Rights in Technical Data and Computer Software (技術データおよびコンピュータ ソフトウェアに関する権利)」の (c) (1) (ii) 項で規定されている制限が適用されます。契約人および製造者は SPSS Inc., 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412 です。
特許番号: 第 7,023,453 号

注意事項: 本文中に記載するその他の製品名は、識別のみを目的として使用されており、各社の商標も含まれています。

Windows は、Microsoft Corporation の登録商標です。

Apple、Mac、および Mac のロゴは、Apple Computer, Inc. の米国およびその他の国で登録されている商標です。

この製品は、WinWrap Basic (Copyright 1993-2007, Polar Engineering and Consulting, http://www.winwrap.com) を使用します。

SPSS Base 16.0 User's Guide
Copyright © 2007 by SPSS Inc.
All rights reserved.

出版者の書面による事前の承諾なくしては、本書のいかなる部分も、再版、検索システムへの登録、または電子的、機械的、光学的、磁気的など、いかなる方法による転送をすることもできません。

ISBN-13: 978-1-56827-834-6
ISBN-10: 1-56827-834-9

1 2 3 4 5 6 7 8 9 0　　10 09 08 07

序文

SPSS 16.0

SPSS 16.0 は、データ分析のための統合システムです。SPSS では、ほとんどのファイル形式のデータをサポートし、取り込んだデータを使用して、グループ化されたレポートおよび図表を生成したり、分布とトレンド、記述統計、および複雑な統計の分析のプロットを作成できます。

この『SPSS Base 16.0 User's Guide』には、SPSS のグラフィカル ユーザー インターフェイスが記載されています。SPSS Base 16.0で実際に行う統計手続きの使用例は、ソフトウェアと共にインストールされるヘルプ システムで参照できます。統計手続きで使用するアルゴリズムは、PDF 形式で提供され、[ヘルプ] メニューから参照することができます。

また、SPSSでは、メニューおよびダイアログ ボックスの背後でコマンド言語を使用しています。システムの拡張機能の中には、コマンド シンタックスを介さないとアクセスできないものがあります。(これらの機能は Student 版では利用できません)。コマンド シンタックスの詳細なリファレンス情報は、2 つの形式で提供されています。1 つは総合的なヘルプ システムに組み込まれているもの、もう 1 つは PDF 形式の独立したドキュメント(『SPSS 16.0 Command Syntax Reference』) です。この PDF ドキュメントは、[ヘルプ] メニューから参照できます。

SPSSオプション

次のオプションは、(Student 版ではなく) 完全バージョンの SPSS Baseシステムへのアドオンの拡張機能として使用できます。

SPSS Regression Models ™は、従来の線型統計モデルに適合しないデータを分析する手法を提供します。プロビット分析、ロジスティック回帰、重み付け推定、2 段階最小 2 乗回帰、および従来の非線型回帰の手続きが含まれています。

SPSS Advanced Models™は、高度な実験科学や生物医学の研究でよく使用される手法に重点をおいています。一般線型モデル (GLM)、線型混合モデル、一般化線型モデル (GZLM)、一般化推定方程式 (GEE)、分散成分分析、対数線型分析、保険統計生命表、Kaplan-Meier の生存分析、Cox 回帰、標準および拡張 Cox 回帰の手続きが含まれます。

SPSS Tables™は、複合スタブ バナー テーブルや多重回答データの表示など、さまざまな高品質の表形式の報告書を作成します。

SPSS Trends ™は、複数の曲線当てはめモデル、平滑法モデル、自己回帰関数の推定方法を使用した包括的な予測分析と時系列分析を行います。

SPSS Categories ® は、コレスポンデンス分析を含む最適尺度法手続きを行います。

SPSS Conjoint ™により、個別の製品属性が消費者や一般市民の嗜好に対して与える影響の現実的な測定手法が提供されます。SPSS Conjoint があれば、購入決定時に消費者が行う、製品属性グループ中での、各製品属性のトレードオフ効果を簡単に測定することができます。

SPSS Exact Tests™は、サンプルが少ないか分布が著しく偏っているために通常の検定では不正確になるような場合に、正確な統計検定の *p* 値を計算します。

SPSS Missing Value Analysis™は、欠損データのパターンの表示、平均値と他の統計量の推定、および欠損観測値への値の代入を行います。

SPSS Maps™は、記号や色、棒グラフ、円グラフ、およびテーマの組み合せを使用して、地理的に分散したデータを高品質のマップに変換します。これにより、発生している事象の内容だけでなく、事象がどの場所で発生しているかが同時に確認できます。

SPSS Complex Samples™では、意識、市場、健康、世論などの調査を行う場合や、社会科学の分野でサンプル調査を使った研究を行う場合に、複雑なサンプルデザインをデータ分析に組み込むことができます。

SPSS Classification Trees ™では、ツリー ベースの分類モデルを作成します。ケースをグループに分類したり、独立 (予測) 変数の値を基に従属 (目的) 変数を予測するためのモデルです。この手続きには、分類を探索的、および確証的に分析するための検証ツールが用意されています。

SPSS Data Preparation ™はデータを素早くスナップショットにして視覚化します。無効なデータ値を識別する検証規則を適用する機能が搭載されています。この機能を使うと、作成した規則に基づいて、範囲外の値、欠損値、空白値などにフラグを付けることができます。また、個々の規則違反や、ケースあたりの規則違反の数を記録して変数に保存することができます。ユーザーがコピーしたり修正したりして利用できる、定義済みの規則の集合も用意されています。

Amos ™(**a**nalysis of **mo**ment **s**tructures) では、構造方程式モデリングを使用して、属性、認識、および振る舞いを生じさせるその他の因子に関わる概念モデルの確認と説明を行います。

SPSS プロダクト ファミリーには、データ入力、テキスト分析、分類、ニューラル ネットワーク、予測企業サービス用のアプリケーションも用意されています。

インストール

SPSS Base システムをインストールするには、SPSS Inc. から受け取った認証コードを使用して、ライセンス認証ウィザードを実行します。詳細は、SPSS Base システムに付属するインストール手順を参照してください。

互換性

SPSS は、さまざまなコンピュータ システムで動作するように設計されています。必要最小構成および推奨構成の詳細情報は、お買い上げのシステムに付属のインストール手順を参照してください。

シリアル番号

シリアル番号は、SPSS Inc. のお客様の ID 番号です。テクニカル サポート、支払い、システムのアップグレードなどに関して SPSS Inc. へ問い合わせる場合は、このシステム番号が必要になります。シリアル番号は、Base システムのパッケージに記載されています。

カスタマ サービス

製品の発送やお支払いに関してご質問がある場合は、SPSS 社までお問い合わせください (SPSS Japan のホームページは http://www.spss.co.jp/support/support.html です)。お問い合せの際には、シリアル番号をご用意ください。

トレーニング セミナー

SPSS Inc. では一般公開およびオンサイトでトレーニング セミナーを実施しています。セミナーでは実践的な講習を行います。セミナーは主要都市で定期的に開催されます。セミナーの詳細については、SPSS 社までお問い合わせください (SPSS Japan のホームページは http://www.spss.co.jp です)。

テクニカル サポート

SPSS のユーザーの方は、SPSS テクニカル サポートのサービスをご利用いただけます。SPSS 製品の使用方法や、対応するハードウェア環境へのインストールに関して問い合わせできます。テクニカル サポートに連絡するには、SPSS ホームページ (http://www.spss.co.jp) をご覧になるか、SPSS 社 (http://www.spss.com/worldwide) までお問い合わせください。お問い合わせの際は、ユーザー名、会社名、およびシリアル番号をご用意ください。

追加の出版物

Prentice Hall 社から、Marija Norušis 著『SPSS Statistical Procedures Companion』が出版されています。

ご意見をお寄せください

お客様のご意見は貴重な情報です。SPSS 製品についての感想をお寄せください。

このマニュアルについて

このマニュアルでは、SPSS Base システム に含まれる手続きのグラフィカル ユーザー インターフェイスについて解説しています。ダイアログ ボックスの画像は、SPSS から取られています。SPSS Base システムのコマンド シンタックスの機能に関する詳細は、ヘルプ システム全体に統合されている場合と、[ヘルプ] メニューから参照できる『SPSS 16.0 Command Syntax Reference』に独立した PDF 形式のドキュメントの場合の 2 つの形で提供されています。

目次

1 概要 1

バージョン 16.0J の新機能 2
ウィンドウ 4
　指定ウィンドウとアクティブ ウィンドウ 6
ステータス バー 7
ダイアログ ボックス 7
ダイアログ ボックス リストの変数名と変数ラベル 7
サイズ変更したダイアログ ボックス 8
ダイアログ ボックスの項目 9
変数の選択 9
データの種類、尺度、および変数リスト アイコン 9
ダイアログ ボックスで変数に関する情報を得る 10
データ分析の基本ステップ 10
統計コーチ 11
詳細 11

2 ヘルプの利用 12

出力項目についてのヘルプを利用するには 13

3 データ ファイル 15

データ ファイルを開く 15
　データ ファイルを開くには 16
　データ ファイルの種類 16
　[ファイルを開く] オプション 17
　Excel 95 以降のファイルを読み込む 17
　旧バージョンの Excel ファイルおよび他のスプレッドシートを読み込むには 17
　dBASE を読み込む 18
　Stata ファイルを読み込む 18
　データベース ファイルを読み込む 18

テキスト ウィザード 33
Dimensions データの読み込み 42
ファイル情報 46
データ ファイルの保存 47
変更したデータ ファイルを保存するには 47
データ ファイルを外部形式で保存するには 48
Excel 形式でのデータ ファイルの保存 51
SAS 形式でのデータ ファイルの保存 51
Stata 形式でのデータ ファイルの保存 53
変数のサブグループの保存 54
データベースへのエクスポート 55
Dimensions へのエクスポート 69
元のデータの保護 70
仮想アクティブ ファイル 70
データ キャッシュの作成 72

4 ディストリビュート アナリシス モード 75

サーバー ログイン 75
サーバー ログイン設定の追加および編集 76
サーバーの選択、切り替え、または追加を行うには 77
利用可能なサーバーの検索 78
リモート サーバーからデータ ファイルを開く 79
ローカル アナリシス モードとディストリビュート アナリシス モードでのファイル アクセス 80
ディストリビュート アナリシス モードでの手続きの使用 81
絶対パスと相対パス指定 81

5 データ エディタ 83

データ ビュー 83
変数ビュー 84
変数の属性を表示または定義するには 85
変数名 85
変数尺度 87
変数の型 87
変数ラベル 89
値ラベル 90

ラベルへの改行の挿入 . . . 91
欠損値 (変数の定義) . . . 91
列幅 . . . 92
変数の配置 . . . 92
変数定義属性を複数の変数に適用する . . . 92
ユーザー指定の変数属性 . . . 94
変数ビューをカスタマイズ . . . 98
スペル チェック変数と値ラベル . . . 98
データの入力 . . . 99
数値データを入力するには . . . 99
非数値データを入力するには . . . 100
データ入力に値ラベルを使用するには . . . 100
データ エディタでのデータ値の制限 . . . 100
データの編集 . . . 100
データ値の置換または修正 . . . 101
データ値の切り取り、コピー、および貼り付け . . . 101
新しいケースを挿入する . . . 102
新しい変数の挿入 . . . 102
データの型を変更するには . . . 103
ケースまたは変数について表示する . . . 103
データおよび属性値の検索および置換 . . . 104
データ エディタでのケースの選択状態 . . . 105
データ エディタ表示オプション . . . 105
データ エディタの印刷 . . . 106
データ エディタの内容を印刷するには . . . 106

6 多重データ ソースの使用方法 108

多重データ ソースの基本的な処理 . . . 109
コマンド シンタックスの複合データセットを使用する作業 . . . 110
データセット間での情報のコピーと貼り付け . . . 111
データセットの名前の変更 . . . 111
複合データセットを抑制 . . . 112

7 データの準備 113

変数プロパティ . . . 113

変数プロパティの定義 114
変数プロパティの定義 114
値ラベルと他の変数プロパティの定義 115
尺度の割り当て 117
ユーザー指定の変数属性 118
変数プロパティのコピー 119
多重回答グループ（多重回答） 120
多重回答グループを定義 121
データ プロパティのコピー 124
データ プロパティをコピーするには 124
ソース変数と目標変数の選択 125
コピーする変数プロパティの選択 127
データセット（ファイル）プロパティのコピー 129
結果 132
重複ケースの特定 132
連続変数のカテゴリ化 135
変数をビン分割するには 136
変数のビン分割 137
ビン分割されたカテゴリの自動生成 140
ビン分割されたカテゴリのコピー 142
連続変数のカテゴリ化でのユーザー欠損値 143

8 データの変換 144

変数の計算 144
変数の計算: IF 条件 146
変数の計算: 型とラベルの定義 147
関数 147
関数の欠損値 147
乱数ジェネレータ 148
ケース内での値の出現数の計算 149
出現数の計算: 集計する値の指定 150
出現数の計算: IF条件 151
値の再割り当て 152
同一変数への値の再割り当て 152
同一の変数への値の再割り当て: 今までの値と新しい値 153
他の変数への値の再割り当て 155
他の変数への値の再割り当て: 今までの値と新しい値 156

ケースのランク付け 157
ケースのランク付け: 手法の選択 158
ケースのランク付け: 同順位の処理 159
連続数への再割り当て 160
日付と時刻ウィザード 163
SPSS での日付と時刻 165
文字列からの日付時刻変数の作成 166
変数のグループからの日付時刻変数の作成 167
日付時刻変数の加算または減算 169
日付時刻変数の一部の抽出 176
時系列データの変換 178
日付の定義 179
時系列の作成 180
欠損値の置き換え 183
予測モデルによるデータのスコアリング 184
保存したモデルの読み込み 186
読み込まれたモデルのリストの表示 188
コマンド シンタックスで利用できる追加機能 188

9 ファイルの操作とファイルの変換 189

ケースの並べ替え 189
変数の並べ替え 190
行と列の入れ替え 192
データ ファイルの結合 192
ケースの追加 193
ケースの追加: 名前の変更 195
ケースの追加: 辞書情報 195
3 つ以上のデータ ソースの結合 196
変数の追加 196
変数の追加: 名前の変更 198
3 つ以上のデータ ソースの結合 199
データのグループ集計 199
データのグループ集計: 集計関数の定義 202
データのグループ集計: 変数名とラベル 202
ファイルの分割 203
ケースの選択 204
ケース選択: IF 条件の定義 206

ケースの選択: 無作為抽出 207
ケースの選択: 範囲の定義 208
ケースの重み付け 209
データの再構成 210
データを再構成するには 210
再構成データ ウィザード: 種類の選択 211
再構成データ ウィザード (変数からケースへ): 変数グループの数 214
再構成データ ウィザード (変数からケースへ): 選択 変数 215
再構成データ ウィザード (変数からケースへ): インデックスの作成 変数 217
再構成データ ウィザード (変数からケースへ): 1 つのインデックス変数の作成 . 220
再構成データ ウィザード (変数からケースへ): 複数のインデックス変数の作成 221
再構成データ ウィザード (変数からケースへ):オプション 222
再構成データ ウィザード (ケースから変数へ): 選択 変数 223
再構成データ ウィザード (ケースから変数へ): ソート データ 224
再構成データ ウィザード (ケースから変数へ): オプション 225
再構成データ ウィザード: 終了 227

10 出力の作業 229

ビューア (一般オプション) 229
結果の表示と非表示 230
出力の移動、削除、コピー 230
初期配置の変更 231
出力項目の配置の変更 231
ビューアのアウトライン 232
項目をビューアに追加する 233
[ビューア] の情報の検索および置換 234
出力を他のアプリケーションにコピー 236
出力項目をコピーして他のアプリケーションに貼り付けるには 236
出力のエクスポート 237
HTML、Word/RTF、および Excel オプション 239
PowerPoint オプション 240
PDF のオプション 240
テキスト オプション 242
図表エクスポート オプション 243
ビューアの印刷 245
出力と図表を印刷するには 245
印刷プレビュー 245
ページ属性: ヘッダーとフッター 246
ページ属性: オプション 248

出力の保存 249
ビューア文書を保存するには 250

11 ピボット テーブル 251

ピボット テーブルの操作 251
ピボット テーブルをアクティブにする 251
テーブルをピボットする 252
次元内の要素の表示順を変更 252
次元要素内の行と列を移動 252
行と列の入れ替え 253
行または列のグループ化 253
行または列のグループ解除 254
行ラベルおよび列ラベルの回転 254
層を使用した作業 254
層の作成と表示 254
層カテゴリに移動 256
項目の表示と非表示 257
テーブルの行と列を非表示にする 257
テーブルの行と列を表示する 258
次元ラベル表示と非表示 258
テーブル表題の表示と非表示 258
テーブルルック 258
テーブルルックを適用または保存するには 259
テーブルルックを編集または作成するには 260
テーブル プロパティ 260
ピボット テーブルのプロパティを変更するには 260
テーブル プロパティ: 全般 261
テーブル プロパティ: 脚注 261
テーブル プロパティ: セル書式 262
テーブル プロパティ: 罫線 264
テーブル プロパティ: 印刷 265
セル プロパティ 266
フォントおよび背景 267
形式値 267
位置合せおよび余白 268
脚注と解説 269
脚注と解説を追加 269
解説を表示か非表示にするには 270
テーブル内の脚注を隠す、または表示するには 270

脚注マーカー（特殊） 270
脚注番号のつけ直し 271
データセルの幅 271
列の幅の変更 271
ピボット テーブル内の隠れた罫線を表示 271
ピボット テーブルの行と列を選択 272
ピボット テーブルの印刷 273
テーブル幅または長さによる分割制御 273
ピボット テーブルからグラフを作成する 274

12 コマンド シンタックスを使用する作業 275

シンタックスの規則 275
ダイアログ ボックスからのシンタックスの貼り付け 277
ダイアログ ボックスからシンタックスを貼り付けるには 277
出力ログのシンタックスのコピー 278
出力ログのシンタックスをコピーするには 279
コマンド シンタックスを実行するには 280
Unicode シンタックス ファイル 280
複数の Execute コマンド 281

13 度数（プロビット分析） 282

度数分布表の統計 283
度数分布表の図表 285
度数分布表の書式 286

14 記述統計記述統計 287

記述統計のオプション 288
DESCRIPTIVES コマンドの追加機能 289

15 探索的分析 291

探索的分析の統計 293
探索的分析の作図 294
探索的分析のべき乗変換 295
探索的分析のオプション 295
EXAMINE コマンドの追加機能. 296

16 クロス集計表 297

クロス集計表の層. 299
クロス集計表のクラスタ棒グラフ. 299
クロス集計表の統計 299
クロス集計表のセル表示の設定. 302
クロス集計表の表書式 303

17 要約 304

ケースの要約のオプション 306
ケースの要約の統計. 306

18 平均値 (信頼性分析) 309

グループの平均のオプション 311

19 OLAP キューブ 314

OLAP キューブの統計 316
OLAP キューブの差分 318
OLAP キューブの表題 319

20 t 検定 320

独立したサンプルの t 検定 320
独立したサンプルの t 検定のグループの定義 322
独立したサンプルの t 検定のオプション 323
対応のある T 検定 323
対応のあるサンプルの t 検定のオプション 325
1 サンプルの t 検定 325
1 サンプルの t 検定のオプション 326
t 検定コマンドの追加機能 327

21 一元配置分散分析 328

一元配置分散分析の対比 329
一元配置分散分析のその後の検定 330
一元配置分散分析のオプション 333
ONEWAY コマンドの追加機能 334

22 GLM - 1 変量分散分析 335

GLM モデル 337
項の構築 (ロジット モデル) 337
平均平方 338
GLM の対比 339
対比の種類 340
GLM のプロファイル プロット 340
GLM のその後の比較 341
GLM の保存 343
GLM のオプション 345
UNIANOVA コマンドの追加機能 346

23 2 変量の相関分析 348

2 変量の相関分析のオプション 350
CORRELATIONS および NONPAR CORR コマンドの追加機能 350

24 偏相関 352

偏相関のオプション 353
PARTIAL CORR コマンドの追加機能 354

25 距離 355

距離行列の非類似度の測定方法 357
距離行列の類似度の測定方法 358
PROXIMITIESL コマンドの追加機能 358

26 線型回帰 360

線型回帰の変数選択方法 362
線型回帰の規則の設定 363
線型回帰の作図 363
線型回帰の新変数の保存 365
線型回帰の統計 367
線型回帰のオプション 369
REGRESSION コマンドの追加機能 370

27 順序回帰 371

順序回帰分析のオプション 372
順序回帰分析の出力 373
順序回帰分析の位置モデル 375
項の構築 (ロジット モデル) 376
順序回帰分析の尺度モデル 376
項の構築 (ロジット モデル) 376
PLUM コマンドの追加機能 377

28 曲線推定 378

曲線推定のモデル 380
曲線推定の保存 381

29 偏 2 段階最小 2 乗回帰 382

モデル 384
オプション 385

30 判別分析 386

判別分析: 範囲の定義 388
判別分析: ケースの選択 388
判別分析: 統計 389
判別分析: ステップワイズ法 390
判別分析: 分類 391
判別分析: 保存 393
DISCRIMINANT コマンドの追加機能 393

31 因子分析 394

因子分析のケースの選択 395
因子分析の記述統計 396
因子分析の因子抽出 397
因子分析の回転 398
因子分析の因子得点 399
因子分析オプション 400
FACTOR コマンドの追加機能 400

32 クラスタリングの手続きの選択 402

33 TwoStep クラスタ分析 403

TwoStep クラスタ分析のオプション 406
TwoStep クラスタ分析の作図 408
TwoStep クラスタ分析の出力 409

34 階層クラスタ分析 411

階層クラスタ分析の方法 413
階層クラスタ分析の統計 414
階層クラスタ分析の作図 415
階層クラスタ分析の新変数の保存 415
CLUSTER コマンド シンタックスの追加機能 416

35 大規模ファイルのクラスタ分析 417

大規模ファイルのクラスタ分析の効率 419
大規模ファイルのクラスタ分析の反復 419
大規模ファイルのクラスタ分析の保存 420
大規模ファイルのクラスタ分析のオプション 420
QUICK CLUSTER コマンドの追加機能 421

36 ノンパラメトリック検定 422

カイ 2 乗検定 422
カイ 2 乗検定の期待範囲および期待度数 424
カイ 2 乗検定のオプション 424
NPAR TESTS コマンドの追加機能 (カイ 2 乗検定) 425
2 項検定 425
2 項検定のオプション 426
NPAR TESTS コマンドの追加機能 (2 項検定) 427

ラン検定 427
ラン検定の分割点 428
ラン検定のオプション 429
NPAR TESTS コマンドの追加機能 (ラン検定) 429
1 サンプルによる Kolmogorov-Smirnov 検定 429
1 サンプルによる Kolmogorov-Smirnov 検定のオプション 431
NPAR TESTS コマンド追加機能 (1 サンプルによる Kolmogorov-Smirnov 検定) 431
2 個の独立サンプルの検定 431
2 個の独立サンプルの検定の種類 432
2 個の独立サンプルの検定: グループの定義 433
2 個の独立サンプルの検定: オプション 434
NPAR TESTS コマンドの追加機能 (2 個の独立サンプルの検定) 434
2 個の対応サンプルの検定 434
2 個の対応サンプルの検定の種類 435
2 個の対応サンプルの検定: オプション 436
NPAR TESTS コマンドの追加機能 (2 個の対応サンプル) 436
複数の独立サンプルの検定 436
複数の独立サンプルの検定の検定の種類 437
複数の独立サンプルの検定の範囲定義 438
複数の独立サンプルの検定のオプション 438
NPAR TESTS コマンドの追加機能 (K 個の独立サンプルの検定) 439
複数の対応サンプルの検定 439
複数の対応サンプルの検定の種類 440
複数の対応サンプルの検定の統計 441
NPAR TESTS コマンドの追加機能 (K 個の対応サンプルの検定) 441

37 多重回答の分析 442

多重回答グループを定義 443
多重回答の度数表 444
多重回答のクロス集計表 446
多重回答のクロス集計表の範囲の定義 448
多重回答のクロス集計表のオプション 448
MULT RESPONSE コマンドの追加機能 449

38 結果の報告 450

報告書の行の集計 450
集計報告書の行の集計を行うには 451
報告書のデータ列/ブレーク列の書式 452
報告書の集計行/最終集計行 452
報告書のブレーク オプション 453
報告書のオプション 454
報告書のレイアウト 454
報告書の表題 455
報告書の列の集計 456
集計報告書の列の集計を行うには 457
データ列の集計関数 458
合計列のデータ列集計 459
報告書の列の書式 460
報告書の列の集計でのブレーク列のオプション 460
報告書の列の集計のオプション 461
報告書の列の集計のレイアウト 461
REPORT コマンドの追加機能 461

39 信頼性分析 463

信頼性分析の統計 465
RELIABILITY コマンドの追加機能 466

40 多次元尺度法 468

多次元尺度法のデータの形式 470
多次元尺度法の尺度の作成 470
多次元尺度法のモデル 471
多次元尺度法のオプション 472
ALSCAL コマンドの追加機能 473

41 比率統計量 474

比率統計量 476

42 ROC 曲線 478

ROC 曲線のオプション . . . 480

43 図表作成機能の概要 481

図表の作成および編集 . . . 481
図表の作成 . . . 481
図表の編集 . . . 485
図表定義のオプション . . . 488
表題および脚注の追加と編集 . . . 488
[全般] オプションの設定 . . . 489

44 ユーティリティ 492

変数情報 (変数) . . . 492
データ ファイルのコメント . . . 493
変数セット . . . 493
変数グループの定義 . . . 494
変数グループの使用 . . . 495
目標変数リストの並べ替え . . . 496

45 オプション 497

[全般] オプション . . . 498
[ビューア] オプション . . . 500
[データ] オプション . . . 501
デフォルト変数の変更 . . . 503
[通貨] オプション . . . 503
通貨書式を作成するには . . . 504
[出力ラベル] オプション . . . 504
[グラフ] オプション . . . 506
データ要素 色 . . . 507
データ要素 線 . . . 507

データ要素 マーカー 508
データ要素 塗りつぶし 509
[ピボット テーブル] オプション 509
ファイル位置オプション 511
[スクリプト] オプション 512

46 メニューとツールバーのカスタマイズ 515

メニュー エディタ 515
ツールバーのカスタマイズ 516
ツールバーの表示 516
ツールバーをカスタマイズするには 517
ツールバーのプロパティ 518
[編集] ツールバー 518
新しいツールの作成 519

47 バッチ ジョブ 521

HTML オプション 525
PowerPoint オプション 525
PDF オプション 525
テキスト オプション 525
ランタイム値 526
ユーザー プロンプト 527
コマンド ラインからのバッチ ジョブの実行 528
バッチ ジョブ ファイル変換 529

48 出力管理システム 531

出力オブジェクトの種類 534
コマンド識別子とテーブルのサブタイプ 535
ラベル 536
OMS オプション 538
ログ 542
ビューアからの出力表示の除外 543

SPSS データ ファイルへの出力の送信 543
例: 単一の 2 次元テーブル 543
例: 層のあるテーブル 544
複数のテーブルから作成されたデータ ファイル 545
列要素を制御してデータ ファイルの変数を制御する 548
OMS で生成されたデータ ファイルの変数名 550
OXML テーブルの構造 551
OMS 識別子 555
ビューアのアウトラインからの OMS 識別子のコピー 556

付録

索引 558

章

1

概要

SPSSでは、記述的なメニューとシンプルなダイアログ ボックスを使用したグラフィカルな環境で、ユーザーに代わってほとんどの仕事をこなしてくれる、強力な統計分析およびデータ管理システムを備えています。ほとんどのタスクは、マウスをポイントしてクリックするだけで実行できます。

統計分析用の簡単なポイント アンド クリック インターフェイスに加え、SPSSは、次のような機能を持っています。

データ エディタ。データ エディタとは、データを定義、入力、編集、および表示するためのスプレッドシートに似た多機能システムです。

ビューア。ビューアを使用すると、結果の参照、出力の表示または非表示の選択、結果の表示順序の変更、および SPSS と他のアプリケーション間での高品質なテーブルおよび図表の移動などを容易に行うことができます。

多次元ピボット テーブル。分析結果は、多次元のピボット テーブルで生き生きと表現できます。テーブルの行、列、層の配置を変え、さまざまな角度で結果を探索していってください。従来のレポート形式では見逃してしまいがちな、重要な発見があるかも知れません。テーブルを分割して一度に 1 つのグループだけ表示されるようにすると、グループの比較も容易になります。

高解像度グラフィックス。高解像度のフルカラーで作成できる円グラフ、棒グラフ、ヒストグラム、散布図、3-D グラフなどが標準機能として組み込まれています。

データベース アクセス。データベースの情報の検索は、複雑な SQL のクエリーの代わりにデータベース ウィザードで行うことができます。

データの変換。変換機能によって、分析用のデータの準備が容易になります。データのサブグループ分け、カテゴリの結合、ファイルの追加、集計、結合、分割、入れ替えなど、さまざまな作業が簡単に行えます。

オンライン ヘルプ。詳細なチュートリアルによって、全体の概要をつかむことができます。また、ダイアログ ボックスの状況依存ヘルプ トピックにより、具体的な作業の手順がわかります。ピボット テーブルのポップアップ定義には統計用語の説明があります。統計コーチは、必要な手続きを見つけるのに役立ち、ケース スタディは統計手続きの使用方法や出力結果の解釈を理解するために実践的な例を提供しています。

コマンド言語。ほとんどのタスクは、マウスをポイントしてクリックするだけで実行できますが、SPSS には、多くの一般的なタスクを保存して自動化できるようにするための強力なコマンド言語が用意されています。コマンド言語を使用すると、メニューやダイアログ ボックスにはない機能も実行できます。

コマンド シンタックスの詳細は、総合的なヘルプ システムに統合されています。また、『Command Syntax Reference』にある PDF 形式の独立したマニュアルもあります。これは、ヘルプ メニューから参照できます。

バージョン 16.0J の新機能

ユーザー インターフェイスの強化。ポイント アンド クリック インターフェイスへの拡張は次の項目を含みます。

- すべてのダイアログ ボックスはサイズの変更が可能になりました。ダイアログ ボックスを広くすると変数一覧も広くなり、多くの変数名および/または記述的なラベルを見ることができます。ダイアログ ボックスを長くすると変数一覧も長くなり、多くの変数名をスクロールしないで見ることができます。
- ドラッグ アンド ドロップ変数選択はすべてのダイアログ ボックスでサポートされています。
- 順序および特性を表示する変数一覧はダイアログ ボックスで即座に変更可能になりました。希望する場合はいつでも、ソートの順序 (アルファベット順、ファイル順序、尺度) を変更、および/または変数名また変数ラベルの表示を切り替えます。 詳細は、7 ページのダイアログ ボックス リストの変数名と変数ラベルを参照してください。

データおよび出力管理。データおよび出力管理の拡張は次の項目を含みます。

- Excel 2007 ファイルを読み込みおよび書き込みます。
- 一度に 1 個のデータセットまたは複数のデータセットを作業するかを選択します。 詳細は、498 ページの 45 章の[全般] オプションを参照してください。
- 複合多次元ピボット テーブルの隠れたレイヤおよび項目を含む ビューア文書の情報を検索および置換します。 詳細は、234 ページの 10 章の[ビューア] の情報の検索および置換を参照してください。
- 定義された文字型変数の幅にかかわらず (以前は文字は 8 バイト以下の定義幅に制限されていました)、欠損値および変数ラベルを割り当てます。
- 新規文字ベースの文字型関数。
- 出力管理システム (OMS) は、ビューア ファイル形式 (.spv) および VML 形式図表および HTML 文書のポップアップ図表を持つイメージ マップをサポートします。 詳細は、531 ページの 48 章の出力管理システムを参照してください。
- データ エディタの [変数ビュー] をカスタマイズします。属性列の表示順を変更し、属性列の表示をコントロールします。 詳細は、98 ページの 5 章の変数ビューをカスタマイズを参照してください。

- 有効なデータセットの変数を、アルファベット順に、または属性（辞書）値により並び替えます。詳細は、190 ページの 9 章の変数の並べ替えを参照してください。
- [変数ビュー] のスペル チェック変数ラベルおよび値ラベル。詳細は、98 ページの 5 章のスペル チェック変数と値ラベルを参照してください。
- 基礎変数型（文字、数値）を変更し、文字変数の定義済み幅を変更し、文字変数の幅を各変数の最長の観測値に自動的に設定します。
- Unicode データおよびシンタックス ファイルの読み込みと書き込みを行ないます。詳細は、498 ページの 45 章の[全般] オプションを参照してください。
- ファイルを検索し、保存するためにデフォルトのディレクトリ位置をコントロールします。詳細は、511 ページの 45 章のファイル位置オプションを参照してください。

パフォーマンス。複合プロセッサまたは複合コア付きのプロセッサのあるコンピュータでは、パフォーマンスの速度がより速くなる複合スレッドがいくつかの手順で利用できます。

統計機能の強化。統計機能の強化には次のものがあります。

- 偏 2 段階最小 2 乗（PLS）。PLS は、通常最小 2 乗法（OLS）回帰、正準相関、構造方程式モデリングに対する代わりの予測技術であり、予測変数が相関しているとき、あるいは予測数がケース数を超えているときは特に役立ちます。詳細は、382 ページの 29 章の偏 2 段階最小 2 乗回帰を参照してください。
- 多層パーセプトロン（MLP）。MLP 手続きは、多層パーセプトロンと呼ばれる特定の種類のニューラル ネットワークを当てはめます。多層パーセプトロンは、前方提供アーキテクチャを使用し、複数の隠れ層を持つことができます。多層パーセプトロンは、非常に柔軟にモデルの種類を当てはめることができます。最も一般的に使われるニューラル ネットワーク アーキテクチャの 1 つです。この手続きは新しいニューラル ネットワーク オプションで利用できます。
- 放射基底関数（RBF）。放射基底関数（RBF）ネットワークは、放射基底関数層と呼ばれる 1 個だけの非表示の層を持つ前方提供、監視カテゴリ ネットワークです。多層パーセプトロン（MLP）のように、RBF ネットワークは予測および分類の両方ができます。MLP より非常に早いですが、当てはめることができるモデルの種類は自由度が高くありません。この手続きは新しいニューラル ネットワーク オプションで利用できます。
- 一般化線型モデルは、順序多項およびTweedie 分布、負の 2 項補助パラメータの制限された最尤法推定、および尤度比統計量を含む多くの新しい機能をサポートします。この手続きは Advanced Models オプションで利用できます。
- Cox 回帰によりモデル情報を XML（PMML）ファイルにエクスポートできるようになります。この手続きは Advanced Models オプションで利用できます。
- コンプレックス サンプルの Cox 回帰。Cox ハザードの比例回帰を生存時間、すなわちンプレックス サンプリング法によって抽出されたサンプルのイベントの出現の前の時間の長さの分析に適用します。この手順は時間共存の連続

およびカテゴリ予測値をサポートしています。この手順は予測変数セットの効果を分析すると同様にサブグループの差異を考慮する簡単な方法を提供します。この手順は、サンプル手順の置換を行う (WR) か、置換を行わない (WOR) かの両方で方法を調整する比例した確率および等しい確率を含むサンプルを選択するために使用されたサンプル デザインを考慮して分散を推定します。この手続きは Complex Samples オプションで利用できます。

プログラマビリティの拡張。プログラマビリティの拡張の強化には次のものがあります。

- R-Plugin。SPSS のべき乗を R 統計ルーチンと組み合わせます。このプラグインは、www.spss.com/devcentral (*http://www.spss.com/devcentral*) からのみダウンロードが可能です。
- 入れ子の **プログラムを開始-プログラムを終了** コマンド構造。
- 複合データセットを作成および管理する機能。

コマンド シンタックス。コマンド シンタックスの追加と変更の完全なリストに関しては、『Command Syntax Reference』の「**はじめに**」の章の「**リリースの履歴**」を参照してください ([ヘルプ] メニューから入手可能)。

この機能はすでにサポートしていません。

- 「インタラクティブ」グラフ用の別の図表エディタはすでにありません。レガシー「インタラクティブな」図表ダイアログ ボックスおよび IGRAPH コマンド シンタックスから作成された図表はすべてのほかの図表と同じ形式で作成され、同じ図表エディタで編集されます。
- レガシー「インタラクティブな」図表ダイアログ ボックスおよび IGRAPH コマンド シンタックスで提供された同じ機能は利用できなくなりました。
- ドラフト ビューアは利用できなくなりました。
- SPSS 16.0 の SPSS(.spo ファイル) の以前のバージョンで作成されたビューア ファイルを開けることはできません。Windows のオペレーティング システムでは、以前のバージョンで作成されたビューア ファイルを表示したり編集するには、レガシー ビューアを含むインストール CD をインストールしてください。
- マップ オプションは利用できなくなりました。
- Trends および Tables オプションのレガシー手順のダイアログ インターフェイスは利用できなくなりました。Trends では、AREG、ARIMA、および EXSMOOTH コマンドがあります。Tables では、TABLES コマンドがあります。レガシー手順を含むこれらのオプションのライセンスがあれば、これらのコマンドのコマンド シンタックスはまだサポートされています。

ウィンドウ

SPSS には異なるタイプのウィンドウが数多くあります。

データ エディタ。データ エディタには、データ ファイルの内容が表示されます。データ エディタを使用すると、新しいデータ ファイルを作成することもできますが、既存のデータ ファイルを修正することもできます。複数のデータ ファイルを開いている場合、各データ ファイルに対し、独立したデータ エディタ ウィンドウが表示されます。

ビューア。すべての統計結果、テーブル、および図表はビューアで表示されます。出力は編集することができ、保存することもできます。[ビューア] ウィンドウは、出力を生成する手続きを最初に実行するときに自動的に開きます。

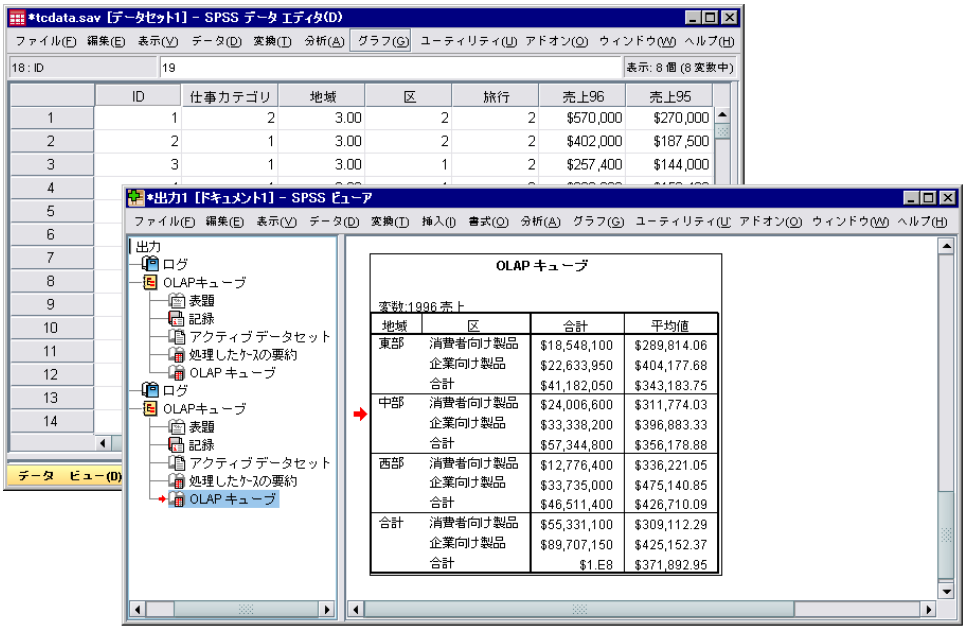
ピボット テーブル エディタ。ピボット テーブルに表示される出力は、ピボット テーブル エディタを使用してさまざまな方法で修正できます。テキストの編集、行と列のデータの交換、色の追加、多次元テーブルの作成、結果の表示と非表示の選択などを行うことができます。

図表エディタ。[図表エディタ] ウィンドウで高解像度図表およびプロットを修正できます。色の変更、各種のフォントと大きさの選択、水平軸と垂直軸の切り替え、3-D 散布図の回転、および図表タイプの変更などを行うことができます。

テキスト出力エディタ。ピボット テーブルに表示されないテキスト出力は、テキスト出力エディタで修正できます。出力の修正およびフォントの指定 (種類、スタイル、色、大きさ) の変更ができます。

シンタックス エディタ。ダイアログ ボックスの選択項目をシンタックス ウィンドウに貼り付けることができます。コマンド シンタックスの形式で選択が表示されます。コマンド シンタックスを編集すると、ダイアログ ボックスでは利用できない特別な機能を利用することができます。コマンドはファイルに保存して、セッションで繰り返し使用できます。

図 1-1
データ エディタとビューア



指定ウィンドウとアクティブ ウィンドウ

複数の [ビューア] ウィンドウを開いている場合、出力は **指定された** [ビューア] ウィンドウへ送られます。複数の [シンタックス エディタ] ウィンドウを開いている場合、コマンド シンタックスは指定された [シンタックス エディタ] ウィンドウに貼り付けられます。指定したウィンドウは、タイトル バーにあるアイコンで、プラス記号で示されます。ウィンドウの指定は、いつでも変更することができます。

指定ウィンドウと現在選択されているウィンドウである **アクティブ** ウィンドウとを混同しないようにしなければなりません。ウィンドウが重なっている場合は、アクティブ ウィンドウが前面に表示されます。ウィンドウを開くと、ウィンドウは自動的にアクティブ ウィンドウになるとともに指定ウィンドウになります。

指定ウィンドウの変更

- ▶ 指定するウィンドウをアクティブ ウィンドウにします (ウィンドウ内でクリックします)。
- ▶ ツールバー (プラス記号アイコン) で [ウィンドウの指定] ボタンをクリックします。

または

- ▶ メニューから次の項目を選択します。

ユーティリティ
ウィンドウの指定

注: データ エディタ ウィンドウでは、アクティブなデータ エディタ ウィンドウによって、この後の計算や分析で使用するデータセットが決定されます。「指定」データ エディタ ウィンドウはありません。 詳細は、109 ページの 6 章の多重データ ソースの基本的な処理を参照してください。

ステータス バー

各 SPSS ウィンドウの下端にあるステータス バーには、次の情報が表示されます。

コマンドの状態。ケース カウンタは、実行する各手続きまたはコマンドの、これまでに処理されたケース数を示します。反復処理を必要とする統計手続きでは、反復回数が表示されます。

フィルタの状態。分析用に無作為抽出またはケースのサブグループを選択した場合、**「フィルタ オン」**というメッセージは、現在有効なのは選択された特定のタイプのケースのみで、データ ファイル内のすべてのケースが分析に取り入れられているわけではないことを示します。

重み付けの状態。「重み付けオン」というメッセージは、分析用ケースに重みを付けるために、重み付け変数が使用されていることを示します。

ファイルの分割の状態。「**分割ファイル オン**」というメッセージは、1 つ以上のグループ化変数に基づいて、データ ファイルが分析用にグループに分割されていることを示します。

ダイアログ ボックス

ほとんどのメニュー選択項目で、ダイアログ ボックスが開かれます。ダイアログ ボックスを使用すると、分析用の変数とオプションを選択することができます。

通常、統計手続きおよび図表のためのダイアログ ボックスには、2 つの基本的なコンポーネントがあります。

ソース変数のリスト。作業データセットの変数を含む変数のリスト。選択された手続きで許可された変数の型だけがソース リストに表示されます。短い文字型 short_string_variable および長い文字型 GLongStringVar 変数の使用は、多くの手続きで制限されています。

目標変数リスト。従属変数リストや独立変数リストのような、分析用に選択した変数を示す 1 つ以上のリスト。

ダイアログ ボックス リストの変数名と変数ラベル

ダイアログ ボックスで変数名または変数ラベルを表示でき、ソース変数リスト上の変数の並び替え潤を制御できます。

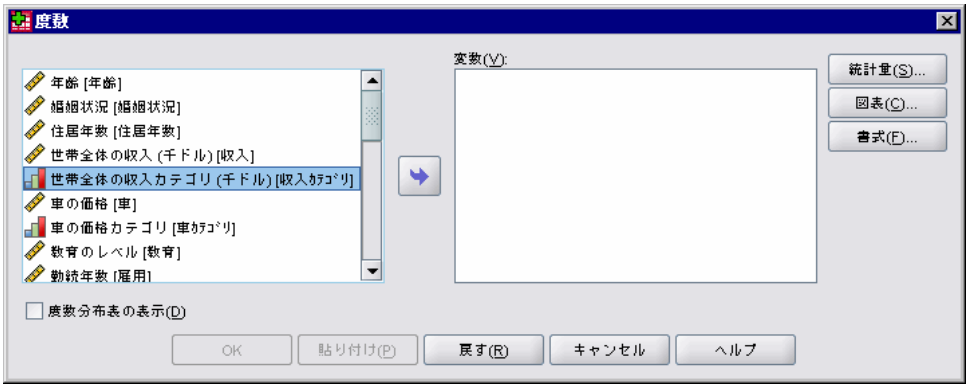
- ソースリストの変数のデフォルト表示属性を制御するには、[編集] メニューの [**オプション**] を選択します。詳細は、498 ページの 45 章の[全般] オプションを参照してください。
- ダイアログ ボックスのソース変数リスト表示属性を変更するには、ソースリストの任意の変数を右クリックして、コンテキスト メニューから表示属性を選択します。変数名または変数ラベルを表示でき (名前は定義済みラベルなしの任意の変数として表示されます)、ファイル順序、アルファベット順、または尺度によりソース リストを並び替えできます。尺度の詳細は、9 ページのデータの種類、尺度、および変数リスト アイコンを参照してください。

図 1-2
ダイアログ ボックスに表示される変数ラベル

サイズ変更したダイアログ ボックス

ダイアログ ボックスは、ウィンドウと同様に、外枠や角をクリックしてドラッグするとサイズ変更できます。たとえば、ダイアログ ボックスの幅を大きくすると、変数リストの幅も大きくなります。

図 1-3
サイズ変更したダイアログ ボックス



ダイアログ ボックスの項目

ほとんどのダイアログ ボックスには、5 つの標準的な項目があります。

OK。手続きを実行します。変数を選択して追加の指定を選択した後で、[OK] をクリックして手続きを実行し、ダイアログ ボックスを閉じます。

貼り付け。ダイアログ ボックスの選択からコマンド シンタックスを生成して、そのシンタックスをシンタックス ウィンドウに貼り付けます。ダイアログ ボックスからは利用できない追加の SPSS 機能を使用してコマンドをカスタマイズできます。

戻す。選択した変数リスト内の変数を選択解除して、ダイアログ ボックスとサブダイアログ ボックス内のすべての指定をデフォルトの状態に戻します。

キャンセル。ダイアログ ボックスを開いてからダイアログ ボックスの設定で行った変更をキャンセルしてダイアログ ボックスを閉じます。ダイアログ ボックスの設定は 1 回のセッション中は有効です。ダイアログ ボックスでは、設定を変更しない限り最後に指定された設定が保持されます。

ヘルプ。状況依存のヘルプです。このコントロールにより、現在のダイアログ ボックスについての情報が含まれたヘルプ ウィンドウが表示されます。

変数の選択

選択する変数が 1 つである場合、ソース変数リスト上で変数を選択して、目標変数リストにドラッグ アンド ドロップで移動します。矢印キーをクリックし、ソース リストから変数を目標リストに移動します。目標変数リストが 1 つだけの場合は、変数を 1 つずつダブルクリックするとソース リストから目標リストへ移動できます。

複数の変数を選択することもできます。

- 変数リストの中で連続している複数の変数を一度に選択するには、最初の変数をクリックして、Shift キーを押しながらグループ内の最後の変数をクリックします。
- 変数リストの中で連続していない複数の変数を一度に選択するには、最初の変数をクリックしてから Ctrl キーを押しながら次の変数をクリックし、これを続けます。(Macintoshでは Command キー)。

データの種類、尺度、および変数リスト アイコン

ダイアログ ボックス リストにある変数の隣に表示されるアイコンにより、変数型と尺度に関する情報が提供されます。

尺度	データ タイプ			
	数値	文字列	日付	時間

スケール データ		利用不可		
順序データ				
名義データ				

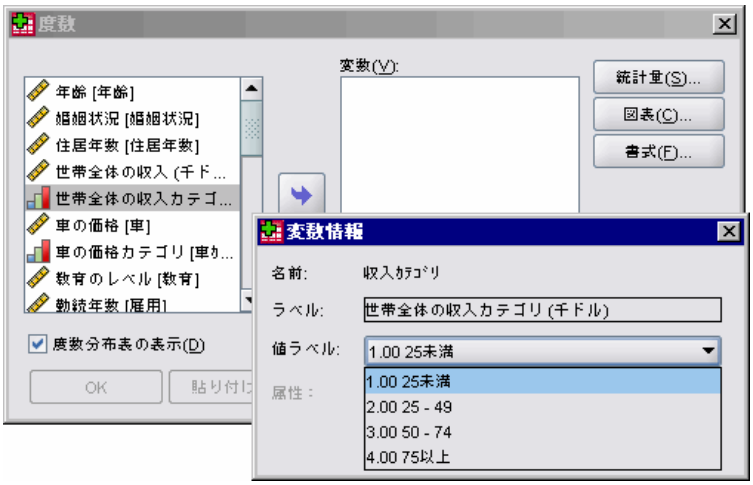
- 尺度の詳細は、87 ページの変数尺度を参照してください。
- 数値、文字列、日付、時刻データの詳細については、87 ページの変数の型を参照してください。

ダイアログ ボックスで変数に関する情報を得る

▶ ソース変数リストまたは目標変数リストの変数を右クリックします。

▶ [**変数情報**] を選択します。

図 1-4
変数情報



データ分析の基本ステップ

SPSS を使用してデータを分析するのは簡単です。必要な操作は、次の処理だけです。

データを SPSS に取り込む。すでに保存済みの SPSS データ ファイルを開くか、スプレッドシート、データベース、またはテキスト データ ファイルを読み込むか、またはデータを直接データ エディタに入力できます。

手続きを選択する。メニューから手続きを選択して、統計量を計算、または図表を作成します。

分析のための変数を選択する。データ ファイル内の変数は、選択された手続きのダイアログ ボックスに表示されます。

手続きを実行して結果を確認する。出力結果はビューアに表示されます。

統計コーチ

SPSS、または利用できる統計手続きに慣れていない場合、統計コーチを使用すると、一般的な用語やイラスト付きの例を使用した簡単な質問に答えていくことで、データに最も適した基本的な統計および図表機能が選択できます。

統計コーチを使用するには、SPSS ウィンドウから次の項目を選択します。

ヘルプ
統計コーチ

統計コーチでは、Base system に含まれるもののうち、手続きの一部を選択してとりあげています。基本的で共通に使用される多くの統計技法のための一般的な助言を提供するようにデザインされています。

詳細

基本に関する総合的な概要の詳細は、オンライン チュートリアルを参照してください。SPSS のメニューから次の項目を選択します。

ヘルプ
チュートリアル

章
2

ヘルプの利用

ヘルプは、次のような多くの異なった形式で提供されます。

[ヘルプ] メニュー。ほとんどの SPSS ウィンドウの [ヘルプ] メニューでは、主要なヘルプ システム、チュートリアル、および技術参考資料へアクセスできます。

- **トピック。**特定のヘルプ トピックの検索に使用できる [目次] タブ、[インデックス] タブ、および [検索] タブへアクセスできます。
- **チュートリアル。**SPSS 多くの基本機能の使用方法に関する、図解付きの段階的な説明が記載されています。チュートリアル全体を最初から最後まで参照する必要はありません。任意の順序でさまざまなトピックに目を通したり、キーワードや目次を使用して特定のトピックを検索するなどして、参照したいトピックを選択できます。
- **ケース スタディ。**さまざまな統計分析の作成方法と処理結果の解釈方法に関する実践的な例が提供されます。例で使用されているサンプル データ ファイルも提供されているので、例を実際に処理してどのように処理結果が出るのか詳細に確認できます。学習する特定の手続きを目次から選択したり、キーワードで関連トピックを検索できます。
- **統計コーチ。**使用する手続きの検索手順を知ることができるウィザード形式のアプローチです。一連の項目を選択すると、統計コーチでは、選択した基準を満たす統計手続き、報告手続き、または図表作成手続きのダイアログ ボックスが表示されます。Base system のほとんどの統計手続きと報告手続き、および多数の図表作成手続きへアクセスできます。
- **Command Syntax Reference。**コマンド シンタックスの詳細なリファレンス情報は、2 つの形式で提供されています。1 つは総合的なヘルプ システムに組み込まれているもの、もう 1 つは PDF 形式の独立したドキュメント (『SPSS Command Syntax Reference』) です。この PDF ドキュメントは、[ヘルプ] メニューから参照できます。
- **統計アルゴリズム。**ほとんどの統計的な手続きで使用されるアルゴリズムについての詳細情報は、2 つの形式で提供されています。1 つはヘルプ システム全体に統合されているもの、もう 1 つはマニュアル CD-ROM に収められている PDF 形式文書、『SPSS Algorithms』です。ヘルプ システムに含まれている、個別のアルゴリズムに関するトピックへのリンク一覧を表示するには、[ヘルプ] メニューの **[アルゴリズム]** を選択 します。

状況依存のヘルプ。ユーザー インターフェイス内の多くの場所で、状況依存のヘルプを参照できます。

- **ダイアログ ボックスのヘルプ ボタン。**ほとんどのダイアログ ボックスには、そのダイアログ ボックスのヘルプ トピックに直接移動できるヘルプ ボタンがあります。ヘルプ トピックは、一般的な情報を表示するために使用され、関連するトピックにリンクしています。
- **ピボット テーブルのコンテキスト メニュー ヘルプ。**ビューアのアクティブなピボット テーブルの項目を右クリックし、コンテキスト メニューから [**用語のヘルプ**] を選択すると、項目の定義が表示されます。
- **コマンド シンタックス。**コマンド シンタックス ウィンドウで、コマンドのシンタックス ブロック内の任意の箇所にカーソルを合わせて F1 キーを押すと、。そのコマンドの完全なコマンド シンタックス表が表示されます。コマンド シンタックスの詳細なマニュアルは、関連項目リストのリンクおよびヘルプの [目次] タブから参照できます。

その他のリソース

テクニカル サポート Web サイト。http://support.spss.com にあるサイトで、よく発生する問題の解決策が見つかります。(テクニカル サポート Web サイトには、ログイン ID とパスワードが必要です。ID とパスワードの取得方法に関する情報は、上記の URL で提供されています)。

SPSS Developer Central。Developer Central は SPSS ユーザーおよび SPSS アプリケーション開発者のすべてのレベルのリソースです。効用、グラフィック例、新規統計モジュール、SPSS テクノロジーの記事をダウンロードします。SPSS ユーザー コミュニティおよび SPSS との対話するにはフォーラムを利用します。SPSS Developer Central (**http://www.spss.com/devcentral**) を参照してください。

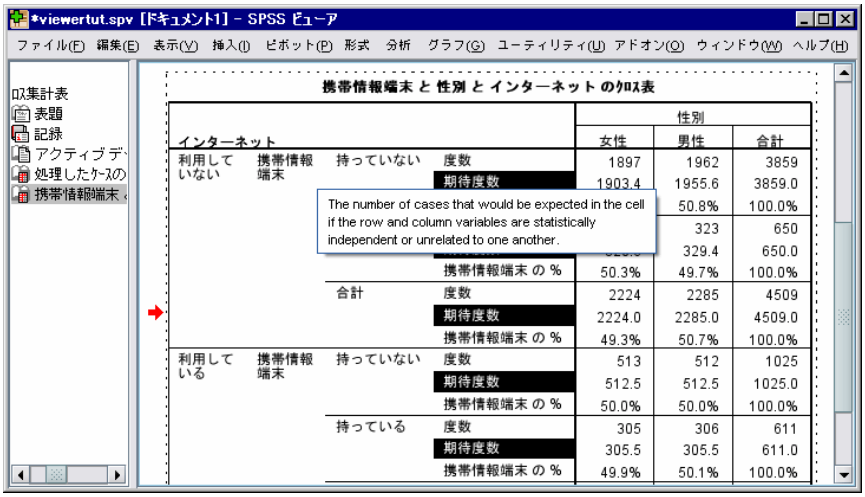
出力項目についてのヘルプを利用するには

[ビューア] のピボット テーブル出力の用語の定義を参照するには次のようにします。

1. ピボット テーブルをダブルクリックして、アクティブにします。
2. 説明を参照したい項目を右クリックします。
3. コンテキスト メニューから [**用語のヘルプ**] を選択します。

項目の定義がポップアップ ウィンドウに表示されます。

図 2-1
マウスの右ボタンを使ったアクティブなピボット テーブル用語のヘルプ



インターネット				性別 女性	男性	合計
利用していない	携帯情報端末	持っていない	度数	1897	1962	3859
			期待度数	1903.4	1955.6	3859.0
					50.8%	100.0%
					323	650
					329.4	650.0
			携帯情報端末 の %	50.3%	49.7%	100.0%
		合計	度数	2224	2285	4509
			期待度数	2224.0	2285.0	4509.0
			携帯情報端末 の %	49.3%	50.7%	100.0%
利用している	携帯情報端末	持っていない	度数	513	512	1025
			期待度数	512.5	512.5	1025.0
			携帯情報端末 の %	50.0%	50.0%	100.0%
		持っている	度数	305	306	611
			期待度数	305.5	305.5	611.0
			携帯情報端末 の %	49.9%	50.1%	100.0%

章

3

データ ファイル

データ ファイルにはさまざまな形式があり、SPSS では、次の形式を始めとするさまざまな形式を処理できます。

- Excel や Lotus で作成したスプレッドシート
- Oracle、SQLServer、Access、dBASEなど、さまざまなデータベースで作成されたデータベース テーブル
- タブ区切り、およびその他の形式の単純テキスト ファイル
- 他のシステムで作成した SPSS データ ファイル
- SYSTAT データ ファイル
- SAS データ ファイル
- Stata データ ファイル

データ ファイルを開く

SPSS 形式で保存したファイル以外に、Excel、SAS、Stata、タブ区切り、その他のファイルについても、ファイルを中間形式に変換しないで、またはデータ定義情報を入力しないで開くことができます。

- データ ファイルは、開いた時点でアクティブ データセットになります。1 つまたは複数のデータ ファイルが既に開いている場合、ファイルは開いたまま残るため、現在のセッションで利用していけます。[データ エディタ] ウィンドウ内で開いている任意のデータ ファイルをクリックすると、そのファイルがアクティブなデータセットになります。 詳細は、108 ページの 6 章の多重データ ソースの使用方法を参照してください。
- リモート サーバーを使ってコマンドを処理し、手続きを実行するディストリビュート アナリシス モードでは、使用できるデータ ファイル、フォルダ、およびドライブは、リモート サーバーで、またはリモート サーバーから何が利用できるかによって異なります。ダイアログ ボックスの一番上に現在のサーバー名が表示されます。ドライブを共有ドライブとして指定し、データ ファイルが格納されているフォルダを共有フォルダとして指定していない限り、ローカル コンピュータ上のデータ ファイルにはアクセスできません。 詳細は、75 ページの 4 章のディストリビュート アナリシス モードを参照してください。

データ ファイルを開くには

▶ メニューから次の項目を選択します。

ファイル
開く
データ...

▶ [データを開く] ダイアログ ボックスで、開くファイルを選択します。

▶ **[開く]** をクリックします。

オプションとして、次の選択が可能です。

- **[観測値に基づいて文字幅を最小化]** を使用した変数の最長観測値に各文字変数に文字幅を自動的に設定します。これは Unicode モードのコード ページ データ ファイルを読み込むときに特に便利です。 詳細は、498 ページの 45 章の[全般] オプションを参照してください。
- スプレッドシート ファイルの先頭行から変数名を読み込む。
- スプレッドシート ファイルに読み取るセルの範囲を指定する。
- 読み込む Excel (Excel 95 以降) ファイル内のワークシートを指定する。

データベースからデータを読み込む方法の詳細は、データベース ファイルを読み込む18 ページのを参照してください。テキスト データ ファイルからデータを読み込む方法の詳細は、テキスト ウィザード33 ページのを参照してください。

データ ファイルの種類

SPSS。 SPSS 形式で保存されたデータ ファイル、また、DOS 製品の SPSS/PC+ 形式で保存されたデータ ファイルを開きます。

SPSS/PC+。 SPSS/PC+ データ ファイルを開きます。

Systat。 SYSTAT データ ファイルを開きます。

SPSS Portable。 ポータブル形式で保存されたデータ ファイルを開きます。ファイルをポータブル形式で保存すると、SPSS 形式でファイルを保存するよりも時間がかかります。

Excel。 Excel ファイルを開きます。

Lotus 1-2-3。 Lotus のリリース 3.0、2.0、または 1A の 1-2-3 書式で保存されたデータ ファイルを開きます。

SYLK。 いくつかのスプレッドシート アプリケーションで使用される SYLK (シンボリック リンク) 形式で保存されたデータ ファイルを開きます。

dBASE。 dBASE IV、dBASE III または III PLUS のいずれか、または dBASE II で使用される dBASE 形式のファイルを開きます。各ケースはレコードです。この形式でファイルを保存すると、変数、値のラベルおよび欠損値の指定が失われます。

SAS。SAS バージョン 6 ～ 9 および SAS トランスポート ファイル。

Stata。Stata バージョン 4 ～ 8。

[ファイルを開く] オプション

変数名の読み込み。スプレッドシートの場合は、ファイルの最初の行、または定義された範囲の最初の行から変数名を読み込むことができます。有効な変数名を作成するために値は必要に応じて変換されます (たとえば、空白は下線に変換されます)。

ワークシート。Excel 95 以降のファイルには、複数のワークシートを登録できます。データ エディタは、デフォルトで最初のワークシートを読み込みます。別のワークシートを読み込むには、ドロップダウン リストからワークシートを選択します。

範囲。スプレッドシート データ ファイルでは、セルの範囲を読み込むこともできます。スプレッドシート アプリケーションでセル範囲を指定する場合と同じ方法を使用します。

Excel 95 以降のファイルを読み込む

Excel 95 以降のファイルの読み込みには、次の規則が適用されます。

データの型と幅。列のひとつひとつがそれぞれ変数となります。各変数のデータの型と幅は、Excel ファイル内でのデータ値の型と幅によって決まります。列に複数のデータの型 (たとえば、日付と数値) が含まれている場合、そのデータの型は文字型に設定され、値はすべて有効な文字型値として読み込まれます。

空白セル。数値型変数では、空白セルはピリオドで示されたシステム欠損値に変換されます。文字型変数では、空白は有効な文字型値であり、空白セルは有効な文字型値として処理されます。

変数名。Excel ファイルの最初の行 (または指定した範囲の最初の行) を変数名として読み込む場合、変数の命名規則に従っていない値は有効な変数名に変換され、元の名前は変数ラベルとして使用されます。Excel ファイルから変数名を読み込まない場合は、デフォルトの変数名が割り当てられます。

旧バージョンの Excel ファイルおよび他のスプレッドシートを読み込むには

バージョン 95 以前の Excel ファイルおよびその他のスプレッドシートのデータを読み込む場合には、次の規則が適用されます。

データの型と幅。各変数のデータの型と幅は、列内にある最初のデータ セルの列幅とデータ型によって決まります。他の型の値はシステム欠損値に変換されます。列内の最初のデータ セルが空白の場合、そのスプレッドシートのグローバル デフォルトのデータの型 (通常は数値) が使用されます。

空白セル。数値型変数では、空白セルはピリオドで示されたシステム欠損値に変換されます。文字型変数では、空白は有効な文字型値であり、空白セルは有効な文字型値として処理されます。

変数名。スプレッドシートから変数名を読み込まない場合は、Excel ファイルおよび Lotus ファイルの変数名として列文字 (A、B、C、...) が使用されます。R1C1 表示形式で保存された SYLK ファイルおよび Excel ファイルでは、文字 C の後に列番号を付けたものが変数名として使用されます (C1、C2、C3、...)。

dBASE を読み込む

データベース ファイルは、SPSS 形式データ ファイルと論理的に非常によく似ています。dBASE ファイルには、次の一般的な規則が適用されます。

- フィールド名は有効な変数名に変換されます。
- dBASE フィールド名でコロンを使用すると、下線に変換されます。
- 削除のマークが付いていても実際には消去されていないレコードは、取り込まれます。SPSS では、削除マークの付いたケースをアスタリスクで表すための新しい文字型変数 D_R が作成されます。

Stata ファイルを読み込む

Stata データ ファイルには、次の一般的な規則が適用されます。

- **変数名。**Stata 変数名は、大文字と小文字が区別される SPSS 変数名に変換されます。大文字と小文字の違いを除けばまったく同じになる Stata 変数名は、下線および連続する添え字が追加されて、有効な SPSS 変数に変換されます (_A、_B、_C、...、_Z、_AA、_AB、...、など)。
- **変数ラベル。**Stata 変数ラベルは、SPSS 変数ラベルに変換されます。
- **値ラベル。**Stata 値ラベルは、「拡張された」欠損値に割り当てられたものを除き、SPSS 値ラベルに変換されます。
- **欠損値。**Stata の「拡張された」欠損値は、システム欠損値に変換されます。
- **日付変換。**Stata の日付書式値は、SPSSDATE 書式 (d-m-y) の値に変換されます。Stata の「時系列の」日付書式値 (週、月、四半期など) は、単純な数値型 (F) 書式に変換されます。ただし、1960 年からの経過期間を週、月、四半期などを単位として表す内部的な元の整数値は保存されます。

データベース ファイルを読み込む

データベースドライバがあるすべてのデータベース形式のデータを読み込むことができます。ローカル アナリシス モードでは、必要なドライバをローカル コンピュータにインストールする必要があります。ディストリビュート アナリシス モード (SPSS サーバーで使用可能) では、リモート サーバーにドライバをインストール

する必要があります。詳細は、75 ページの 4 章のディストリビュート アナリシス モードを参照してください。

データベース ファイルを読み込むには

- メニューから次の項目を選択します。

  ファイル
  データベースを開く
  新規クエリー...

- データソースを選択します。
- 必要であれば、(データ ソースに応じて) データベース ファイルを選択するか、またはログイン名、パスワード、およびその他の情報を入力するか、あるいはその両方を行います。
- テーブルとフィールドを選択します。OLE DB データ ソース (Windows のオペレーティング システム上でのみ可能) の場合、1 つのテーブルのみ選択することができます。
- テーブル間の関係を指定します。
- 次のオプションが選択できます。
  - データの選択基準を指定する。
  - ユーザーの入力用プロンプトを追加して、パラメータ クエリーを作成する。
  - 作成したクエリーを実行前に保存する。

保存済みのデータベース クエリーを編集するには

- メニューから次の項目を選択します。

  ファイル
  データベースを開く
  クエリーの編集...

- 編集するクエリー ファイル (*.spq) を選択します。
- 指示に従ってクエリーを新規作成します。

保存済みのクエリーを使用してデータベース ファイルを読み込むには

- メニューから次の項目を選択します。

  ファイル
  データベースを開く
  クエリーの実行...

- 実行するクエリー ファイル (*.spq) を選択します。

- ▶ 必要であれば、(データベース ファイルに応じて) ログイン名とパスワードを入力します。
- ▶ クエリーにプロンプトが埋め込まれている場合には、必要であれば、情報を入力してください (たとえば、売上高を取り込む場合は、その四半期名)。

データ ソースの選択

データベース ウィザードの最初の画面を使用して、読み込むデータ ソースの種類を選択します。

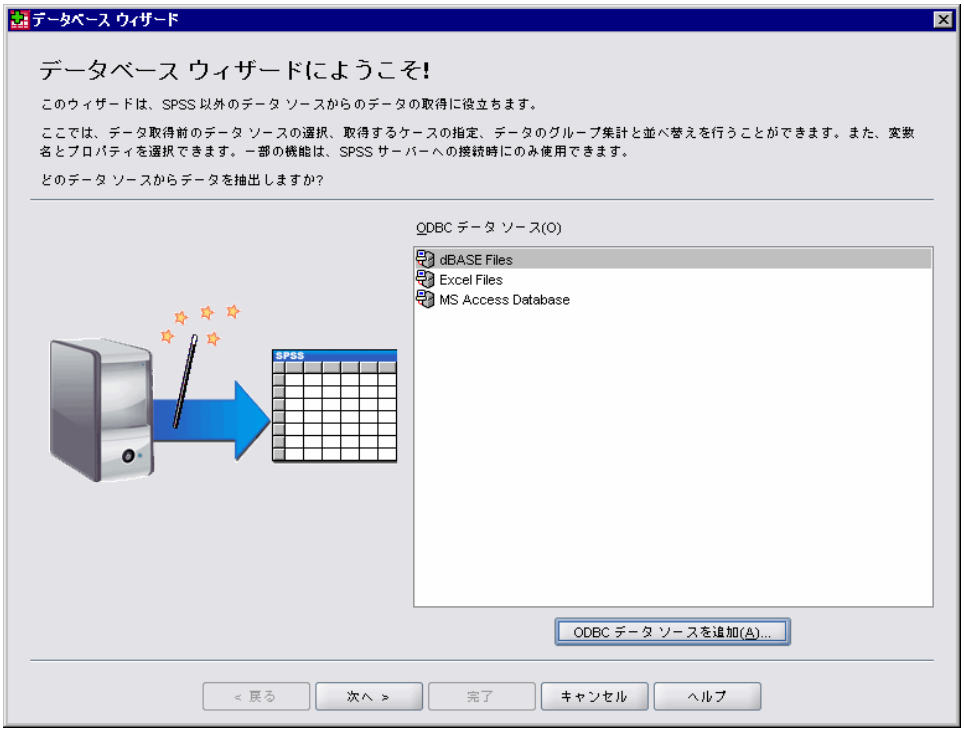
ODBC データ ソース

ODBC データ ソースが設定されていない場合、または新しいデータ ソースを追加する場合は、[ODBC データ ソースの追加] をクリックします。

- Linux オペレーティング システムでは、このボタンは利用できません。ODBC データ ソースは odbc.ini で指定され、ODBCINI 環境変数はファイルの位置に設定されなければいけません。詳細は、データベース ドライバのドキュメントを参照してください。
- ディストリビュート アナリシス モード (SPSS サーバーで使用可能) では、このボタンはありません。ディストリビュート アナリシス モードでデータ ソースを追加する方法については、システム管理者に問い合わせてください。

ODBC データ ソースは、2 つの重要な情報から成り立っています。1 つはデータのアクセスに使用されるドライバについての情報、もう 1 つはアクセスするデータベースの場所についての情報です。データ ソースを指定するには、まず、適切なドライバをインストールしておく必要があります。ローカル アナリシス モードでは、SPSS のインストール CD からいろいろなデータベース形式のドライバをインストールできます。

図 3-1
データベース ウィザード



OLE DB データ ソース

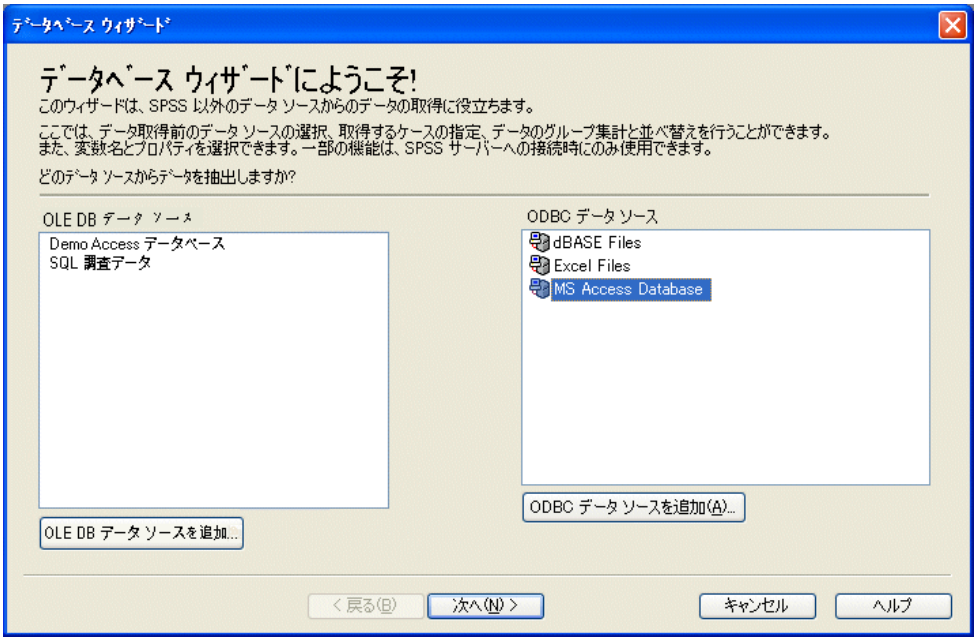
OLE DB データ ソース (Windows のオペレーティング システム上でのみ可能) にアクセスするには、次の項目がインストールされてなければいけません。

- .NET framework
- Dimensions Data Model および OLE DB Access

本リリースの SPSS と互換性があるこれらのコンポーネントのバージョンは、インストール CD からインストールすることができます。また、Autoplay メニューからインストールすることもできます。

- 複数テーブルの結合は、OLE DB データ ソースではサポートされていません。1 度に読み込めるデータ ファイルは 1 つだけです。
- OLE DB データ ソースは、ローカル アナリシス モードでだけ追加できます。ディストリビュート アナリシス モードで、Windows サーバーに OLE DB データ ソースを追加する方法については、システム管理者に問い合わせてください。
- ディストリビュート アナリシス モード (SPSS サーバーで使用可能) では、OLE DB データ ソースは Windows サーバー上でだけ利用できます。また、.NET および Dimensions Data Model と OLE DB Access を、サーバー上へインストールすることが必要です。

図 3-2
OLE DB データ ソースへのアクセス機能のあるデータベース ウィザード



OLE DB データ ソースを追加するには

- [OLE DB データ ソースを追加] をクリックします。
- [データ リンク プロパティ] ダイアログ ボックスで、[プロバイダ] タブをクリックし、OLE DB プロバイダを選択します。
- [次へ] または [接続] タブをクリックします。
- ディレクトリ位置とデータベース名を入力するか、ボタンをクリックしてデータベースを参照して、データベースを選択します。(ユーザー名とパスワードが必要になることがあります)。
- 必要な情報をすべて入力し終えたら、[OK] をクリックします。(指定したデータベースが利用可能であることを確認するには、[接続テスト] ボタンをクリックします)。
- データベース接続情報の名前を入力します。(この名前は、利用可能な OLE DB データ ソースの一覧に表示されます)。

図 3-3
[OLE DB 接続情報に名前を付けて保存] ダイアログ ボックス

▶ [OK] をクリックします。

これにより、データベース ウィザードの最初の画面に戻ります。ここで、OLE DB データ ソースの一覧から保存した名前を選択して、ウィザードの次のステップに進むことができます。

OLE DB データ ソースの削除

OLE DB データ ソースの一覧からデータ ソース名を削除するには、次の場所にあるデータ ソースの名前が付いた UDL ファイルを削除します。

[ドライブ名]:¥Documents and Settings¥[ログイン ユーザー名]¥Local Settings¥Application Data¥SPSS¥UDL

データ フィールドの選択

[データの選択] ステップでは、どのテーブルとフィールドを読み込むかが制御されます。データベース フィールド (列) は、変数として読み込まれます。

テーブル内のフィールドが 1 つ以上選択されている場合は、そのテーブル内にある他のフィールドもすべて [データベース ウィザード] ウィンドウに表示されます。しかし、変数としてインポートされるのは、選択されているものだけです。これによりテーブル結合を作成し、ここでインポートしないフィールドを使用するときの基準を指定できます。

図 3-4
データベース ウィザード、データの選択

フィールド名の表示。テーブル内のフィールドを表示するには、テーブル名の左にあるプラス記号 (+) をクリックします。フィールドを表示しないようにするには、テーブル名の左にあるマイナス記号 (–) をクリックします。

フィールドを追加するには。[使用可能なテーブル] リストでフィールドをダブルクリックするか、または [読み込むフィールドの順序] リストまでドラッグします。フィールドリスト内でドラッグ アンド ドロップを行うとフィールドが並べ替えられます。

フィールドを削除するには。[読み込むフィールドの順序] リストでフィールドをダブルクリックするか、または [使用可能なテーブル] リストまでドラッグします。

フィールド名の並べ替え。このチェック ボックスをオンにすると、データベース ウィザードで使用できるフィールドが文字順 (五十音順) に表示されます。

デフォルトでは、[使用可能なテーブル] リストには標準のデータベース テーブルのみが表示されます。リストに表示する項目の種類は次の方法で指定できます。

- **テーブル。**標準のデータベース テーブルです。
- **ビュー。**ビューは、クエリーで定義した仮想「テーブル」またはダイナミック「テーブル」です。ビューには、複数のテーブルを結合して含めたり、他のフィールド値を基に計算した結果から得られた複数のフィールドを結合して含めたり、これらのテーブルとフィールドの両方を結合して含めたりすることができます。

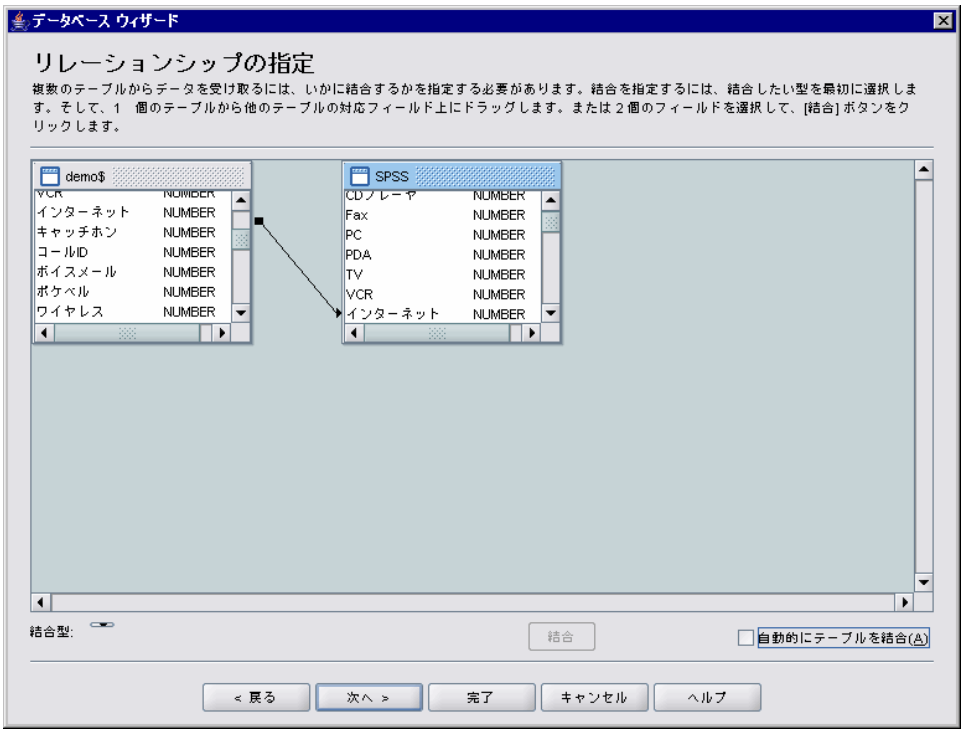
- **シノニム。**シノニムは、テーブルまたはビューのエイリアスです。通常、クエリーで定義します。
- **SYSTEM TABLE。**システム テーブルでは、データベースのプロパティを定義します。一部の標準データベース テーブルは、システム テーブルに分類される場合があります。これらのデータベース テーブルは、このオプションを選択した場合にのみ表示されます。多くの場合、実際のシステム テーブルへのアクセスはデータベース管理者に制限されます。

注: OLE DB データ ソース (Windows オペレーティング システム上のみで可能) の場合、1 つのテーブルのみ選択できます。複数テーブルの結合は、OLE DB データ ソースではサポートされていません。

テーブル間のリレーションシップの作成

ODBC データ ソースでは、[リレーションシップの指定] ダイアログ ボックスを使用すると、テーブル間のリレーションシップを定義できます。複数のテーブルからなるフィールドを選択する場合は、1 つ以上の結合を定義する必要があります。

図 3-5
データベース ウィザード、リレーションシップの指定



リレーションシップの設定。リレーションシップを作成するには、任意のテーブルから結合したいフィールドへ、1 つのフィールドをドラッグします。データベース ウィザードにより、2 つのフィールド間に **結合線**が引かれ、そのリレーションシップが示されます。各フィールドは、データの型が同じでなければなりません。

自動的にテーブルを結合。主キー/外部キーまたは同じ名前とデータ型を持つテーブルを自動的に結合します。

結合種類。外部結合がドライバでサポートされている場合、内部結合、左外部結合、または右外部結合のいずれかを指定できます。

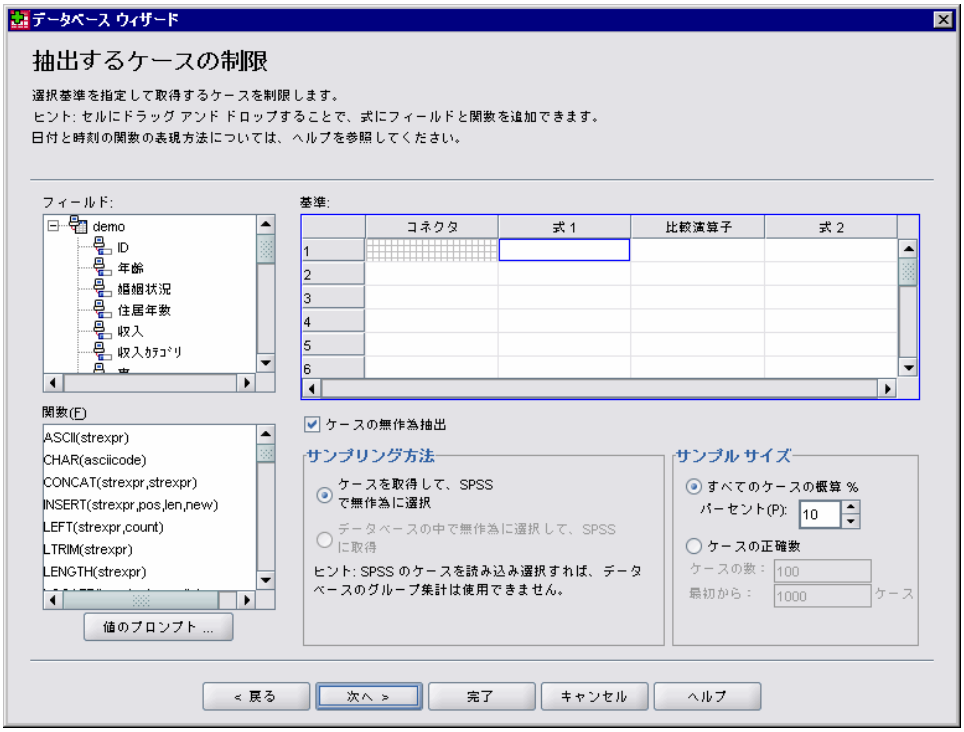
- **内部結合。**内部結合は、関連フィールドが等しい行だけを含みます。この例では、2 つのテーブルのうち、対応する ID 値を持つすべての行が含まれます。
- **外部結合。**内部結合の一対一対応に加え、外部結合を使用することで、一対多対応方法でテーブルを結合することができます。たとえば、データ値を表すレコードがわずかしかないテーブルと、調査回答者を表す数百や数千のレコードが含まれるテーブルにある、値を持つ関連する説明ラベルとを対応付けることができます。左外部結合は、左にあるテーブルからはすべてのレコードを取り入れ、右にあるテーブルからは関連フィールドが等しいレコードだけを取り入れます。右外部結合では、右のテーブルからすべてのレコードをインポートして、左にあるテーブルからは関連フィールドが等しいレコードだけをインポートします。

抽出するケースの制限

[抽出するケースの制限] ダイアログ ボックスを使用すると、ケース (行) のサブセットを選択するための基準を指定できます。ケースを制限するということは、通常、基準グリッドに基準を記入することです。基準は、2 つの式とその間にある関係で構成されます。式により、各ケースに **真の値**、**偽の値**、または **欠損値**のいずれかの結果が返されます。

- 結果が **真**の場合は、ケースが選択されます。
- 結果が **偽**または **欠損**の場合は、ケースが選択されません。
- ほとんどの基準は、6 つの関係演算子 (<、>、<=、>=、=、および <>) のうち、1 つ以上を使用します。
- 式にはフィールド名、定数、算術演算子、数値型の関数およびその他の関数、論理変数、および関係演算子を使用できます。変数としてインポートする予定のないフィールドも使用できます。

図 3-6
データベース ウィザード、抽出するケースの制限



基準を作成するには、最低 2 つの式とそれらを結び付ける関係が必要です。

▶ 式を構築するには、次の方法のいずれかを選択してください。
- [数式] セルに、フィールド名、定数、算術演算子、数値型の関数およびその他の関数、または、論理変数を入力できます。
- [フィールド] リストでフィールドをダブルクリックします。
- フィールドを [フィールド] リストから、[数式] セルまでドラッグします。
- 任意のアクティブな [数式] セルのドロップダウン メニューからフィールドを選択します。

▶ 関係演算子 (= や > など) を選択するときは、カーソルを [条件式] セルにおいて演算子を入力するか、ドロップダウン メニューから演算子を選択します。

ケースを選択する式を持つ WHERE 節が SQL に含まれている場合、式の日付と時刻は、特殊な方法で指定する必要があります (次の例に示されている中かっこを含む)。

- 日付リテラルは、一般的な {d 'yyyy-mm-dd'} 形式を使用して指定する必要があります。

- 時刻リテラルは、一般的な {t 'hh:mm:ss'} 形式を使用して指定する必要があります。
- 日時リテラル (タイムスタンプ) は、一般的な {ts 'yyyy-mm-dd hh:mm:ss'} 形式を使用して指定する必要があります。
- 日付または時刻、あるいはその両方の値は、全体を一重引用符で囲む必要があります。年は 4 桁の形式で指定し、日付および時刻は値の部分ごとに 2 桁の形式で指定する必要があります。たとえば、2005 年 1 月 1 日午前 1 時 5 分の場合は次のように指定します。

  {ts '2005-01-01 01:05:00'}

関数。算術関数、論理演算関数、文字列関数、日付関数、時間関数が組み込みで用意されています。関数をリストから式にドラッグしたり、有効な SQL 関数を入力することができます。有効な SQL 関数については、データベースの説明書を参照してください。標準関数のリストは、次の場所にあります。

http://msdn2.microsoft.com/en-us/library/ms711813.aspx

無作為抽出を使用。これは、データ ソースからケースを無作為抽出するオプションです。データ ソースが大きい場合、小さなサンプルを代表とすれば、手続きの実行時間を大幅に短縮することができます。データ ソースで固有の無作為抽出が利用できる場合、その無作為抽出は、SPSSでの無作為抽出よりも速く実行されます。SPSSでの無作為抽出は、サンプルを無作為に抽出するために、データ ソース全体を読み込む必要があるためです。

- **近似 (任意のサンプルの選択)。**近似的に指定されたケースの割合で任意のサンプルを選択します。SPSS は各ケースに対して独立した疑似無作為決定を行うので、選択されたケースのパーセントを指定したパーセントに近づけることができます。データ ファイル内にケースが多くあるほど、選択されたケースのパーセントは指定したパーセントに近くなります。
- **正確に (任意のサンプルの選択)。**ケースに指定された合計数から指定されたケース数の任意のサンプルを選択します。指定されたケースの総数がデータ ファイル内にあるケースの総数よりも多いと、その数に比例してサンプルのケースは要求された数よりも少なくなってしまいます。

注: 無作為抽出を使用している場合は、(SPSS サーバーを使用して分散モードで使用できる) 集計を利用できません。

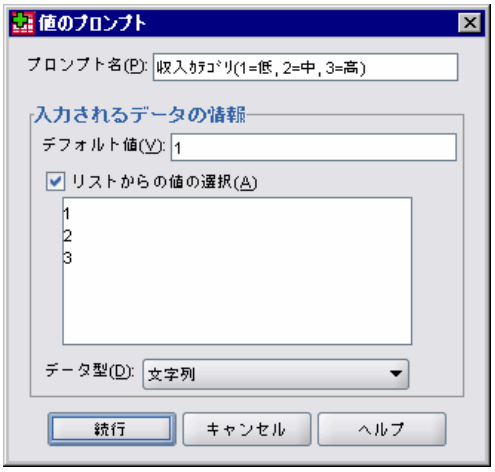
値のプロンプト。クエリーにプロンプトを埋め込んで、**パラメータ クエリー**を作成できます。ユーザーがクエリーを実行すると、ここで指定した情報を入力するように求められます。同じデータを別の視点で見る必要が場合には、この方法が有効です。たとえば、同じクエリーを実行して、別の会計四半期の売上高を見るような場合です。

▶ この場合、カーソルを [数式] セルにおき、[**値のプロンプト**] をクリックしてプロンプトを作成します。

パラメータ クエリーの作成

[値のプロンプト] ダイアログ ボックスを使用して、他の人があなたのクエリーを実行するたびに情報の入力を求めるダイアログを作成します。この機能は、異なる基準を使用して同じデータ ソースを参照する場合に役に立ちます。

図 3-7
値のプロンプト



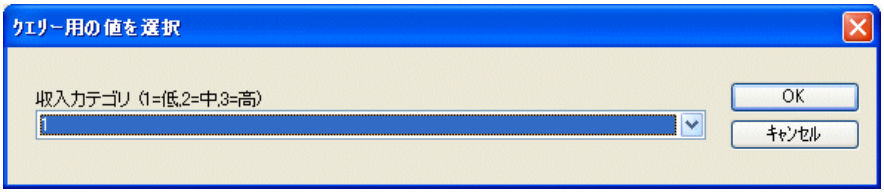
プロンプトを作成するには、プロンプト文字列およびデフォルト値を入力します。プロンプト文字列は、ユーザーがあなたのクエリーを実行するたびに表示されます。文字列は、入力する情報の種類を指定する必要があります。ユーザーがリストから選択しない場合、入力の形式がわかるよう、文字列によりヒントを提示する必要があります。たとえば、[**四半期 (Q1、Q2、Q3、...) を入力してください。**] とします。

リストからの値の選択。このチェック ボックスを選択しないと、ここに入力した値から、ユーザーを制限することができます。値がリターンで区切られるようにしてください。

データ型。データの型として、[**数値**]、[**文字列**]、[**日付**] のいずれかを選択します。

最終結果は、次のようになります。

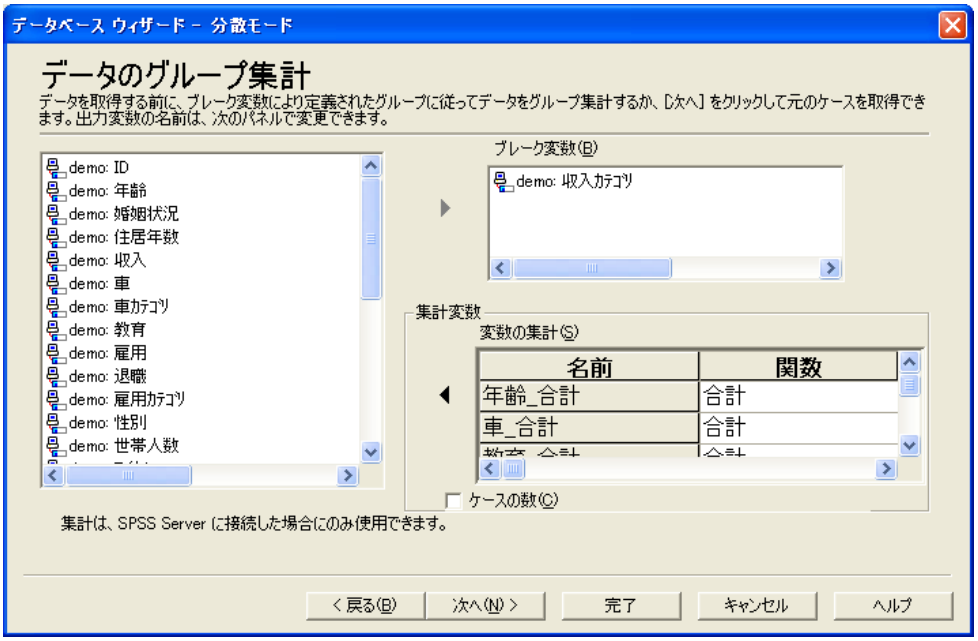
図 3-8
ユーザー定義プロンプト



データの集計

分散モードでリモート サーバー (SPSS サーバーで使用可能) に接続している場合は、データを SPSSに読み込む前に集計できます。

図 3-9
データベース ウィザード、データの集計



データは SPSSに読み込んだ後でも集計できますが、データ ソースが大きい場合は前もって集計することにより時間を節約できることがあります。

- ▶ 集計データを作成するため、ケースのグループ化方法を定義する 1 つ以上のブレーク変数を選択します。
- ▶ 1 つ以上の集計変数を選択します。
- ▶ [関数] をクリックし、各集計変数に集計関数を選択します。
- ▶ 必要に応じて、各ブレーク グループ内のケース数を持つ変数を作成してください。

注:SPSS無作為抽出を使用する場合は、集計を使用できません。

変数を定義する

変数名とラベル。完全なデータベースのフィールド (列) 名は変数ラベルとして使用できます。変数名を変更しない限り、[データベース ウィザード] は、次の 2 種類の方法のどちらかでデータベースから変数名をそれぞれの列に割り当てます。

- データベース フィールドの名前が有効な一意の変数名であれば、変数名として使用されます。
- データベース フィールドの名前が有効な一意の変数名でない場合は、新しい一意の名前が自動的に作成されます。

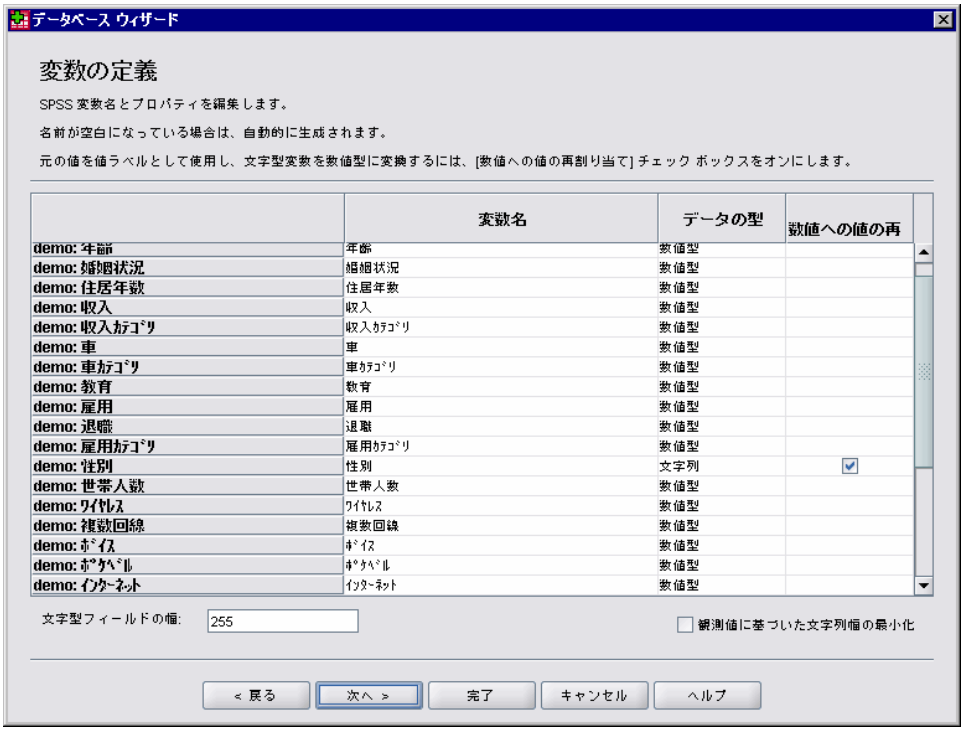
変数名を編集するには、該当するセルをクリックします。

文字列の数値への変換。自動的に数値型変数に変換する文字型変数の **[数値への値の再割り当て]** チェック ボックスを選択します。文字列は、元の値のアルファベット順で、連続した整数値に変換されます。元の値は、新しい変数の値ラベルとして保持されます。

変数の幅 - 文字列フィールドの幅。このオプションにより、変数の幅と文字列の幅の値を制御します。デフォルトでは、幅は 255 バイトで最初の 255 バイト (通常、1 バイト文字の言語で 255 文字) のみを読み込みます。この幅は 32,767 バイトまで広げることができます。文字型値が切り捨てられないようにしたいところですが、処理効率を落とすことになるので、不必要に大きな値を指定しないようにします。

観測値に基づいて文字幅を最小化します。最長観測値に各文字変数に文字幅を自動的に設定します。

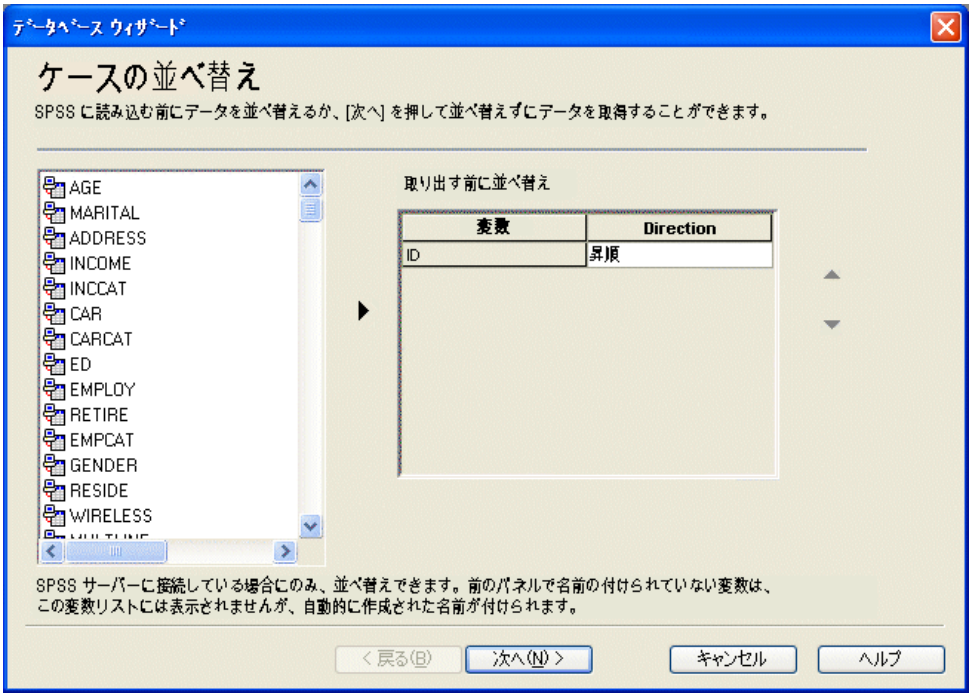
図 3-10
データベース ウィザード、変数の定義



ケースの並べ替え...

分散モードでリモート サーバー (SPSS サーバーで使用可能) に接続している場合は、データを SPSSに読み込む前に並べ替えできます。

図 3-11
データベース ウィザード、ケースの並べ替え



データは SPSSに読み込んだ後でも並べ替えることができますが、データ ソースが大きい場合は前もって並べ替えることにより時間を節約できることがあります。

結果

[結果] ダイアログ ボックスには、クエリーの SQL Select ステートメントが表示されます。

- クエリーを実行する前に SQL Select ステートメントを編集できますが、[**戻る**] ボタンをクリックして前のステップを変更すると、Select ステートメントの変更が失われます。
- クエリーを今後の使用のために保存しておくには、[**クエリーをファイルに保存**] セクションを使用します。
- GET DATA シンタックスをシンタックス ウィンドウに貼り付けるには、[**続けて修正するためシンタックス エディタに貼り付け**] を選択します。[結果] ウィンドウから Select ステートメントをコピーして貼り付けても、必要なコマンド シンタックスは貼り付けられません。

注: 貼り付けたシンタックスには、ウィザードで作成された SQL の各行の終わり引用符の前に、空白が含まれています。このような空白は、不要なものではありません。コマンドが処理されるとき、SQL ステートメントのすべての行は、そのままリテラルな形式で結合されます。スペースがないと、行の最後の文字と次の行の最初の文字との間にスペースがなくなります。

図 3-12
データベース ウィザード、結果パネル

テキスト ウィザード

テキスト ウィザードを使用すると、次のようなさまざまな書式のテキスト データ ファイルを読み込むことができます。

- タブ区切りファイル
- スペース区切りファイル
- カンマ区切りファイル
- 固定フィールド書式ファイル

自由書式ファイルでは、値と値の間に入れる区切り文字として他の文字を指定することもでき、さらに複数の区切り文字を指定することもできます。

テキスト データ ファイルを読み込むには

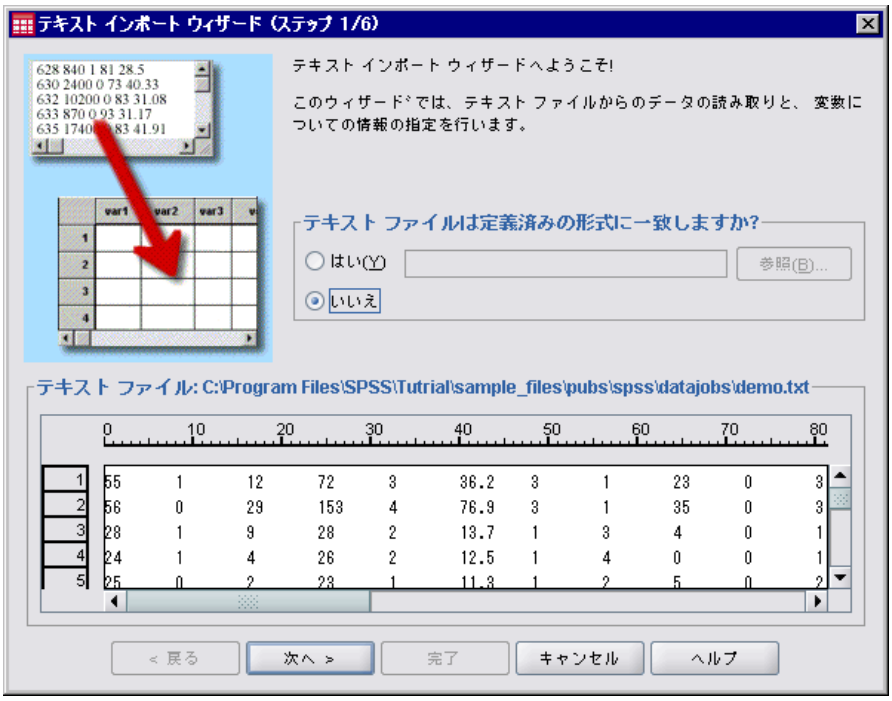
- メニューから次の項目を選択します。

  ファイル
  テキスト データの読み込み

- [ファイルを開く] ダイアログ ボックスで、テキスト ファイルを選択します。
- テキスト ウィザードの手順に従って、データ ファイルの読み込み方法を定義します。

テキスト ウィザード: ステップ 1

図 3-13
テキスト ウィザード: ステップ 1



プレビュー ウィンドウにテキスト ファイルが表示されます。定義済みの形式 (テキスト ウィザードですでに保存されている形式) を適用したり、テキスト ウィザードのステップに従って、データの読み込み方法を指定できます。

テキスト ウィザード: ステップ 2

図 3-14
テキスト ウィザード: ステップ 2

このステップでは、変数に関する情報を指定します。変数は、データベースのフィールドに似ています。たとえば、アンケートの各項目は変数です。

元データの形式データを正しく読み込むために、テキスト ウィザードでは、ある変数のデータ値がどこで終わり、次の変数のデータ値がどこで始まるかを判別する方法を指定する必要があります。変数の並べ方によって、前の変数と次の変数を区別する方法を定義します。

- **自由書式。**スペース、カンマ、タブ、またはその他の文字を使用して変数を区切ります。変数は各ケースに対して同じ順序で記録されますが、必ずしも同じ列位置ではありません。
- **固定書式。**各変数は、データ ファイルの各ケースに対して同じレコード (行) 上の同じ列位置に記録されます。変数と変数との間に区切り文字は必要ありません。実際、一般のコンピュータ プログラムで生成されたテキスト データ ファイルの多くは、データ値はスペースで区切られることもなく、列記されて表示されます。列位置により、読み込まれる変数が決まります。

ファイルの先頭に変数名を含んでいますか?データ ファイルの先頭の行に各変数を説明するラベルが含まれている場合は、それらのラベルを変数名として使用できます。変数の命名規則に従っていない値は、有効な変数名に変換されます。

テキスト ウィザード: ステップ 3 (自由書式)

図 3-15
テキスト ウィザード: ステップ 3 (自由書式用)

このステップでは、ケースに関する情報を指定します。ケースは、データベースのレコードに似ています。たとえば、アンケートのそれぞれの回答者は、1 つのケースです。

最初のケースの取り込み開始行番号。データ値が格納されているデータ ファイルの 1 行目を示します。データ ファイルの最初の行にデータ値を表さない説明ラベルまたはその他のテキストが含まれている場合、この値は 1 行目には**なりません**。

ケースの表される方法。各ケースの終わる位置と次のケースの始まる位置をテキスト ウィザードが判別する方法を設定します。

- **各行が 1 つのケースを表す。**各行には 1 つのケースしか入りません。大量の変数が含まれるデータ ファイルで非常に長い行になることがあっても、各ケースが 1 行に収められていることはよくあります。すべての行に同一個数のデータ値が含まれるわけではない場合、各ケースの変数の個数は、最大個数のデータ値が含まれる行によって決まります。データ値が少ないケースには、追加変数の欠損値が割り当てられます。
- **特定の数の変数が 1 つのケースを表す。**テキスト ウィザードは、各ケースの変数の指定された数から、前のケースの読み込みを停止する位置と次のケースの読み込みを開始する位置を判断します。同一行に多数のケースが含まれることがあり、またケースが前の行の真中から始まり、次の行まで続くということも

あります。テキスト ウィザードは、行数に関係なく、読み込んだ値の個数に基づいて各ケースの終わりを判別します。各ケースは、すべての変数に対してデータ値 (または区切り文字が示す欠損値) を含んでいる必要があります。そうでない場合、データ ファイルは正常に読み込まれません。

インポートするケース数。データ ファイル内のすべてのケース、最初の *n* 個のケース (*n* は指定した数)、または指定したパーセントの無作為抽出サンプルをインポートできます。無作為抽出ルーチンは各ケースに対して独立した疑似無作為決定を行うので、選択されたケースのパーセントを指定したパーセントに近づけることができます。データ ファイル内にケースが多くあるほど、選択されたケースのパーセントは指定したパーセントに近くなります。

テキスト ウィザード: ステップ 3 (固定書式)

図 3-16
テキスト ウィザード: ステップ 3 (固定書式用)

このステップでは、ケースに関する情報を指定します。ケースは、データベースのレコードに似ています。たとえば、アンケートの回答者は、1 つのケースになります。

最初のケースの取り込み開始行番号。データ値が格納されているデータ ファイルの 1 行目を示します。データ ファイルの最初の行にデータ値を表さない説明ラベルまたはその他のテキストが含まれている場合、この値は 1 行目には**なりません**。

1 つのケースを表す行数。各ケースの終わる位置と次のケースの始まる位置をテキスト ウィザードが判別する方法を設定します。各変数は、その変数のケース内の行数とその変数の列位置によって定義されます。データを正常に読み込むには、ケースごとに行数を指定する必要があります。

インポートするケース数。データ ファイル内のすべてのケース、最初の *n* 個のケース (*n* は指定した数)、または指定したパーセントの無作為抽出サンプルをインポートできます。無作為抽出ルーチンは各ケースに対して独立した疑似無作為決定を行うので、選択されたケースのパーセントを指定したパーセントに近づけることができます。データ ファイル内にケースが多くあるほど、選択されたケースのパーセントは指定したパーセントに近くなります。

テキスト ウィザード: ステップ 4 (自由書式)

図 3-17
テキスト ウィザード: ステップ 4 (自由書式用)

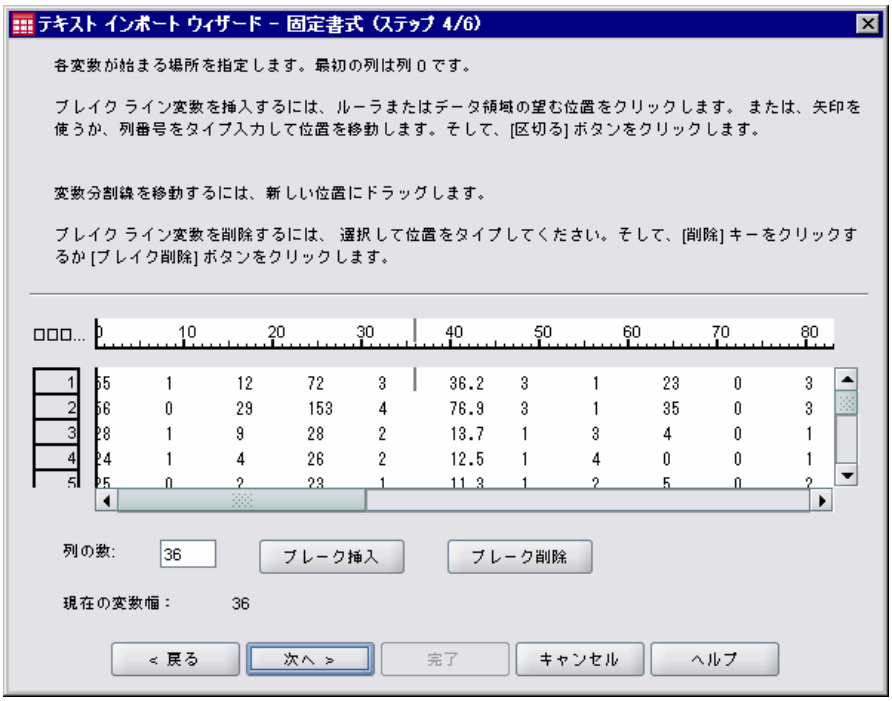
このステップでは、最良と思われるデータ ファイルの読み込み方法が表示されます。また、データ ファイルからテキスト ウィザードに変数を読み込ませる方法を修正できます。

変数間に使用する区切り記号。データ値を区切る文字または記号を示します。スペース、カンマ、セミコロン、タブ、その他の文字を組み合せて選択できます。データ値が間に挟まっていない複数の連続する区切り文字は、欠損値として取り扱われます。

テキスト修飾子について。区切り記号を含む値を囲むために使用する文字のことです。たとえば、カンマが区切り記号の場合、値にカンマが含まれていると、テキスト修飾子で値を囲まなければ、その値は正しく読み込まれません。テキスト修飾子で値を囲むことにより、値の中のカンマが区切り記号と解釈されないようにします。Excel からエクスポートされた CSV 形式のデータ ファイルは、テキスト修飾子として二重引用符 (") を使用しています。テキスト修飾子は値の始まりと終わりの両方に使用され、値全体を囲みます。

テキスト ウィザード: ステップ 4 (固定書式)

図 3-18
テキスト ウィザード: ステップ 4 (固定書式用)



このステップでは、最良と思われるデータ ファイルの読み込み方法が表示されます。また、データ ファイルからテキスト ウィザードに変数を読み込ませる方法を修正できます。プレビュー ウィンドウの垂直線は、そのファイルでテキスト ウィザードが現在認識している各変数の開始位置を示しています。

変数を区切るため、必要に応じて変数分割行を挿入、移動、および削除できます。各ケースに複数行が使用されると、行の終わりに追加した次の行に各ケース 1 行としてデータが表示されます。

注:

コンピュータ生成のデータ ファイルが、間にスペースやその他の識別文字が入らない、連続するデータ値列で構成されていると、各変数はどこから始まるのか判別が難しい場合があります。このようなデータ ファイルは、通常、各変数の行および列位置を指定したデータ定義ファイルまたはその他の記述書に依存します。

テキスト ウィザード: ステップ 5

図 3-19
テキスト ウィザード: ステップ 5

このステップでは、テキスト ウィザードで各変数を読み込むときに使用する変数名とデータ形式、さらに最終的なデータ ファイルに含める変数を設定します。

変数名。デフォルトの変数名を使用する変数名で上書きできます。データ ファイルから変数名を読み込むとき、変数の命名規則に従っていない変数名はテキスト ウィザードによって自動的に変更されます。プレビュー ウィンドウで変数を選択して、変数名を入力します。

データ形式。プレビュー ウィンドウで変数を選択し、ドロップダウン リストからデータ形式を選択します。複数の変数を続けて選択するには Shift キーを押しながらクリックします。複数の変数を個別に選択するには Ctrl キーを押しながらクリックします。

テキスト ウィザードのデータ形式設定オプション

テキスト ウィザードで変数を読み込むためのデータ形式の設定オプションには、次のものがあります。

インポートしない。インポートされたデータ ファイルから、選択された変数を除去します。

数値型。有効な値には、数字、数字の前のプラスまたはマイナス符号、および小数点があります。

文字列。有効な値には、キーボード上にあるほぼすべての文字および埋め込まれた空白があります。自由書式ファイルでは、値の文字数を最大 32,767 文字まで指定できます。デフォルトでは、テキスト ウィザードは、選択された変数について見つかった最長の文字列値に、文字数を設定します。固定書式ファイルでは、文字列値の文字数は、ステップ 4 で変数分割行を配置することにより定義されます。

日時。有効な値には、dd-mm-yyyy、mm/dd/yyyy、dd.mm.yyyy、yyyy/mm/dd、hh:mm:ss といった一般的な書式の日付や、その他の各種日付および時刻書式があります。月は、数字、ローマ数字、3 文字略語で示すこともでき、さらに略さずに綴ることもできます。リストからデータ形式を選択します。

ドル記号。有効な値は、数字の前のオプションのドル記号、および 1000 単位を表す記号として、オプションのカンマを含む数字です。

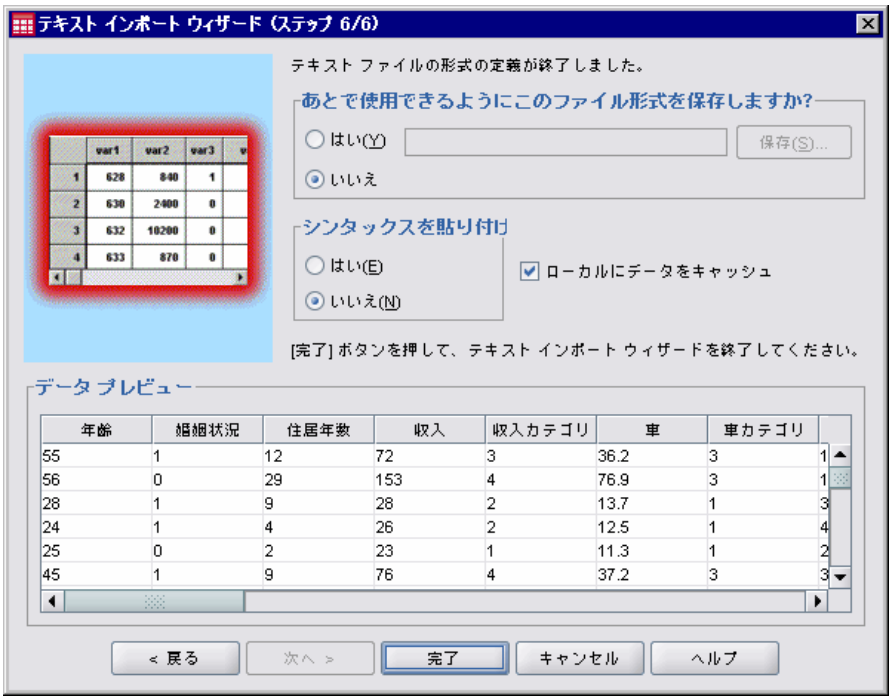
カンマ。有効な値には、小数点としてピリオドを使用し、1000 単位を表す記号としてカンマを使用する数値があります。

ドット。有効な値には、小数点としてカンマを使用し、1000 単位を表す記号としてピリオドを使用する数値があります。

注: 選択された書式に合わない無効な文字を含む値は、欠損値として処理されます。指定された区切り記号を含む値は、多重値として取り扱われます。

テキスト ウィザード: ステップ 6

図 3-20
テキスト ウィザード: ステップ 6



これは、テキスト ウィザードの最終ステップです。これまでの指定をファイルに保存しておき、類似のテキスト データ ファイルをインポートするときに使用できます。また、テキスト ウィザードで生成されたシンタックスを、シンタックス ウィンドウに貼り付けることもできます。その後で、他のセッションやバッチ機能で使用できるように、シンタックスをカスタマイズまたは保存できます。

ローカルにデータをキャッシュ。データ キャッシュはデータ ファイルの完全なコピーで、ディスクの一時領域に保持されます。データ ファイルのキャッシュを作成すると、性能を向上させることができます。

Dimensions データの読み込み

Microsoft Windows のオペレーティング システム上は、Quanvert、Quancept、mrInterview など SPSS Dimensions 製品からデータを読み込むことができます。(**注**: この機能は、Microsoft Windows のオペレーティング システム上にインストールされた SPSSと一緒のときにだけ使用できます)。

Dimensions データ ソースを読み込むには、次のコンポーネントがインストールされている必要があります。

- .NET framework
- Dimensions Data Model および OLE DB Access

本リリースの SPSS と互換性があるこれらのコンポーネントのバージョンは、インストール CD からインストールすることができます。また、Autoplay メニューからインストールすることもできます。Dimensions データ ソースは、ローカル アナリシス モードでだけ読み込めます。この機能は、SPSS サーバーを使用するディストリビュート アナリシス モードでは利用できません。

Dimensions データ ソースからデータを読み込む手順は次のとおりです。

▶ 開いている SPSS ウィンドウから、次のメニューを選択します。
ファイル
Dimensions データを開く

▶ [データ リンク プロパティ] の [接続] タブで、メタデータ ファイル、ケース データ タイプ、およびケース データ ファイルを指定します。

▶ [OK] をクリックします。

▶ [Dimensions データのインポート] ダイアログ ボックスで、取り込む変数、およびケースの選択基準を選択します。

▶ [OK] をクリックすると、データが読み込まれます。

[データ リンク プロパティ] の [接続] タブ

Dimensions データを読み込むには、次の項目を指定する必要があります。

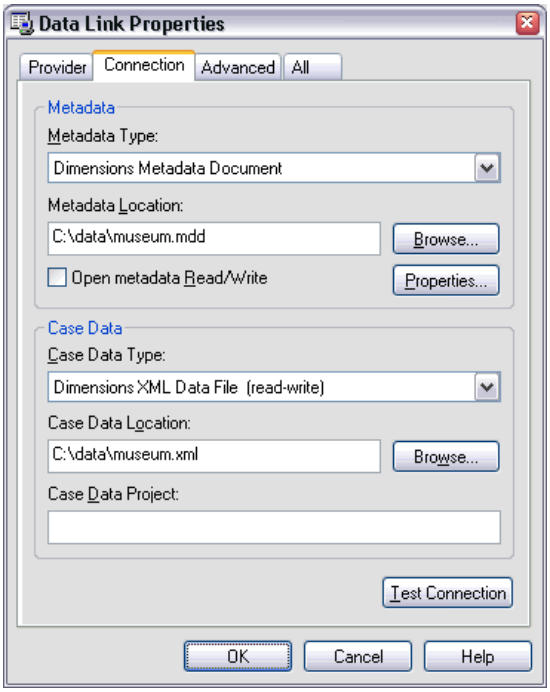
メタデータの場所。調査票定義情報を含むメタデータ文書ファイル (*.mdd*)。

ケース データ タイプ。ケース データ ファイルの形式。使用できる形式は次のとおりです。

- Quancept Data File (DRS)。Quancept の **.drs** ファイル、**.drz** ファイル、または **.dru** ファイル形式のケース データ。
- Quanvert Database。Quanvert データベース形式のケース データ。
- Dimensions Database (MS SQL Server)。SQL Server における SPSS MR リレーショナル データベース形式のケース データ。このオプションは、mrInterview で収集されたデータを読み込む場合に使用できます。
- Dimensions XML Data File。XML ファイル形式のケース データ。

ケース データの場所。ケース データを含むファイル。このファイルの形式は、選択したケース データ タイプと対応している必要があります。

図 3-21
[データ リンク プロパティ] の [接続] タブ



注: [接続] タブのその他の設定内容、および [データ リンク プロパティ] のその他のタブにおける設定内容が、Dimensions データを SPSS に読み込む上でどの程度の影響を及ぼすか、または影響を及ぼさないのかについては不明です。そのため、これらの設定内容は変更しないことをお勧めします。

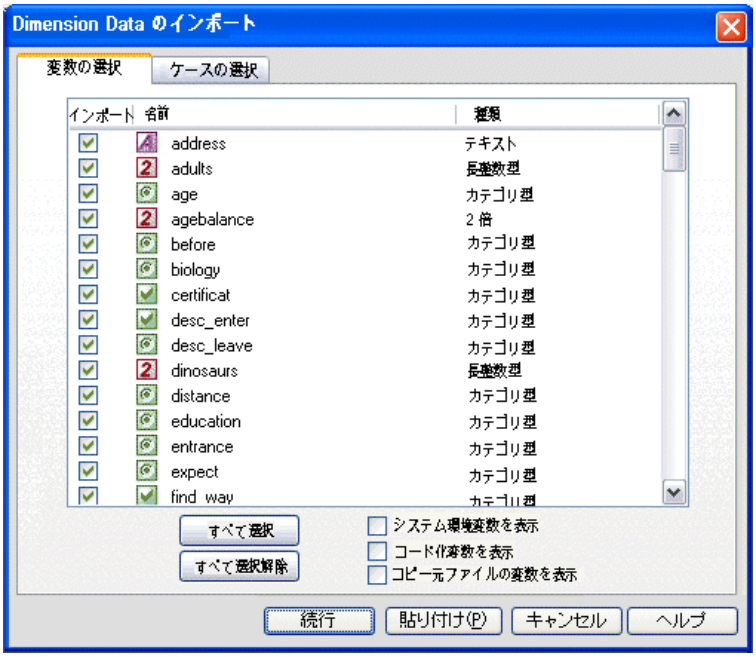
[変数の選択] タブ

読み込みの対象とする変数は複数をまとめて選択できます。デフォルトでは、データ ソース内の標準的な変数はすべて表示/選択されます。

- **システム変数を表示。**インタビューのステータス (**処理中**、**完了**、**完了日**など) を表す変数などの「システム」変数が表示されます。またその中から、取り込むシステム変数を選択することができます。デフォルトでは、すべてのシステム変数が対象から除外されます。

- **コード変数を表示。**カテゴリ変数に対する自由回答形式の「その他」という回答に使用されるコードを表す変数が表示されます。またその中から、取り込むコード変数を選択することができます。デフォルトでは、すべてのコード変数が対象から除外されます。
- **ソース ファイル変数を表示。**スキャンされた回答のイメージのファイル名を含む変数が表示されます。またその中から、取り込むソース ファイル変数を選択することができます。デフォルトでは、すべてのソース ファイル変数が対象から除外されます。

図 3-22
[Dimensions データのインポート: 変数の選択] タブ



[ケースの選択] タブ

システム変数を含む Dimesions データ ソースに対しては、さまざまなシステム変数基準に基づいてケースを選択できます。読み込む変数のリストに目的のシステム変数を含める必要はありませんが、選択基準を適用するためには、必要なシステム変数がソース データの中に存在する必要があります。ソース データの中に必要なシステム変数が存在しない場合、目的の選択基準は適用されません。

データ コレクションのステータス。回答者データ、テスト データ、またはその両方を選択できます。ケースの選択は、次のインタビュー ステータス パラメータの組み合わせを基準にして行うこともできます。

- 正常に完了
- アクティブ / 処理中

- タイム アウト
- スクリプトによって停止
- 回答者によって停止
- インタビュー システムの停止
- 通知 (スクリプト内の通知ステートメントによって停止)

データ コレクションの完了日付。データ コレクションの完了日付に基づいてケースを選択できます。

- **開始日付。**指定した日付またはそれ以降にデータ コレクションが完了したケースが含まれます。
- **終了日付。**指定した日より前にデータ コレクションが完了したケースが含まれます。この場合、終了日付にデータ コレクションが完了したケースは**含まれません**。
- 開始日付と終了日付をともに指定した場合、開始日付から終了日付までの範囲 (ただし終了日付は含まない) がすべて完了日として定義されます。

図 3-23
[Dimensions データのインポート: ケースの選択] タブ

ファイル情報

データ ファイルに記録されているのは、生データだけではありません。次の変数定義情報も含まれています。

- 変数名
- 変数の形式
- 記述的な変数ラベルと値ラベル

上記の情報は、データ ファイルの辞書部分に保持されます。データ エディタを使用して、変数定義情報を表示することもできます。アクティブなデータセットまたは任意のデータ ファイルの完全な辞書情報を表示することもできます。

データ ファイル情報を表示するには

▶ [データ エディタ] ウィンドウのメニューから、次の項目を選択します。

ファイル
データ ファイル情報の表示

▶ 現在開いているデータ ファイルの情報を表示する場合は、[**作業ファイル**] を選択します。

▶ その他のデータ ファイルの情報を表示するには、[**外部ファイル**] を選択し、目的のデータ ファイルを選択します。

データ ファイル情報が、ビューアに表示されます。

データ ファイルの保存

データ ファイルを SPSS 形式で保存するだけでなく、データを次に示すさまざまな外部形式で保存することもできます。

- Excel などのスプレッドシート形式
- タブ区切りのテキスト ファイルおよび CSV テキスト ファイル
- SAS
- Stata
- データベース テーブル

変更したデータ ファイルを保存するには

▶ データ エディタをアクティブ ウィンドウにします (アクティブにするウィンドウのどこかをクリックします)。

▶ メニューから次の項目を選択します。

ファイル
データの上書き保存

修正されたデータ ファイルが保存されて、ファイルの以前のバージョンが上書きされます。

注: Unicode モードでデータ ファイルを保存すると 16.0 以前の SPSS のバージョンでは読み込むことができません。以前のリリースにより読み込まれることができるフォーマットの Unicode データ ファイルを保存するには、コード ページ モードのファイルを開けて、再度保存します。ファイルは現在のロケールのエンコードで保存されます。ファイルが現在のロケールによって認識されない文字を含んでいる場合は、データが損失するかもしれません。Unicode モードおよびコード ページ モード間の切り替えの情報は、「全般」を参照してください 498 ページの。

データ ファイルを外部形式で保存するには

- データ エディタをアクティブ ウィンドウにします (アクティブにするウィンドウのどこかをクリックします)。
- メニューから次の項目を選択します。

  ファイル
  名前を付けて保存...
- 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。
- 新しいデータ ファイルのファイル名を入力します。

スプレッドシートまたはタブ区切りデータ ファイルの最初の行に変数名を書き込むには

- [名前を付けてデータを保存] ダイアログ ボックスの **[スプレッドシートへの変数名の書き込み]** をクリックします。

データ値ではなく、値ラベルを Excel ファイルで保存するには

- [名前を付けてデータを保存] ダイアログ ボックスの **[データ値の替わりに定義場所のデータラベルを保存]** をクリックします。

値ラベルを SAS シンタックス ファイル (SAS ファイルの種類を選択している場合にだけアクティブになります) に保存するには

- [名前を付けてデータを保存] ダイアログ ボックスの **[値ラベルを .sas ファイルに保存]** をクリックします。

データをデータベースにエクスポートする方法の詳細は、55 ページのデータベースへのエクスポートを参照してください。

データを Dimensions アプリケーション用にエクスポートする方法の詳細は、69 ページのDimensions へのエクスポートを参照してください。

データの保存: データ ファイルの種類

データは次の形式で保存できます。

SPSS (*.sav)。 SPSS 形式。

- SPSS 形式で保存されたデータ ファイルは、7.5 以前のバージョンの SPSS では読み込めません。Unicode モードでデータ ファイルを保存すると 16.0 以前の SPSS のリリースでは読み込むことができません。詳細は、498 ページの 45 章の[全般] オプションを参照してください。
- SPSS 10.x または 11.x で 8 バイトを超える変数名を持つデータ ファイルを使用すると、8 バイトに切り詰められた一意な変数名が使用されます。ただし、元の変数名は、リリース 12.0 以降で使用できるように保持されます。10.0 より前のリリースでは、データ ファイルを保存すると、元の長い変数名は失われます。
- 13.0J より前のリリースで 255 バイトを超える文字型変数を持つデータ ファイルを使用すると、それらの文字型変数が 255 バイトごとに細分化され複数の文字型変数になります。

バージョン 7.0 (*.sav)。 バージョン 7.0 形式。バージョン 7.0 形式で保存されたデータ ファイルは、SPSS 7.0 および以前のバージョンでも読み込めますが、定義された多重回答グループまたは Data Entry for Windows の情報は含まれません。

SPSS/PC+ (*.sys)。 SPSS/PC+ 形式。データ ファイルに 500 以上の変数が含まれている場合は、最初の 500 だけが保存されます。定義されたユーザー欠損値が複数ある変数では、すべてのユーザー欠損値は最初に定義されたユーザー欠損値に再割り当てされます。

SPSS Portable (*.por)。 他のオペレーティング システム上で他のバージョンの SPSS を使用して読み取ることができる SPSS ポータブル形式。変数名は 8 バイトに制限され、必要に応じて 8 バイトの一意な名前に自動的に変換されます。SPSS 形式のデータ ファイルはプラットフォーム/オペレーティング システムに依存するため、ほとんどの場合、ポータブル形式で保存されたデータは必要ではなくなっています。Unicode モードではポータブル形式でデータを保存することはできません。詳細は、498 ページの 45 章の[全般] オプションを参照してください。

タブ区切り (*.dat)。 各値がタブで区切られているテキスト ファイル。((**注**: 文字列値に埋め込まれたタブ文字は、タブ区切りファイル内でもタブ文字として保存されます。ただし値に埋め込まれたタブ文字は、値の区切り記号としてのタブ文字とは区別されません)。

カンマ区切り (*.csv)。 各値がカンマまたはセミコロンで区切られているテキスト ファイル。SPSS の小数点としてピリオドが使用されている場合、値の区切り記号にはカンマが使用されます。小数点としてカンマが使用されている場合、値の区切り記号にはセミコロンが使用されます。

固定 ASCII (*.dat)。 すべての変数に対してデフォルトの書き込み書式を使用した固定書式のテキスト ファイル。変数フィールド間にはタブもスペースもありません。

Excel 2007 (*.xlsx)。 Microsoft Excel 2007 XLSX-形式ワークブック。最大変数値は 16,000 で、16,000 を超える追加の変数は削除されます。データセットが 100 万単位ものケースが含まれるような場合は、ワークブックで複数のシートが作成されます。

Excel 97 ～ 2003 (*.xls)。 Microsoft Excel 97 ワークブック。最大変数値は 256 で、256 を超える追加の変数は削除されます。データセットが 65,356 以上のケースが含まれるような場合は、ワークブックで複数のシートが作成されます。

Excel 2.1(*.xls)。 Microsoft Excel .2.1 のスプレッドシート ファイル。最大変数値は 256 で、最大行数は 16,384 です。

1-2-3 リリース 3.0 (*.wk3)。 Lotus 1-2-3 スプレッドシート ファイル、リリース 3.0。保存できる変数の最大数は、256 です。

1-2-3 リリース 2.0 (*.wk1)。 Lotus 1-2-3 スプレッドシート ファイル、リリース 2.0。保存できる変数の最大数は、256 です。

1-2-3 リリース 1.0 (*.wks)。 Lotus 1-2-3 スプレッドシート ファイル、リリース 1A。保存できる変数の最大数は、256 です。

SYLK (*.slk)。 Microsoft Excel および Multiplan スプレッドシート ファイル用のシンボリック リンク形式。保存できる変数の最大数は、256 です。

dBASE IV (*.dbf)。 dBASE IV 形式。

dBASE III (*.dbf)。 dBASE III 形式。

dBASE II (*.dbf)。 dBASE II 形式。

SAS v7+ Windows short extension (*.sd7)。 SAS バージョン 7 ～ 8 for Windows の短いファイル名形式。

SAS v7+ Windows long extension (*.sas7bdat)。 SAS バージョン 7 ～ 8 for Windows の長いファイル名形式。

SAS v7+ for UNIX (*.ssd01)。 SAS v8 for UNIX。

SAS v6 for Windows (*.sd2)。 SAS V6 for Windows/OS2 ファイル形式。

SAS v6 for UNIX (*.ssd01)。 SAS V6 for UNIX (Sun、HP、IBM) ファイル形式。

SAS v6 for Alpha/OSF (*.ssd04)。 SAS V6 for Alpha/OSF (DEC UNIX) ファイル形式。

SAS Transport (*.xpt)。 SAS トランスポート ファイル。

Stata バージョン 8 Intercooled (*.dta)。

Stata バージョン 8 SE (*.dta)。

Stata バージョン 7 Intercooled (*.dta)。

Stata バージョン 7 SE (*.dta)。

Stata バージョン 6 (*.dta)。

Stata バージョン 4 ～ 5 (*.dta)。

ファイルの保存のオプション

スプレッドシート ファイル、タブ区切りファイル、およびカンマ区切りファイルでは、ファイルの最初の行に変数名を書き込むことができます。

Excel 形式でのデータ ファイルの保存

データは、3 種類の Microsoft Excel ファイル形式のいずれかの形式で保存できます。Excel 2.1、Excel 97、および Excel 2007。

- Excel 2.1 および Excel 97 は 256 列に制限されるので、最初の 256 変数のみが含まれます。
- Excel 2007 は 16,000 列に制限されるので、最初の 16,000 変数のみが含まれます。
- Excel 2.1 は 16,384 行に制限されるので、最初の 16,384 ケースのみが含まれます。
- Excel 97 および Excel 2007 はまたシートごとの行数制限がありますが、ワークブックは複数のシートがあり、単一シートの最大数を超えた場合複数のシートが作成されます。

変数の種類

次の表は、SPSS の元のデータと Excel にエクスポートされたデータ間で一致する変数の型を示します。

SPSS変数の型	Excel データ形式
数値	0.00; #,##0.00; ...
カンマ	0.00; #,##0.00; ...
ドル記号	$#,##0_); ...
日付	d-mmm-yyyy
時間	hh:mm:ss
文字列	一般文字列

SAS 形式でのデータ ファイルの保存

データを SAS ファイルで保存する場合には、データのさまざまな面に、特別な処理が施されます。次にその内容を示します。

- @、#、$ など、SPSS の変数名では許可されている特定の文字が SAS では無効です。データをエクスポートすると、これらの無効な文字は、アンダースコアに置き換えられます。
- 日本や中国などで使用されるマルチバイト文字を含む SPSS 変数名は、一般的な Vnnn 形式の変数名に変換されます (nnn は整数値)。

- 40 文字以上で構成される SPSS の変数ラベルは、SAS v6 ファイルにエクスポートすると切り捨てられます。
- SPSS の変数ラベルが存在する場合は、SAS の変数ラベルにマッピングされます。SPSS データ内に変数ラベルが存在しない場合は、変数名が SAS 変数ラベルにマッピングされます。
- SAS では、システム欠損値に対して 1 つの値しか許可されていませんが、SPSS では多数許可されています。そのため、SPSS のすべてのシステム欠損値が、SAS ファイル内にある 1 つのシステム欠損値にマッピングされます。

値ラベルの保存

データ ファイルと関連付けられた値および値ラベルをシンタックス ファイルに保存するオプションがあります。たとえば、**cars.sav** データ ファイルの値ラベルをエクスポートすると、生成されるシンタックス ファイルには次のものが格納されます。

```
libname library '\spss\' ;

proc format library = library ;
   value ORIGIN /* Country of Origin */
      1 = 'American'
      2 = 'European'
      3 = 'Japanese' ;
   value CYLINDER /* Number of Cylinders */
      3 = '3 Cylinders'
      4 = '4 Cylinders'
      5 = '5 Cylinders'
      6 = '6 Cylinders'
      8 = '8 Cylinders' ;
   value FILTER__ /* cylrec = 1 | cylrec = 2 (FILTER) */
      0 = 'Not Selected'
      1 = 'Selected' ;

proc datasets library = library ;
modify cars;
   format    ORIGIN ORIGIN.;
   format  CYLINDER CYLINDER.;
   format  FILTER__ FILTER__.;
quit;
```

この機能は、SAS トランスポート ファイルではサポートされていません。

変数の種類

次の表は、SPSS の元のデータと SAS にエクスポートされたデータ間で一致する変数の型を示します。

SPSS変数の型	SAS 変数の型	SAS データ形式
数値	数値	12
カンマ	数値	12
ドット (.)	数値	12
科学的表記法	数値	12
日付	数値	(日付) 例、MMDDYY10, ...
日付 (時間)	数値	Time18
ドル記号	数値	12
通貨書式	数値	12
文字列	文字	$8

Stata 形式でのデータ ファイルの保存

- データの書き出しは、Stata 5 ～ 8 形式および Intercooled と SE の両形式 (バージョン 7 と 8 のみ) で行うことができます。
- Stata 5 形式で保存したデータ ファイルは、Stata 4 で読み込むことができます。
- 変数ラベルの最初の 80 バイトが Stata 変数ラベルとして保存されます。
- 数値型変数では、値ラベルの最初の 80 バイトが Stata 値ラベルとして保存されます。文字列型変数では、値ラベルは破棄されます。
- バージョン 7～8 では、大文字小文字を区別する形式の変数名の最初の 32 バイトが、Stata 変数名として保存されます。それより前のバージョンでは、変数名の最初の 8 バイトが Stata 変数名として保存されます。英字でも、数字でも、下線でもない文字は、下線に変換されます。
- 日本や中国などで使用されるマルチバイト文字を含む SPSS 変数名は、一般的な Vnnn 形式の変数名に変換されます (nnn は整数値)。

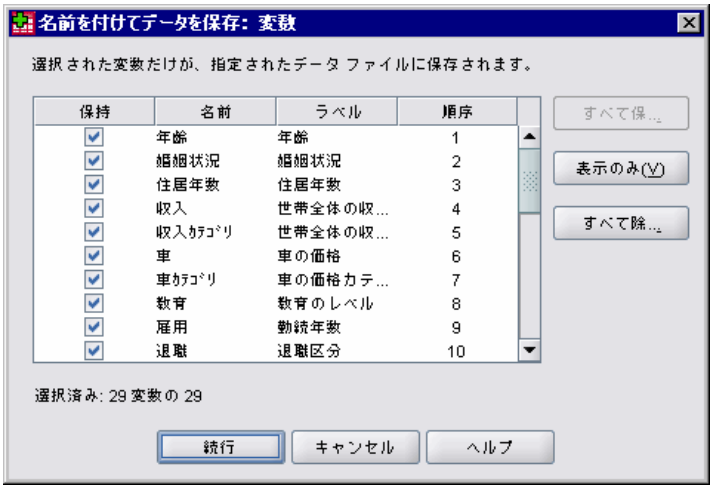
- バージョン 5 ～ 6 および Intercooled バージョン 7 ～ 8 では、文字列型値の最初の 80 バイト が保存されます。Stata SE バージョン 7 ～ 8 では、文字列型値の最初の 244バイト が保存されます。
- バージョン 5 ～ 6 および Intercooled バージョン 7 ～ 8では、最初の 2,047 個の変数だけが保存されます。Stata SE バージョン 7 ～ 8 では、最初の 32,767 個の変数だけが保存されます。

SPSS変数の型	Stata 変数の型	Stata データ形式
数値	数値	g
カンマ	数値	g
ドット (.)	数値	g
科学的表記法	数値	g
日付*、日時	数値	D_m_Y
時刻、日時	数値	g (秒数)
曜日	数値	g (1 ～ 7)
月	数値	g (1 ～ 12)
ドル記号	数値	g
通貨書式	数値	g
文字列	文字列	s

* 日付、アメリカ式日付、ヨーロッパ式日付、SDate、ユリウス日付、四半期と年、月と年、週と年

変数のサブグループの保存

図 3-24
[名前を付けてデータを保存: 変数] ダイアログ ボックス



[名前を付けてデータを保存: 変数] ダイアログ ボックスでは、新しいデータ ファイルに保存する変数を選択できます。デフォルトでは、すべての変数が保存されます。保存しない変数の選択を解除するか、**[すべて除外]** をクリックしてから、保存する変数を選択します。

表示のみ。現在使用中の変数グループに属する変数だけを選択します。詳細は、495 ページの 44 章の変数グループの使用を参照してください。

変数のサブセットを保存するには

- データ エディタをアクティブ ウィンドウにします (アクティブにするウィンドウのどこかをクリックします)。
- メニューから次の項目を選択します。

  ファイル
  名前を付けて保存...
- **[変数]** をクリックします。
- 保存する変数を選択します。

データベースへのエクスポート

データベースへのエクスポート ウィザードでは次の処理を実行できます。

- 既存のデータベース テーブルのフィールド (列) にある値の置換、またはテーブルへの新しいフィールドの追加。
- データベース テーブルへの新しいレコード (行) の追加。
- データベース テーブル全体の入れ替え、またはテーブルの新規作成。

データをデータベースにエクスポートするには

- エクスポートするデータが含まれるデータセットの [データ エディタ] ウィンドウから、次のメニュー項目を選択します。

  ファイル
  データベースにエクスポート
- データベース ソースを選択します。
- エクスポート ウィザードの指示に従って、データをエクスポートします。

SPSS 変数からのデータベース フィールドの作成

フィールドを新規作成する (フィールドを既存のデータベース テーブルへ追加する、テーブルを新規作成する、テーブルを入れ替える) 場合は、フィールドの名前、データ型、および幅 (設定可能な場合) を指定できます。

フィールド名。デフォルトのフィールド名には、SPSS 変数名が使用されます。このフィールド名は、データベースでの書式に従ったものであれば、任意の名前に変更できます。多くのデータベースでは、空白文字など、変数名には使用できない文字でもフィールド名に使用できます。したがって、たとえば CallWaiting という変数名を Call Waiting というフィールド名に変更することもできます。

型。エクスポート ウィザードでは、標準の ODBC データ型、または選択したデータベース書式で使用できるもののうち定義済みの SPSS データ形式に最も近いデータ型に基づいて、データ型の初期割り当てが行われます。ただし、データベースと SPSS のそれぞれで使用されるデータ型の種類は、完全には対応していません。たとえば、SPSS の数値はそのほとんどが、倍精度浮動小数点値として保存されるのに対し、データベースでは数値のデータ型として、浮動小数点型 (倍精度)、整数型、実数型などが使用できます。また多くのデータベースでは、SPSS の時間書式に対応する書式がありません。データ型は、ドロップダウン リストから選択できる任意のデータ型に変更できます。

一般に、変数の基本的なデータ型 (文字型や数値型) は、データベース フィールドの基本的なデータ型に対応しています。データベースでは解決できないようなデータ型の不整合がある場合は、エラーが発生し、データベースに対するデータのエクスポートは行われません。たとえば、数値データ型を持つデータベース フィールドに文字型変数をエクスポートした場合、その文字型変数のいずれかの値に非数値型文字が含まれていると、エラーが発生することになります。

幅。文字型 (文字列型、可変長文字列型) フィールドに対して定義されている幅を変更できます。数値型フィールドの幅は、それぞれのデータ型により定義されます。

SPSS 変数の書式からデータベース フィールドのデータ型へのマッピングは、デフォルトでは、次のような一般的な規則に従って行われます。データベース フィールドの実際のデータ型は、使用するデータベースによっては異なる場合があります。

SPSS変数の形式	データベース フィールドのデータ型
数値	浮動小数点型または倍精度浮動小数点型
カンマ	浮動小数点型または倍精度浮動小数点型
ドット (.)	浮動小数点型または倍精度浮動小数点型
科学的表記法	浮動小数点型または倍精度浮動小数点型
日付	日付型、日時型、またはタイムスタンプ型
日時	日時型またはタイムスタンプ型
時刻、日時	浮動小数点型または倍精度浮動小数点型 (秒単位)
曜日	整数型 (1 ～ 7)
月	整数型 (1 ～ 12)
ドル記号	浮動小数点型または倍精度浮動小数点型
通貨書式	浮動小数点型または倍精度浮動小数点型
文字列	文字列型または可変長文字列型

ユーザー欠損値

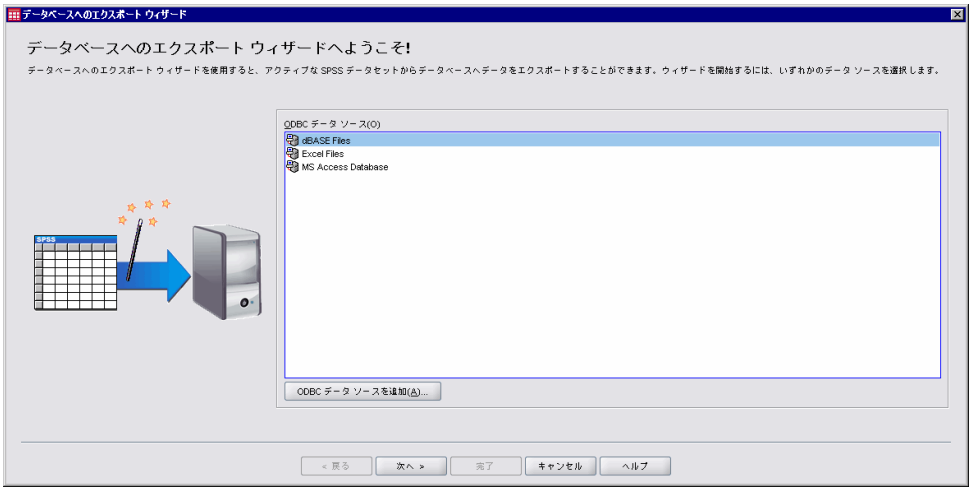
変数のデータをデータベース フィールドにエクスポートする場合のユーザー欠損値の処理に対しては、次の 2 つのオプションを使用できます。

- **有効値としてエクスポート。**ユーザー欠損値は、通常の有効な非欠損値として処理されます。
- **数値型のユーザー欠損値を null、文字型のユーザー欠損値を空白スペースとしてエクスポート。**数値型のユーザー欠損値が、システム欠損値と同様に処理されます。また文字型のユーザー欠損値は、空白文字に変換されます (文字列はシステム欠損値として扱うことはできません)。

データ ソースの選択

データベースへのエクスポート ウィザードの最初の画面で、データのエクスポート先となるデータ ソースを選択します。

図 3-25
データベースへのエクスポート ウィザード、データ ソースの選択



データのエクスポートは、適切な ODBC ドライバがインストールされている任意のデータベース ソースに対して行えます。(**注**: OLE DB データ ソースへのエクスポートはサポートされていません)。

ODBC データ ソースが設定されていない場合、または新しいデータ ソースを追加する場合は、[**ODBC データ ソースの追加**] をクリックします。

- Linux オペレーティング システムでは、このボタンは利用できません。ODBC データ ソースは odbc.ini で指定され、ODBCINI 環境変数はファイルの位置に設定されなければいけません。詳細は、データベース ドライバのドキュメントを参照してください。
- ディストリビュート アナリシス モード (SPSS サーバーで使用可能) では、このボタンはありません。ディストリビュート アナリシス モードでデータ ソースを追加する方法については、システム管理者に問い合わせてください。

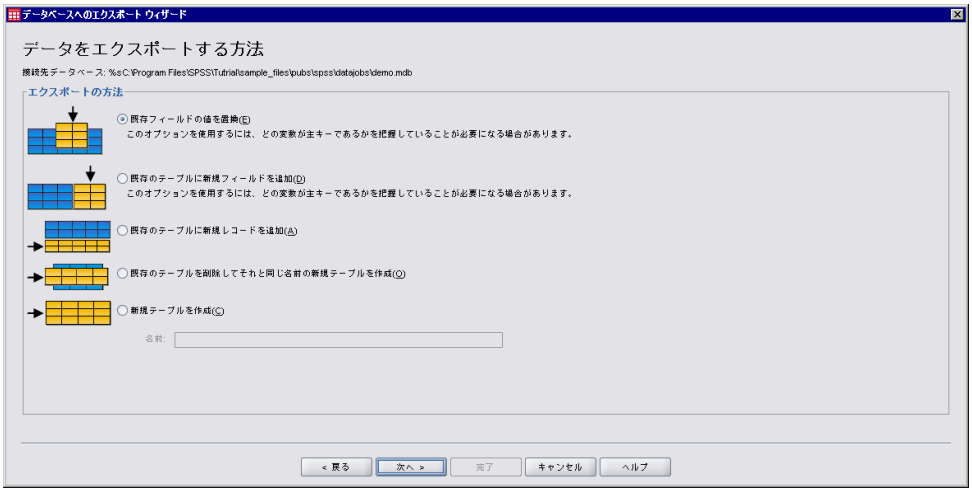
ODBC データ ソースは、2 つの重要な情報から成り立っています。1 つはデータのアクセスに使用されるドライバについての情報、もう 1 つはアクセスするデータベースの場所についての情報です。データ ソースを指定するには、まず、適切なドライバをインストールしておく必要があります。ローカル アナリシス モードでは、SPSS のインストール CD からいろいろなデータベース形式のドライバをインストールできます。

一部のデータ ソースでは、次の手順に進む前に、ログイン ID とパスワードの入力が必要になることがあります。

データのエクスポート方法の選択

データ ソースを選択したら、次にデータのエクスポート方法を指定します。

図 3-26
データベースへのエクスポート ウィザード、エクスポート方法の選択



データをデータベースにエクスポートする場合、次のオプションを選択できます。

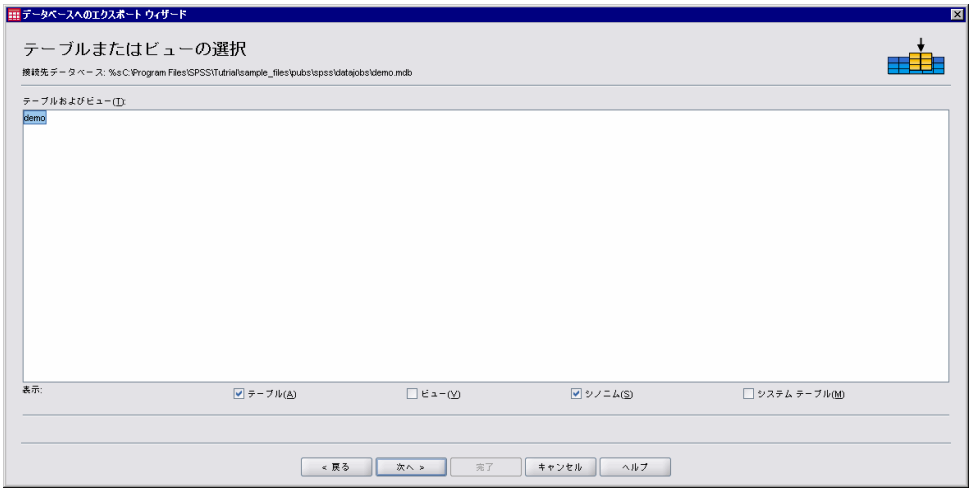
- **既存フィールドの値を置換。**既存テーブルから選択したフィールドの値を、アクティブなデータセットで選択した変数の値に置き換えます。 詳細は、62 ページの既存フィールドの値の置換を参照してください。

- **既存のテーブルに新規フィールドを追加。**既存のテーブルに新規フィールドを作成し、そこへアクティブなデータセットで選択した変数の値を格納します。詳細は、63 ページの新規フィールドの追加を参照してください。このオプションは、Excel ファイルに対しては使用できません。
- **既存のテーブルに新規レコードを追加。**既存テーブルに新規レコード (行) を追加し、そこへアクティブなデータセット内のケースの値を格納します。詳細は、64 ページの新規レコード (ケース) の追加を参照してください。
- **既存のテーブルを削除してそれと同じ名前の新規テーブルを作成。**指定したテーブルを削除して、それと同じ名前の新規テーブルを作成し、そこへアクティブなデータセットで選択した変数を格納します。フィールド プロパティの定義 (主キー、データ型など) を含め、元のテーブルに関する情報はすべて失われます。詳細は、66 ページの新規テーブルの作成またはテーブルの置換を参照してください。
- **新規テーブルを作成。**新規テーブルをデータベースに作成し、そこへアクティブなデータセットで選択した変数のデータを格納します。名前には、データソースで有効なテーブル名であれば、どのような値でも使用できます。ただし、データベース内の既存テーブルまたは既存ビューと重複する名前は指定できません。詳細は、66 ページの新規テーブルの作成またはテーブルの置換を参照してください。

テーブルの選択

データベース内のテーブルを修正するまたは置き換える場合は、その対象となるテーブルを選択する必要があります。データベースへのエクスポート ウィザードのこの画面には、選択したデータベースに含まれるテーブルやビューが一覧表示されます。

図 3-27
データベースへのエクスポート ウィザード、テーブルまたはビューの選択



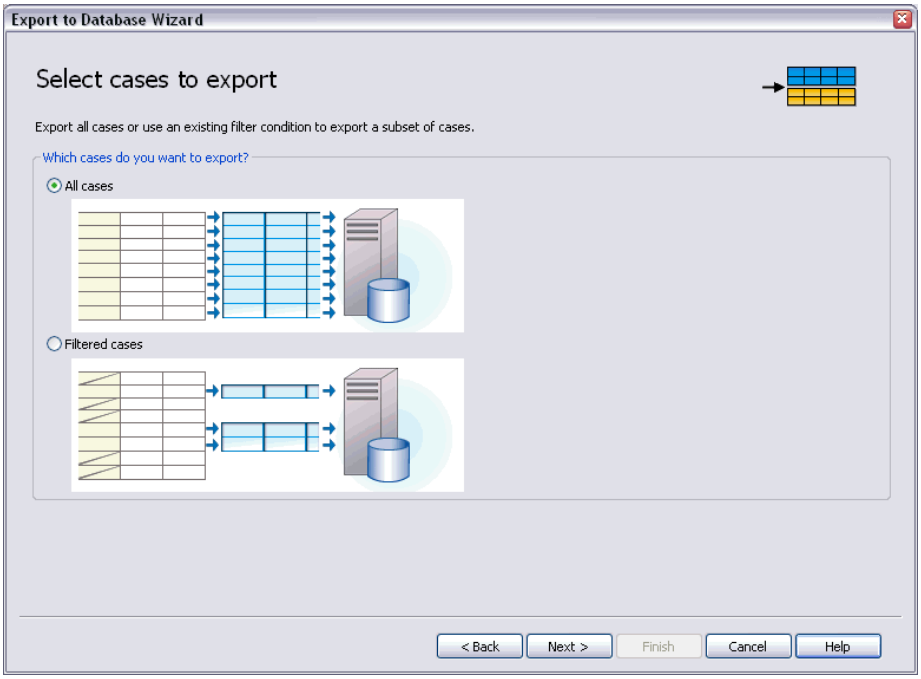
デフォルトでは、標準のデータベース テーブルだけが表示されます。リストに表示する項目の種類は次の方法で指定できます。

- **テーブル。**標準のデータベース テーブルです。
- **ビュー。**ビューは、クエリーで定義した仮想「テーブル」またはダイナミック「テーブル」です。ビューには、複数のテーブルを結合して含めたり、他のフィールド値を基に計算した結果から得られた複数のフィールドを結合して含めたり、これらのテーブルとフィールドの両方を結合して含めたりすることができます。ビューでは、レコードの追加や既存フィールドの値の置換を行うことはできますが、ビューの構成方法によっては、修正できるフィールドが限定される場合があります。たとえば、派生フィールドの修正、ビューへのフィールドの追加、ビューの置き換えは行えません。
- **シノニム。**シノニムは、テーブルまたはビューのエイリアスです。通常、クエリーで定義します。
- SYSTEM TABLE。システム テーブルでは、データベースのプロパティを定義します。一部の標準データベース テーブルは、システム テーブルに分類される場合があります。これらのデータベース テーブルは、このオプションを選択した場合にだけ表示されます。多くの場合、実際のシステム テーブルへのアクセスはデータベース管理者に制限されます。

エクスポートするケースの選択

データベースへのエクスポート ウィザードでケースを選択する方法は、すべてを選択するか、事前に定義されたフィルタ条件を使用して選択するかのどちらかに限られます。ケースのフィルタリングがすべて無効な場合、この画面は表示されず、アクティブなデータセットに含まれるケースがすべてエクスポートされます。

図 3-28
データベースへのエクスポート ウィザード、エクスポートするケースの選択



ケースの選択に使用するフィルタ条件の定義の詳細は、204 ページのケースの選択を参照してください。

ケースとレコードとの結合

既存テーブルにフィールド (列) を追加したり、既存フィールドの値を置換したりする場合は、アクティブ データセット内の各ケース (行) を、データベース内の対応するレコードと正確に関連付ける必要があります。

- データベースではほとんどの場合、各レコードを一意に識別するためのフィールドまたはフィールドの組み合わせが **主キー**として指定されます。
- 主キー フィールドなど、各レコードを一意に識別するためのフィールドに対応するのはどの変数なのかを特定する必要があります。
- こうしたフィールドは、必ずしもデータベースの主キーである必要はありませんが、フィールドの値またはそれらの組み合わせが各ケースに対して一意である必要があります。

データベースで各レコードを一意に識別するフィールドと変数を結合するには

- 変数を、対応するデータベース フィールド上にドラッグ アンド ドロップします。

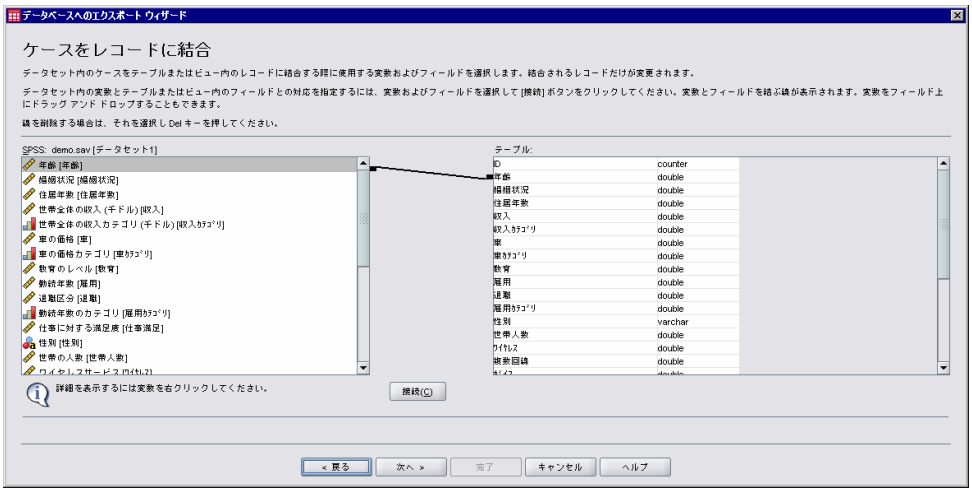
または

- 変数のリストから変数を選択し、データベース テーブルから対応するフィールドを選択して、[**接続**] をクリックします。

結合線を削除するには

- 結合線を選択し、Delete キーを押します。

図 3-29
データベースへのエクスポート ウィザード、ケースとレコードの結合



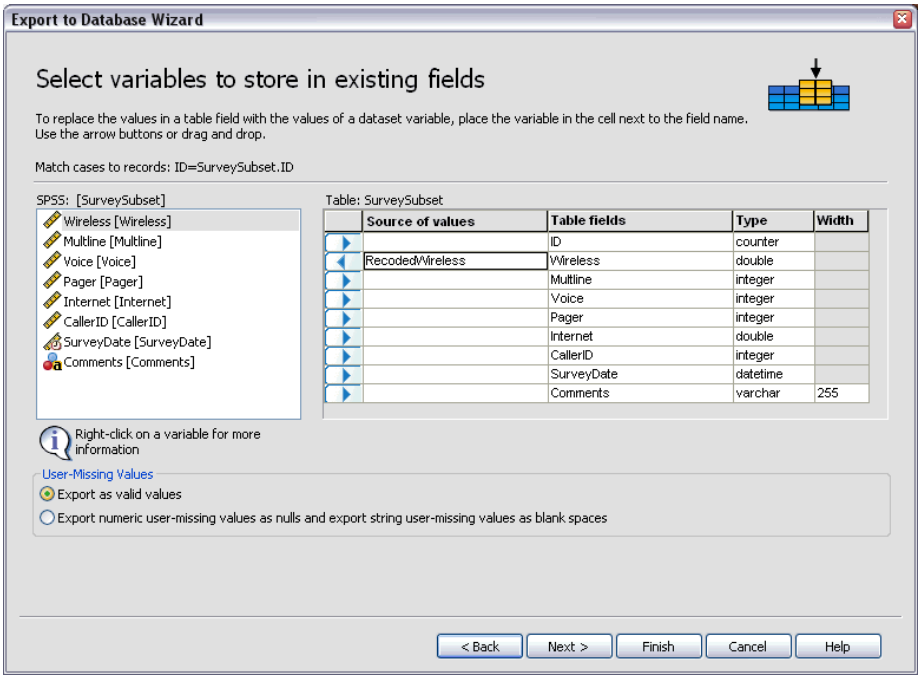
注:SPSS 変数名とデータベース フィールド名は必ずしも一致しませんが (データベース フィールド名には SPSS 変数名で使用できない文字が含まれることがあるため)、修正するデータベース テーブルからアクティブなデータセットが作成された場合は通常、変数名と変数ラベルのどちらか一方は、データベース フィールド名に類似したものとなります。

既存フィールドの値の置換

データベースの既存フィールドの値を置換するには

- データベースへのエクスポート ウィザードの [**データをエクスポートする方法**] 画面で、[**既存フィールドの値を置換**] を選択します。
- [**テーブルまたはビューの選択**] 画面で、データベース テーブルを選択します。
- [**ケースをレコードに結合**] 画面で、各ケースを一意に識別する変数を、対応するデータベース フィールド名に結合します。
- 値を置換するフィールドごとに、新しい値を含む変数を、対応するデータベース フィールド名の隣にある [**値のソース**] 列にドラッグ アンド ドロップします。

図 3-30
データベースへのエクスポート ウィザード、既存フィールドの値を置換



- 一般に、変数の基本的なデータ型 (文字型や数値型) は、データベース フィールドの基本的なデータ型に対応しています。データベースでは解決できないようなデータ型の不整合がある場合は、エラーが発生し、データベースに対するデータのエクスポートは行われません。たとえば、数値データ型 (倍精度整数型、実数型、整数型など) を持つデータベース フィールドに文字型変数をエクスポートした場合、その文字型変数のいずれかの値に非数値型文字が含まれていると、エラーが発生します。変数の隣のアイコンに表示される *a* の文字は、その変数が文字型変数であることを示しています。
- フィールドの名前、データ型、幅は修正できません。元のデータベース フィールドの属性は維持されます。置換されるのは値だけです。

新規フィールドの追加

既存のデータベース テーブルに新規フィールドを追加するには

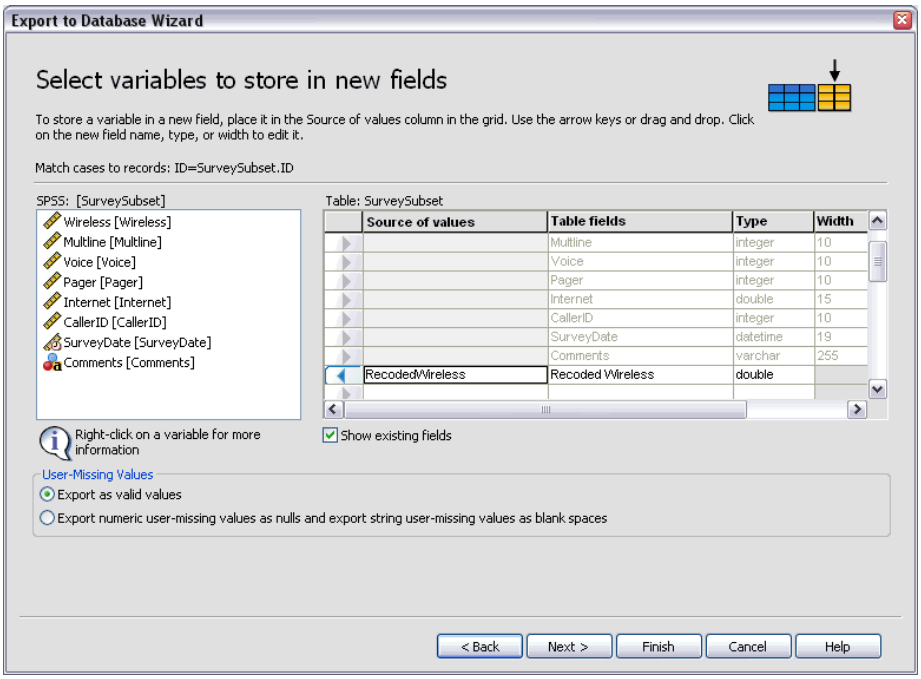
▶ データベースへのエクスポート ウィザードの [**データをエクスポートする方法**] 画面で、[**既存のテーブルに新規フィールドを追加**] を選択します。

▶ [**テーブルまたはビューの選択**] 画面で、データベース テーブルを選択します。

▶ [**ケースをレコードに結合**] 画面で、各ケースを一意に識別する変数を、対応するデータベース フィールド名に結合します。

► 新規フィールドとして追加する変数を [**値のソース**] 列にドラッグ アンド ドロップします。

図 3-31
データベースへのエクスポート ウィザード、既存テーブルに新規フィールドを追加



フィールド名およびデータ型の詳細は、55 ページのデータベースへのエクスポートに記載されている、SPSS 変数からデータベース フィールドを作成する方法についての説明を参照してください。

既存のフィールドを表示。このオプションを選択すると、既存フィールドのリストが表示されます。データベースへのエクスポート ウィザードのこの画面は、既存フィールドの置換には使用できませんが、テーブル内に存在するフィールドを把握する場合に便利です。既存フィールドの値を置換する場合は、62 ページの既存フィールドの値の置換を参照してください。

新規レコード (ケース) の追加

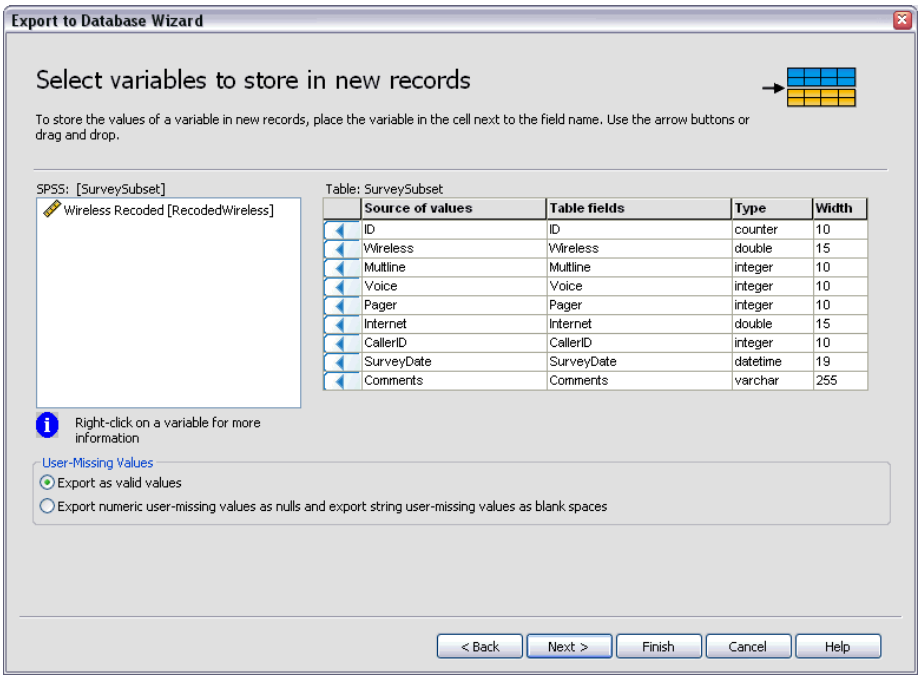
データベース テーブルに新規レコード (ケース) を追加するには

► データベースへのエクスポート ウィザードの [**データをエクスポートする方法**] 画面で、[**既存のテーブルに新規レコードを追加**] を選択します。

► [**テーブルまたはビューの選択**] 画面で、データベース テーブルを選択します。

▶ 変数を [値のソース] 列にドラッグ アンド ドロップして、アクティブなデータセット内の変数をテーブル フィールドに結合します。

図 3-32
データベースへのエクスポート ウィザード、テーブルへのレコード (ケース) の追加



データベースへのエクスポート ウィザードでは、アクティブなデータセットに格納されている元のデータベース テーブル (使用可能な場合) や、フィールド名と同じ変数名についての情報を基に、既存フィールドに結合するすべての変数が自動的に選択されます。自動的に結合を行うこの機能は、最初に大まかな処理を行っておくためのものであり、変数とそれを結合するデータベース フィールドとの対応関係は、自由に変更できます。

既存テーブルに新規レコードを追加する場合には、以下の基本的な規則/制限が適用されます。

- アクティブなデータセットのすべてのケース (または選択したすべてのケース) がテーブルに追加されます。これらのケースの中に、データベース内の既存のレコードと重複するものがある場合、重複するキー値が検出されるとエラーが発生します。選択したケースだけをエクスポートする方法の詳細は、60 ページのエクスポートするケースの選択を参照してください。

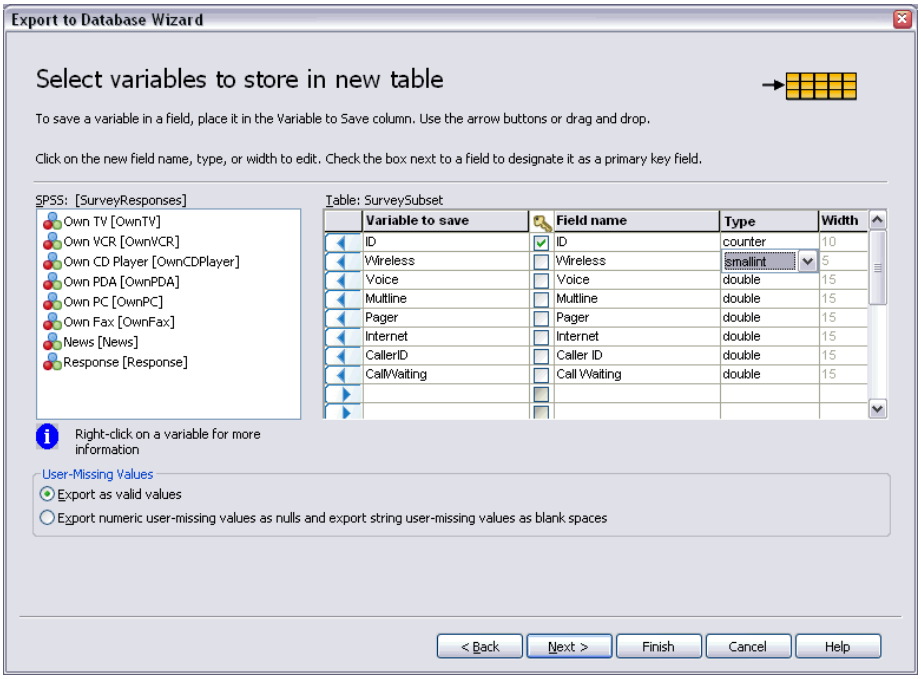
- 既存フィールドの値には、セッション内で新たに作成された変数の値を使用できますが、新規フィールドを追加したり、既存フィールドの名前を変更したりすることはできません。データベース テーブルに新規フィールドを追加する方法の詳細は、63 ページの新規フィールドの追加を参照してください。
- 除外されたデータベース フィールドや変数に結合されていないフィールドには、データベース テーブルに追加されたレコードに対する値は含まれません。([**値のソース**] セルが空の場合は、そのフィールドに結合されている変数はありません)。

新規テーブルの作成またはテーブルの置換

新規のデータベース フィールドを作成、または既存のデータベース テーブルを置換するには

- エクスポート ウィザードの [**データをエクスポートする方法**] 画面で、[**既存のテーブルを削除してそれと同じ名前の新規テーブルを作成**] または [**新規テーブルを作成**] を選択し、新規テーブルの名前を入力します。
- 既存テーブルを置換する場合は、[**テーブルまたはビューの選択**] 画面でデータベース テーブルを選択します。
- 変数を [**保存する変数**] 列にドラッグ アンド ドロップします。
- 必要であれば、主キーを定義する変数/フィールドの指定、フィールド名の変更、およびデータ型の変更を行えます。

図 3-33
データベースへのエクスポート ウィザード、新規テーブルの変数の選択



主キー。変数をデータベース テーブルの主キーとして指定するには、キー アイコンが表示されている列のボックスを選択します。

- 主キーの値はすべて一意であることが必要です。そうでない場合は、エラーが発生します。
- 主キーとして単独の変数を選択した場合は、すべてのレコード (ケース) で、その変数に一意の値が設定されている必要があります。
- 主キーとして複数の変数を選択した場合は、複合主キーが定義されます。この場合、各ケースに対しては、選択した変数の値の **組み合わせ**が一意であることが必要です。

フィールド名およびデータ型の詳細は、55 ページのデータベースへのエクスポートに記載されている、SPSS 変数からデータベース フィールドを作成する方法についての説明を参照してください。

データベースへのエクスポート ウィザードの終了

データベースへのエクスポート ウィザードの最後の画面には、エクスポートされるデータおよびそのエクスポート方法がまとめて表示されます。またこの画面では、データをエクスポートするのか、基本のコマンド シンタックスをシンタックス ウィンドウに貼り付けるのかを選択することもできます。

図 3-34
データベースへのエクスポート ウィザード: 完了画面

要約情報

- **データセット。**データのエクスポートに使用するデータセットの SPSS セッション名。この情報は主に、複数のデータ ソースを開いている場合に使用します。グラフィカル ユーザー インターフェイス (データベース ウィザードなど) を使って開いたデータ ソースには、DataSet1、DataSet2、... という形の名前が自動的に割り当てられていきます。また、コマンド シンタックスを使って開いたデータ ソースには、データセット名が (明示的に指定された場合に限って) 割り当てられます。
- **テーブル。**修正または作成されるテーブルの名前。
- **エクスポートするケース。**すべてのケースがエクスポートされるか、または事前に定義されたフィルタ条件に基づいて選択されたケースだけがエクスポートされます。 詳細は、60 ページのエクスポートするケースの選択を参照してください。
- **アクション。**データベースを修正する方法 (新規テーブルの作成、既存テーブルへのフィールドまたはレコードの追加など) が表示されます。
- **ユーザー欠損値。**ユーザー欠損値は、有効な値としてエクスポートできるほか、数値型変数に対してはシステム欠損値と同様に処理し、文字型変数に対しては空白文字に変換することができます。このオプションは、エクスポートする変数を選択した画面で設定されます。

Dimensions へのエクスポート

[Dimensions にエクスポート] ダイアログ ボックスでは、SPSS データ ファイルおよび Dimesions メタデータ ファイルを作成することができます。これらのファイルは、mrInterview および mrTables などの Dimensions アプリケーションにデータを読み込む際に使用できます。この機能は、SPSS と Dimensions アプリケーションとの間でデータを「やり取り」する場合に便利です。たとえば、元のメタデータ属性をすべて維持したまま、mrInterview データ ソースを SPSS に読み込み、新しい変数を計算し、さらにはそのデータを mrTables で読み込める形式で保存することができます。

データを Dimensions アプリケーション用にエクスポートするには

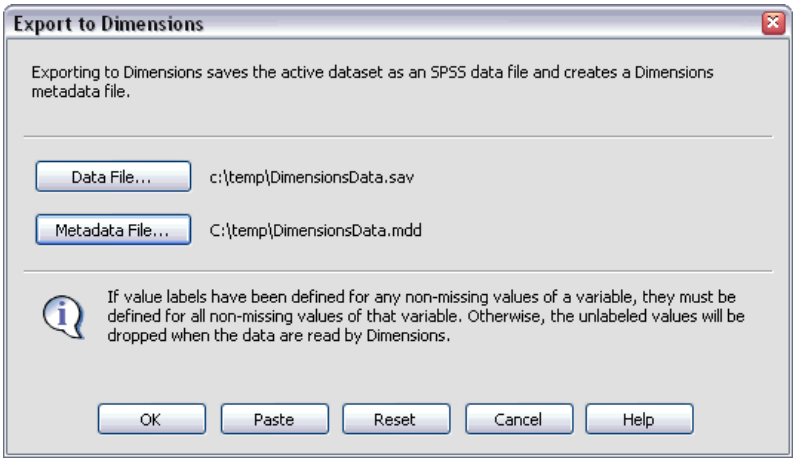
- エクスポートするデータが表示されている [データ エディタ] ウィンドウのメニューから、次の操作を行います。

  ファイル
  Dimensions にエクスポート

- **[データ ファイル]** をクリックし、SPSS データ ファイルの名前および場所を指定します。

- **[メタデータ ファイル]** をクリックし、Dimensions メタデータ ファイルの名前および場所を指定します。

図 3-35
[Dimensions にエクスポート] ダイアログ ボックス



Dimensions データ ソースから作成されなかった新規変数およびデータセットに対する SPSS 変数属性は、Dimensions Development Library の SPSS SAV DSC ドキュメントに記載されている方法に従って、メタデータ ファイル内の Dimensions メタデータ属性にマッピングされます。

アクティブなデータセットが Dimensions データ ソースから作成されている場合は、次のような処理が行われます。

- 元のメタデータ属性を、新規変数のメタデータ属性、およびメタデータ属性に影響を与える可能性のある、元の変数への変更処理（値ラベルの追加や修正など）と結合することにより、新規のメタデータ ファイルが作成されます。
- 元の変数が Dimensions データ ソースから読み込まれた場合、SPSS で認識されないメタデータ属性は、元の状態のまま維持されます。たとえば、グリッド変数は SPSS により通常の SPSS 変数に変換されますが、新規のメタデータ ファイルが保存される際も、これらのグリッド変数を定義するメタデータはそのまま維持されます。
- Dimensions 変数の名前が、SPSS 変数の命名規則に合わせて自動的に変更された場合、メタデータ ファイルでは、変換後の名前が元の Dimensions 変数名にマッピングされます。

値ラベルの有無により、変数のメタデータ属性、ひいては Dimensions アプリケーションによるこれらの変数の読み込み方法が変わります。ある変数において、欠損値以外の値に値ラベルが定義されている場合は、その変数の欠損値以外のすべての値に対し値ラベルを定義するようにしてください。ラベルのない値は、データ ファイルが Dimensions に読み込まれる際、対象から除外されます。

元のデータの保護

元のデータが誤って変更または削除されることを防ぐために、ファイルを読み取り専用にできます。

▶ データ エディタ メニューから次の項目を選択します。

ファイル
ファイルを読み取り専用にマーク

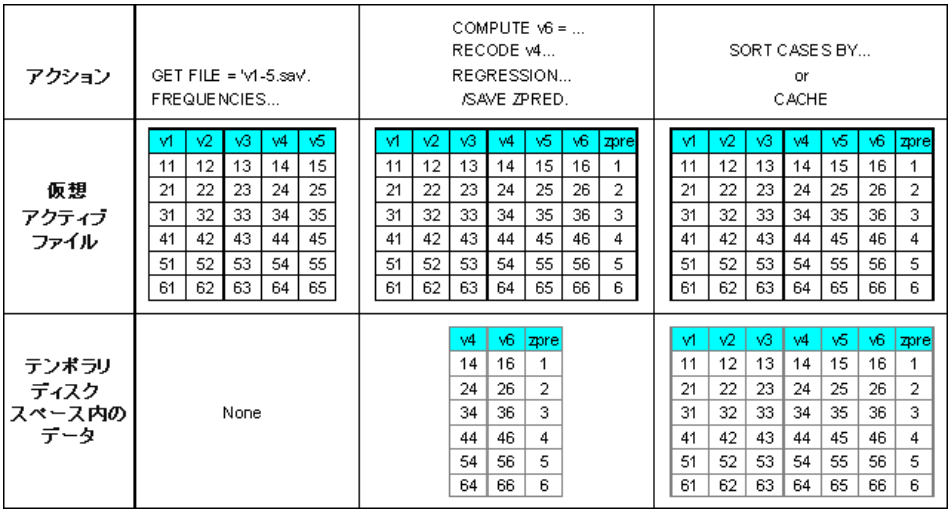
これ以降、データを編集してデータ ファイルを保存しようとすると、別のファイル名でしかデータを保存できなくなります。これにより、元のデータが保護されます。

ファイル許可を「読み取り/書き込み」に戻すには、[ファイル] メニューから **[ファイルを読み取りおよび書き込み用にマーク]** を選択します。

仮想アクティブ ファイル

仮想アクティブ ファイルを使用すると、大きなデータ ファイルを操作する場合に、同程度（またはそれ以上）のテンポラリ ディスク スペースが不要になります。ほとんどの分析および図表作成手続きでは、別の手続きを実行するごとに、元データのソースを再度読み込みます。データを修正する手続きは、変更をトレースするために、ある程度のテンポラリ ディスク スペースを必要とします。データ ファイルの全体をコピーするだけの十分なディスク スペースを常に必要とする処理もあります。

図 3-36
テンポラリ ディスク スペースの必要条件



v1	v2	v3	v4	v5
11	12	13	14	15
21	22	23	24	25
31	32	33	34	35
41	42	43	44	45
51	52	53	54	55
61	62	63	64	65

v1	v2	v3	v4	v5	v6	zpre
11	12	13	14	15	16	1
21	22	23	24	25	26	2
31	32	33	34	35	36	3
41	42	43	44	45	46	4
51	52	53	54	55	56	5
61	62	63	64	65	66	6

v1	v2	v3	v4	v5	v6	zpre
11	12	13	14	15	16	1
21	22	23	24	25	26	2
31	32	33	34	35	36	3
41	42	43	44	45	46	4
51	52	53	54	55	56	5
61	62	63	64	65	66	6

v4	v6	zpre
14	16	1
24	26	2
34	36	3
44	46	4
54	56	5
64	66	6

v1	v2	v3	v4	v5	v6	zpre
11	12	13	14	15	16	1
21	22	23	24	25	26	2
31	32	33	34	35	36	3
41	42	43	44	45	46	4
51	52	53	54	55	56	5
61	62	63	64	65	66	6

テンポラリ ディスク スペースを必要としない処理には、次のものがあります。

- SPSS データ ファイルの読み込み
- 2 つ以上の SPSS データ ファイルの結合
- データベース ウィザードによるデータベース テーブルの読み込み
- SPSS データ ファイルとデータベース テーブルとの結合
- データを読み込む手続きの実行 (たとえば、[度数分布表]、[クロス集計表]、[探索的分析])

テンポラリ ディスク スペース内にデータの列を 1 つ以上作成する処理には、次のものがあります。

- 新しい変数の計算
- 既存の変数の再割り当て
- 変数を作成または修正する手続きの実行 (たとえば、[線型回帰] の予測値の保存)

テンポラリ ディスク スペース内にデータ ファイル全体のコピーを作成する処理には、次のものがあります。

- Excel ファイルの読み込み
- データを並べ替える手続きの実行 (例、[ケースの並び替え]、[ファイルの分割])
- GET TRANSLATE または DATA LIST コマンドによるデータの読み込み

- ［データをキャッシュ］機能または CACHE コマンドの使用
- データ ファイルを読み込むような、他のアプリケーションの SPSS からの起動 (たとえば、AnswerTree、DecisionTime)

注:GET DATA コマンドには、DATA LIST に相当する機能がありますが、テンポラリ ディスク スペース内にデータ ファイルの全体のコピーを作成しません。コマンド シンタックスの SPLIT FILE コマンドは、データ ファイルの並べ替えを行いません。したがって、データ ファイルのコピーは作成されません。ただし、このコマンドが正常に動作するためには並べ替えを済ませたデータが必要です。そのため、この手続き用のダイアログ ボックス インターフェイスがデータ ファイルの並べ替えを自動的に行うので、データ ファイルの完全なコピーが作成されます。(コマンド シンタックスは、Student 版では利用できません)。

デフォルトにより、データ ファイル全体のコピーを作成するアクションには次のものがあります。

- データベース ウィザードによるデータベースの読み込み
- テキスト ウィザードによるテキスト ファイルの読み込み

テキスト ウィザードには、自動的にデータをキャッシュするオプション設定があります。デフォルトで、このオプションが選択されます。**[ローカルにデータをキャッシュ]** をオフにすると、このオプションの選択が解除されます。データベース ウィザードでは、生成されたコマンド シンタックスを貼り付けて、CACHE コマンドを削除することができます。

データ キャッシュの作成

仮想アクティブ ファイルを利用すると、必要なテンポラリ ディスク スペースの容量が激減しますが、「アクティブな」ファイルの一時的なコピーがないということは、手続きごとに元データのソースを読み直す必要があるということでもあります。外部ソースから読み込まれる大規模なデータ ファイルの場合、そのデータの一時的なコピーを作成するとパフォーマンスが向上します。たとえば、データベース ソースから読み込んだデータ テーブルの場合、データの読み込みを必要とするコマンドまたは手続きで、データベースから情報を読み込む SQL クエリーを再実行する必要があります。実質的にすべての統計分析手続きおよび図表作成手続きがデータを読み込む必要があるため、手続きを実行するごとに SQL クエリーを実行し直すことになります。

分析を実行するコンピュータ (ローカル コンピュータまたはリモート サーバーのどちらか) 上に十分なディスク スペースがあれば、アクティブ ファイルのデータ キャッシュを作成できるため、SQL クエリーを何回も実行しなくて済み、処理時間が短縮されます。データ キャッシュとは、データ全体の一時的なコピーです。

注: データベース ウィザードは、デフォルトで自動的にデータ キャッシュを作成します。ただし、データベースの読み込みに、コマンド シンタックスで GET DATA コマンドを使用している場合は、データ キャッシュが自動作成されません。(コマンド シンタックスは、Student 版では利用できません)。

データ キャッシュを作成するには

▶ メニューから次の項目を選択します。

ファイル
データをキャッシュ

▶ [OK] または [**今すぐキャッシュ**] をクリックします。

[OK] をクリックすると、プログラムが次回データを読み込むとき (たとえば、次回統計手続きを実行するとき) にデータ キャッシュを作成します。通常は、追加のデータ パスを必要としないので、[OK] をクリックします。[**今すぐキャッシュ**] をクリックすると、即座にデータ キャッシュが作成されます。ほとんどの状況では、即座にデータ キャッシュを作成する必要はないと思われます。[**今すぐキャッシュ**] は、主に次の 2 つの理由で使用されます。

- データ ソースが「ロック」されていて、自分のセッションを終了するか、別のデータ ソースを開くか、またはデータをキャッシュするまで、別の人が更新操作を実行できない。
- データ ソースが大きい場合、データをキャッシュすると、データ エディタ内の [データ ビュー] タブのスクロール速度が非常に速くなる。

データを自動的にキャッシュするには

SET コマンドを使用すると、アクティブ データ ファイルで変更回数が指定されている回数に達した後にデータ キャッシュを自動的に作成できます。デフォルトでは、アクティブ データ ファイルが 20 回変更された後、アクティブ データ ファイルが自動的にキャッシュされます。

▶ メニューから次の項目を選択します。

ファイル
新規作成
シンタックス

▶ シンタックス ウィンドウに、「SET CACHE n」と入力します (ここで、*n* は、データ ファイルをキャッシュする前にアクティブなデータ ファイルに加えられる変更の数を表します)。

▶ シンタックス ウィンドウのメニューから、次の項目を選択します。

実行
すべて

注: キャッシュ設定は異なるセッション間で有効にはなりません。新しいセッションを開始するたびに、値はデフォルトの 20 回にリセットされます。

章

4

ディストリビュート アナリシス モード

ディストリビュート アナリシス モードにすれば、メモリーを大量に使用する作業をローカル (デスクトップ) コンピュータ以外のコンピュータで行うことができます。通常、ディストリビュート アナリシスに使用されるリモート サーバーはローカル コンピュータに比べて大容量のメモリーを備え、高速な計算が可能なので、ディストリビュート アナリシス モードを活用すると、コンピュータの処理時間を大幅に短縮できます。リモート サーバーによるディストリビュート アナリシスは、次の場合に力を発揮します。

- 大規模なデータ ファイル群、特にデータベース ソースから読み込まれるデータを伴う作業。
- メモリーを大量に消費する作業。ローカル アナリシス モードでは時間がかかる作業には、ディストリビュート アナリシスが適しています。

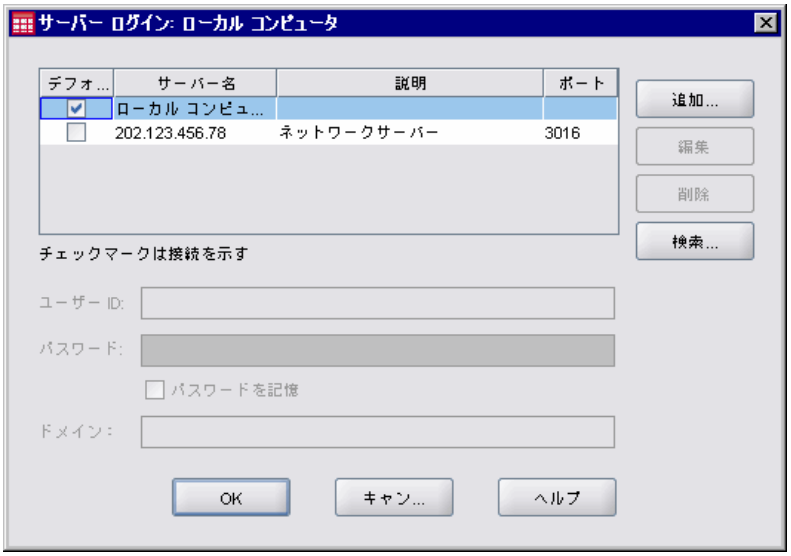
ディストリビュート アナリシスの影響を受けるのは、データの読み込み、データの変換、新しい変数の計算、統計量の計算といったデータ関係の作業に限られます。ディストリビュート アナリシスは、ピボット テーブルの操作や図表の修正などの出力の編集に関係する作業には有効ではありません。

注: ディストリビュート アナリシスを利用できるのは、ローカル バージョンの SPSS が動作しており、かつ、リモート サーバー上にインストールされたサーバー バージョンの SPSS ソフトウェアにアクセス可能な場合に限られます。さらに、ディストリビュート アナリシスが Windows デスクトップ コンピュータでのみ利用可能です。Mac および Linux では利用できません。

サーバー ログイン

[サーバー ログイン] ダイアログ ボックスでは、コマンドを処理して手続きを実行するコンピュータを選択できます。ローカル コンピュータまたはリモート サーバーのどちらかを選択することができます。

図 4-1
[サーバー ログイン] ダイアログ ボックス



リモート サーバーを、リストに追加、変更、またはリストから削除できます。リモート サーバーでは通常、ユーザー ID とパスワードを入力する必要があります。また、ドメイン名が必要な場合もあります。使用可能なサーバー、ユーザー ID、パスワード、ドメイン名、およびその他の接続情報については、システム管理者に問い合せてください。

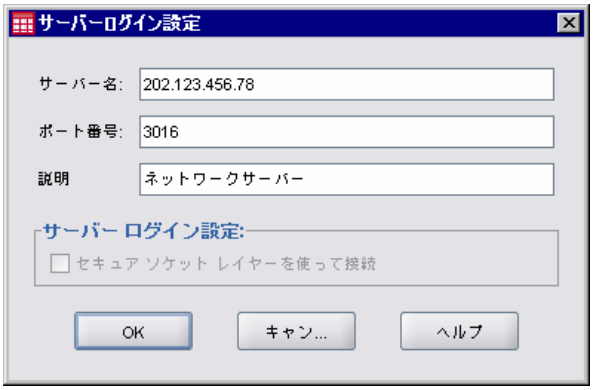
デフォルト サーバーを選択し、どのサーバーとであっても、関連するユーザー ID、ドメイン名、およびパスワードを保存できます。新しいセッションを開始すると、自動的にデフォルト サーバーに接続されます。

SPSS Adaptor for Predictive Enterprise Services を使用するライセンスがあり、SPSS Predictive Enterprise Services 3.5 以降を実行しているなら、ネットワーク上で利用可能なサーバーのリストを表示するには [**検索...**] をクリックします。Predictive Enterprise Repository にログしていないなら、サーバーのリストを表示する前に接続情報を入力する必要があります。

サーバー ログイン設定の追加および編集

ディストリビュート アナリシス モードで使用するリモート サーバーの接続情報を追加または編集するには、[サーバー ログイン設定] ダイアログ ボックスを使用します。

図 4-2
[サーバー ログイン設定] ダイアログ ボックス



使用可能なサーバー、サーバーのポート番号、および追加接続情報については、システム管理者に問い合わせてください。管理者から指示がない限り、セキュア ソケット レイヤーを使用しないでください。

サーバー名。サーバーの「名前」は、コンピュータに割り当てられた英数字名 (たとえば、NetworkServer) またはコンピュータに割り当てられた一意的な IP アドレス (たとえば、202.123.456.78) です。

ポート番号。ポート番号は、サーバー ソフトウェアが通信用に使用するポートです。

説明。サーバー リストに表示するオプションの説明を入力することができます。

セキュア ソケット レイヤーを使って接続。セキュア ソケット レイヤー (SSL) を使用すると、ディストリビュート アナリシス要求がリモート SPSS サーバーに送信されるときに暗号化されます。SSL を使用する場合は、事前に管理者に確認してください。このオプションを使用するためには、自分のデスクトップ コンピュータとサーバーの両方で SSL が構成されている必要があります。

サーバーの選択、切り替え、または追加を行うには

▶ メニューから次の項目を選択します。

ファイル
サーバーの切り替え...

デフォルトのサーバーを選択するには

▶ サーバーのリストで、使用したいサーバーの隣にあるチェック ボックスをオンにします。

▶ 管理者から提供されるユーザー ID、ドメイン名、およびパスワードを入力します。

注: 新しいセッションを開始すると、自動的にデフォルト サーバーに接続されます。

別のサーバーに切り替えるには

▶ リストからサーバーを選択します。

▶ ユーザー ID、ドメイン名、およびパスワード (必要な場合) を入力します。

注: セッション中にサーバーを切り替えると、開いているすべてのウィンドウが閉じます。ウィンドウを閉じる前に、変更を保存するよう求められます。

サーバーを追加するには

▶ 管理者から、サーバー接続情報を取得します。

▶ [**追加**] をクリックして、[サーバー ログイン設定] ダイアログ ボックスを開きます。

▶ 接続情報とオプション設定を入力し、[OK] をクリックします。

サーバーを編集するには

▶ 管理者から、変更された接続情報を取得します。

▶ [**編集**] をクリックして、[サーバー ログイン設定] ダイアログ ボックスを開きます。

▶ 変更内容を入力して、[OK] をクリックします。

利用可能なサーバーを検索するには

注: 利用可能なサーバーの検索機能は、SPSS Adaptor for Predictive Enterprise Services のライセンスがあり、SPSS Predictive Enterprise Services 3.5 以降を実行しているときにのみ利用できます。

▶ [**検索...**] をクリックして、[サーバー検索] ダイアログ ボックスを開きます。Predictive Enterprise Repository にログしていないなら、接続情報を入力する必要があります。

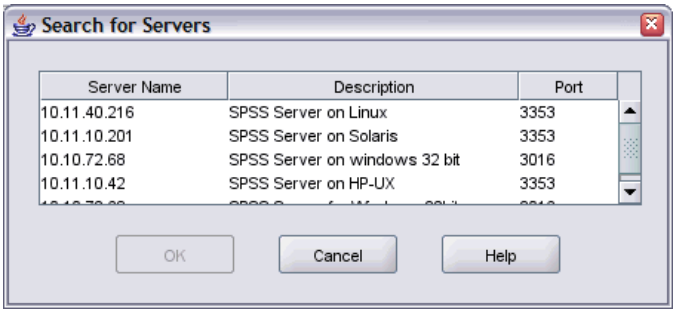
▶ 1 つ以上の利用可能なサーバーを選択して、[OK] をクリックします。[サーバー ログイン] ダイアログ ボックスに、サーバーが表示されません。

▶ サーバーの 1 個に接続するには、「他のサーバーに切り替えるには」の指示に従ってください。

利用可能なサーバーの検索

ネットワークで可能な 1 個以上のサーバーを選択するには [サーバーの検索] ダイアログ ボックスを使用します。[サーバー ログイン] ダイアログ ボックスの [**検索...**] をクリックすると、ダイアログ ボックスが表示されます。

図 4-3
[サーバー検索] ダイアログ ボックス



[サーバー ログイン] ダイアログ ボックスに追加するには、1 個以上のサーバーを選択し、[OK] をクリックします。[サーバー ログイン] ダイアログ ボックスに手動でサーバーを追加すると、利用可能なサーバーを検索しながら適切なサーバー名とポート番号を要求せずにサーバーに接続できます。この情報は、自動的に提供されます。しかし、ユーザー名、ドメイン、そしてパスワードのような適切なログイン情報が必要です。

リモート サーバーからデータ ファイルを開く

ディストリビュート アナリシス モードでは、標準の [ファイルを開く] ダイアログ ボックスの代わりに [リモート ファイルを開く] ダイアログ ボックスが表示されます。

- リモート サーバー上で、またはリモート サーバーからどのようなファイル、フォルダ、およびドライブが使用できるかにより、表示される一覧の内容が異なります。ダイアログ ボックスの一番上に現在のサーバー名が表示されます。
- ディストリビュート アナリシス モードでは、ドライブを共有ドライブとして指定していない場合、またはデータ ファイルが格納されているフォルダを共有フォルダとして指定していない場合は、ローカル コンピュータ上のデータ ファイルにはアクセスできません。ローカル コンピュータ上のフォルダをサーバー ネットワークと「共有」する方法については、オペレーティング システムのドキュメントを参照してください。
- サーバーが別のオペレーティング システムを稼動している場合は (たとえば、ローカルのコンピュータで Windows を稼動していて、サーバーでは UNIX が稼動している場合)、ローカル データ ファイルが共有フォルダにあったとしても、おそらく、ディストリビュート アナリシス モードでローカル データ ファイルにはアクセスできません。

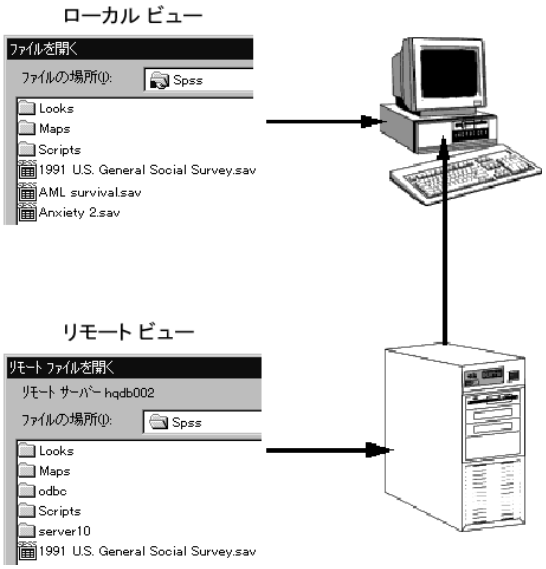
ローカル アナリシス モードとディストリビュート アナリシス モードでのファイル アクセス

ローカル コンピュータとネットワーク上のデータ ファイル、フォルダ (ディレクトリ)、およびドライブの表示は、コマンドの処理および手続きの実行に現在使用しているコンピュータの観点からの表示になります。このコンピュータは、あなたの目の前にあるコンピュータである必要はありません。

ローカル アナリシス モード。ローカル コンピュータを自分の「サーバー」として使用する場合、(データ ファイルを開くための) ファイル アクセス用のダイアログ ボックスに現れるデータ ファイル、フォルダ、およびドライブの表示は他のアプリケーションや Windows エクスプローラの表示に似ています。自分のコンピュータ内のすべてのデータ ファイルおよびフォルダと、マウントされているネットワーク ドライブのファイルおよびフォルダを表示できます。

ディストリビュート アナリシス モード。別のコンピュータを「リモート サーバー」として使用し、コマンドと手続きを実行する場合、データ ファイル、フォルダ、およびドライブの表示は、リモート サーバー コンピュータからの表示になります。馴染みのあるのフォルダ名 (Program Files など) や、ドライブ (C など) が表示される場合がありますが、自分のコンピュータ上のフォルダやドライブ**ではありません。**リモート サーバー上のフォルダとドライブです。

図 4-4
ローカル ビューとリモート ビュー



ディストリビュート アナリシス モードでは、ドライブを共有ドライブとして指定しているか、またはデータ ファイルが格納されているフォルダを共有フォルダとして指定していない限り、ローカル コンピュータ上のデータ ファイルにはアクセ

スできません。サーバーが別のオペレーティング システムを稼動している場合は (たとえば、ローカルのコンピュータで Windows を稼動していて、サーバーでは UNIX が稼動している場合)、ローカル データ ファイルが共有フォルダにあったとしても、おそらく、ディストリビュート アナリシス モードでローカル データ ファイルにはアクセスできません。

ディストリビュート アナリシス モードとは、ネットワーク上の他のコンピュータに置かれているデータ ファイルにアクセスするということと同じではありません。ローカル アナリシス モードでもディストリビュート アナリシス モードでも、他のネットワーク装置にあるデータ ファイルにアクセスできます。ローカル モードでは、ローカル コンピュータから他の装置にアクセスします。ディストリビュート モードでは、リモート サーバーから他のネットワーク装置にアクセスします。

現在のモードが、ローカル アナリシス モードなのかディストリビュート アナリシス モードなのか不明な場合は、データ ファイル アクセス用のダイアログ ボックスのタイトル バーを参照してください。ダイアログ ボックスのタイトルに「**リモート**」という文字が含まれている場合 ([**リモート ファイルを開く**] など)、または [**リモート サーバー**: [**サーバー名**]] というテキストがダイアログ ボックスの最上段に表示されている場合は、ディストリビュート アナリシス モードを使用しています。

注: これは、データ ファイルにアクセスするためのダイアログ ボックス (たとえば、[データを開く]、[データを保存]、[データベースを開く]、[データ ディクショナリーの適用]) だけに当てはまることです。他のファイルの種類 (たとえば、ビューア ファイル、シンタックス ファイル、およびスクリプト ファイル) についてはすべて、ローカル ビューが常に使用されます。

ディストリビュート アナリシス モードでの手続きの使用

ディストリビュート アナリシス モードでは、ローカル バージョンとリモート サーバー上のリモート バージョンの両方にインストールされている手続きのみが使用できます。

ローカル コンピュータにはインストールされているが、リモート サーバーでは利用できないオプションのコンポーネントがある場合、ローカル コンピュータからリモート サーバーに切り替えると、関係する手続きがメニューから削除され、対応するコマンド シンタックスでエラーが発生します。ローカル モードに戻すと、関係する手続きがすべて戻ります。

絶対パスと相対パス指定

ディストリビュート アナリシス モードでは、データ ファイルの相対パス指定およびコマンド シンタックス ファイルは現在のサーバーに関係するものであり、ローカル コンピュータに関係しません。相対パス指定 (/mydocs/mydata.sav など) は、ローカル ドライブのディレクトリおよびファイルを指しているのではなく、リモート サーバーのハードディスク上のディレクトリおよびファイルを指しているということです。

Windows UNC パス指定

Windows サーバー バージョンを使用している場合、コマンド シンタックスでデータとシンタックス ファイルにアクセスするときに、UNC (universal naming convention) 指定を使用できます。UNC 指定の一般的書式は次のとおりです。

```
\\servername\sharename\path\filename
```

- *servername* は、データ ファイルが格納されているコンピュータの名前です。
- *sharename* は、共有フォルダとして指定されているそのコンピュータ上のフォルダ (ディレクトリ) です。
- *path* は、共有フォルダの下の追加フォルダ (サブディレクトリ) パスです。
- *filename* は、データ ファイルの名前です。

一例を以下に挙げます。

```
GET FILE='\\hqdev001\public\july\sales.sav'.
```

コンピュータに名前が割り当てられていない場合には、次のように、IP アドレスを使用できます。

```
GET FILE='\\204.125.125.53\public\july\sales.sav'.
```

UNC パス指定を使用していても、共有として指定されている装置およびフォルダのデータおよびシンタックス ファイルにしかアクセスできません。ディストリビュート アナリシス モードを使用している場合には、これに、ローカル コンピュータ上のデータおよびシンタックス ファイルが加わります。

UNIX 絶対パス指定

UNIX サーバー バージョンでは、UNC パスに相当するものがないため、相対パスは使用できません。すべてのディレクトリ パスは、サーバーのルートから始まる絶対パスで指定する必要があります。たとえば、データ ファイルが /bin/spss/data にあり、現在のディレクトリも /bin/spss/data である場合、`GET FILE='sales.sav'` は無効です。次のようにパス全体を指定する必要があります。

```
GET FILE='/bin/spss/sales.sav'.
INSERT FILE='/bin/spss/salesjob.sps'.
```

章
5

データ エディタ

データ エディタは、データ ファイルを作成、編集するためのスプレッドシートに似た便利な機能を提供します。データ エディタ ウィンドウは、セッションを開始すると自動的に開きます。

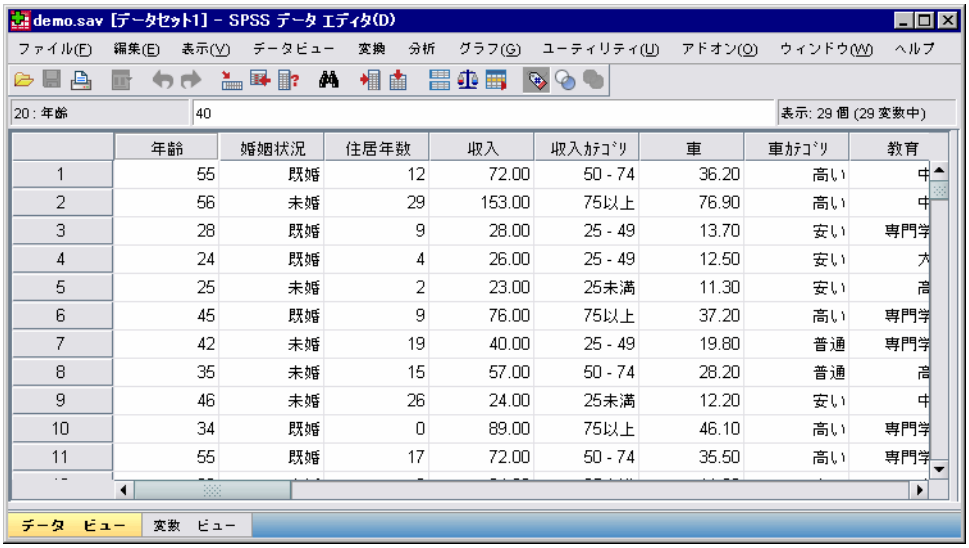
データ エディタは、次の 2 種類のデータ ビューを提供します。

- **データ ビュー。**実際のデータ値または定義済みの値ラベルを表示します。
- **変数ビュー。**定義済みの変数ラベルおよび値ラベル、データの型 (たとえば、文字型、日付型、または数値型)、尺度 (名義、順序、スケール)、およびユーザー指定の欠損値などの、変数定義情報を表示します。

どちらのビューでも、データ ファイルに含まれる情報を追加、変更、および削除できます。

データ ビュー

図 5-1
データ ビュー



データ ビューの機能の多くは、スプレッドシート アプリケーションの機能に似ています。しかし、大きく異なる点もいくつかあります。

- 行はケースです。各行は、ケースまたは観測値を表します。たとえば、アンケートの各回答者は、1 つのケースになります。
- 列は変数です。各列は、測定している変数または特性を表します。たとえば、アンケートの各項目は 1 つの変数になります。
- セルには値が入っています。各セルには、ケースが取る変数の値が 1 つ入っています。セルは、ケースと変数の交点です。セルにはデータ値しか入りません。スプレッドシート プログラムと違い、データ エディタのセルには数式を書き込むことはできません。
- データ ファイルは矩形行列です。データ ファイルの大きさは、ケースと変数の数によって決まります。データは、どのセルにも入力できます。定義されたデータ ファイルの領域外のセルにデータを入力すると、そのセルを含む行または列を取り入れるためにデータ矩形が拡張されます。データ ファイルの枠線内には「空の」セルはありません。数値型変数では、空白セルはシステム欠損値に変換されます。文字型変数では、空白は有効値とみなされます。

変数ビュー

図 5-2
変数ビュー

変数ビューには、データ ファイル内の各変数の属性が記述されます。変数ビューで

- 行は変数。
- 列は変数の属性。

変数の追加や削除ができます。また、次に示す変数の属性を変更できます。

- 変数名
- データの型
- 桁数や文字数
- 小数桁数
- 記述的な変数ラベルと値ラベル
- ユーザー指定の欠損値
- 列幅
- 尺度

これらの属性は、データ ファイルを保存するときにすべて保存されます。

変数プロパティの定義は、[変数ビュー] で行うだけでなく、次の 2 つの方法で定義することもできます。

- [データ プロパティのコピー] ウィザードでは、外部の SPSS データ ファイルまたは現在のセッションで使用可能な別のデータセットをテンプレートとし、アクティブなデータセット内でファイル プロパティと変数プロパティを定義することができます。またアクティブなデータセットにある変数は、同じデータセットにある別の変数のテンプレートとしても利用できます。[データ プロパティのコピー] は、データ エディタ ウィンドウの [データ] メニューで利用できます。
- [変数プロパティの定義] (これもデータ エディタ ウィンドウの [データ] メニューで使用可能) では、データをスキャンして選択した変数の一意的なデータ値をすべて表示し、ラベルのない値を示します。また自動ラベル機能を提供しています。この方法は 0 = **男性**、1 = **女性**など、カテゴリを表す数値コードを使用したカテゴリ変数で特に有効です。

変数の属性を表示または定義するには

▶ データ エディタ ウィンドウをアクティブにします。

▶ [データ ビュー] の列の最上段にある変数名をダブルクリックするか、または [**変数ビュー**] タブをクリックします。

▶ 新しい変数を定義するには、空白行に変数名を入力します。

▶ 定義または修正する属性を選択します。

変数名

変数名には、次の規則が適用されます。

- 各変数名は一意でなければなりません。
- 変数名の長さは最大 64 バイト分で、先頭文字には @、#、$ のいずれかの文字を使用する必要があります。それ以降には、文字、数字、非句読文字、およびピリオド (.) を使用できます。コード ページ モードでは、64 バイトは、1 バイト言語 (たとえば、英語、フランス語、ドイツ語、スペイン語、イタリア語、ヘブライ語、ロシア語、ギリシャ語、アラビア語、タイ語) では 64 文字に相当し、2 バイト言語 (たとえば、日本語、中国語、韓国語) では 32 文字に相当します。コード ページ モードで 1 バイトしか使用しない文字列の多くは、Unicode モードでは 2 バイト以上使用します。たとえば、é はコード ページ形式では 1 バイトですが、Unicode 形式では 2 バイトです。そのため、r ésum é はコード ページ ファイルでは 6 バイト、Unicode モードでは 8 バイトです。

  注: 上記の「文字」には、プラットフォームの文字セットでサポートされている言語において、通常の表記語の中で使用される非句読文字がすべて含まれます。
- 変数名に空白文字を使用することはできません。
- 変数名の先頭にある # 文字は、スクラッチ変数であることを表します。スクラッチ変数は、コマンド シンタックスによってのみ作成できます。変数を新規に作成するためのダイアログ ボックスでは、変数の先頭の文字に # を指定することはできません。
- 変数名の先頭にある $ 文字は、その変数がシステム変数であることを表します。ユーザー定義の変数の先頭文字に $ 記号は使用できません。
- ピリオド、アンダースコア、$、#、および @ は、変数名に使用できます。たとえば、**A._$@#1** は有効な変数名です。
- ただし、ピリオドはコマンド終了符として解釈されることがあるため、変数名の末尾には使用しないでください。変数名がピリオドで終わる変数は、コマンド シンタックスによってのみ作成できます。変数を新規作成するためのダイアログ ボックスでは、変数名がピリオドで終わる変数を作成することはできません。
- また、アンダースコアも変数名の末尾には使用しないようにしてください。変数名の末尾にアンダースコアを使用すると、コマンドや手続きにより自動的に作成された変数の名前と重複する場合があります。
- 予約キーワードを変数名として使用することはできません。予約キーワードには、ALL、AND、BY、EQ、GE、GT、LE、LT、NE、NOT、OR、TO、WITH があります。
- 変数名は、大文字と小文字の組み合わせを使用できます。大文字と小文字は、表示目的で保持されます。
- 長い変数名を出力する際、複数行に渡って表示するために途中で改行する必要がある場合、アンダースコア、ピリオド、および小文字から大文字に変化する場所で行が分割されます。

変数尺度

尺度をスケール (間隔または比率尺度の数値データ)、順序、または名義として指定できます。名義データおよび順序データには、文字型か数値型のいずれかを指定します。

- **名義データ。**値がランキングなどを持たないカテゴリを表しているとき、名義 (変数) として取り扱うことができます。たとえば、従業員の会社の所属などです。名義変数の例としては、地域やジップ コードや所属宗教などがあります。
- **順序データ。**値がランキングをもったカテゴリを表しているとき、変数を順序として取り扱うことができます。たとえば、「かなり不満」から「かなり満足」までのようなサービス満足度のレベルなどです。順序変数の例としては、満足度や信頼度を表す得点や嗜好得点などです。
- **スケール データ。**値が有意な基準を持った順序カテゴリを表しているとき、変数を尺度として扱うことができます。値間の距離の比較などに適切です。スケール変数の例としては、年齢や、千ドル単位で表した所得があります。

注: 文字型順序変数の場合は、文字型の値が文字順 (五十音順) に並んだものが、カテゴリの正しい順序だとみなされます。たとえば、値が **低**、**中**、**高**である文字型変数では、カテゴリの順序は **高**、**低**、**中**と解釈されますが、これは正しい順序ではありません。一般に、順序データを表す場合には、数値型コードを使用した方が信頼性が高いといえます。

セッションに測定尺度水準が割り当てらている間に、新規数値型変数が作成されました。外部ファイル形式から読み込まれるデータ、およびバージョン 8.0 より前で作成された SPSS データ ファイルに関しては、デフォルトの尺度は次の規則に基づいて割り当てられます。

- 一意の値の数 24 より少ない数値型変数、および文字列方変数は、名義データに設定されます。
- 一意の値の数が 24 以上ある数値型変数は、スケール データに設定されます。

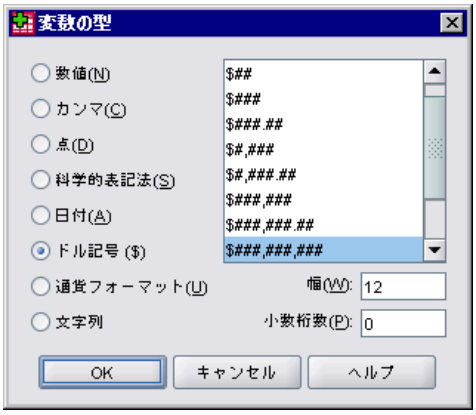
数値型変数に関するスケール / 名義データの分割値は、[オプション] ダイアログ ボックスで変更することができます。 詳細は、501 ページの 45 章の[データ] オプションを参照してください。

[データ] メニューから [変数プロパティの定義] ダイアロ グボックスを開いて、正しい尺度を割り当てることができます。 詳細は、117 ページの 7 章の尺度の割り当てを参照してください。

変数の型

[変数の型] では、各変数に対するデータの型を指定します。デフォルトで、すべての新しい変数は数値とみなされます。[変数の型] を使用して、データの型を変更できます。[変数の型] ダイアログ ボックスの内容は、選択したデータの型によって決まります。幅と小数桁数を入力するためのテキスト ボックスがあるデータの型と、スクロールしてリストから書式を選択するデータの型があります。

図 5-3
[変数の型] ダイアログ ボックス



利用できるデータの型としては、次のものがあります。

数値型。値が数値である変数。値は、標準の数値書式で表示されます。データ エディタでは、標準書式、または科学的表記法の数値が有効です。

カンマ。3 桁ごとにカンマで区切り、小数点にピリオドを使用した値が表示される数値変数。データ エディタは、カンマ付きまたはカンマなしのカンマ変数の数値、または科学的表記法の数値を受け付けます。値では、小数点の右側にカンマを含めることはできません。

ドット。3 桁ごとにピリオドで区切り、小数点にカンマを使用した値が表示される数値変数。データ エディタは、ピリオド付きまたはピリオドなしのドット変数の数値や科学的表記法の数値を受け付けます。値では、小数点の右側にピリオドを含めることはできません。

科学的表記法。E の後ろに、10 のべき乗指数が符号付きで示されて値が表示される数値変数。データ エディタは、指数付き、または指数なしの変数の数値が有効です。指数の前には、E または D (符号つきまたは符号なし)、あるいは符号のみが付きます (たとえば、123、1.23E2、1.23D2、1.23E+2、1.23+2)。

日付。カレンダ日付形式や時計時刻形式の 1 つで表示される数値型変数。リストから書式を選択します。日付は、スラッシュ、ハイフン、ピリオド、または空白スペースを区切り文字として入力できます。2 桁表示の年で表される 100 年間の範囲は、[オプション] ([編集] メニューの **[オプション]** を選択し、**[データ]** タブをクリック) で設定されます。

ドル記号。ドル記号 ($) で始まり、3 桁ごとにカンマで区切り、小数点にピリオドを使用して表示される数値変数。入力するデータ値の前に、ドル記号を付けても付けなくてもかまいません。

通貨書式。[オプション] ダイアログ ボックスの [通貨] タブで定義したユーザー指定の通貨書式で値が表示される数値変数。定義されたユーザー指定の通貨文字はデータ入力時には使用できませんが、データ エディタには表示されます。

文字列。値が数値ではなく、したがって計算には使用されない変数。値の長さは、定義された文字数までです。大文字と小文字は異なるものとみなされます。英数字変数とも呼ばれます。

変数の型を定義するには

- 定義する変数の [**型**] セル内のボタンをクリックします。
- [変数の型] ダイアログ ボックスでデータの型を選択します。
- [OK] をクリックします。

入力書式と表示書式

書式によっては、[データ ビュー] 内の値の表示が、入力されて内部的に保存されている実際の値と異なることもあります。一般的なガイドラインを次に示します。

- 数値、カンマ、およびドット書式では、小数点以下 16 桁までの値を入力でき、値全体が内部的に保存されます。[データ ビュー] では、定義された小数桁数のみを表示し、それ以上桁数のある値を丸めます。しかし、計算では常に完全な値が使用されます。
- 文字型変数では、値はすべて最大幅の左端に寄せられます。最大幅が 3 の文字列変数では、**No** の値は `'No '` として内部的に保存され、`' No'` と同じではありません。
- 日付型書式では、スラッシュ、ダッシュ、スペース、カンマ、またはピリオドを日、月、および年の値の間の区切り記号として使うことができ、さらに月の値として数字、3 文字の略語、省略しない月名を入力することもできます。一般書式 *dd-mmm-yy* の日付は、区切り記号としてダッシュを使い、月は 3 文字の略語で表示されます。一般書式 *dd/mm/yy* および *mm/dd/yy* の日付は、区切り記号としてスラッシュを使い、月は数字で表示されます。内部的に、ほとんどの日付書式変数は 1582 年 10 月 14 日からの秒数として保管されます。[オプション] で、2 桁の年数を取る西暦範囲を設定できます ([編集] メニューの [**オプション**] をクリックして、[**データ**] タブをクリックします)。
- 時間型書式では、時間、分、および秒間の区切り記号としてコロン、ピリオド、またはスペースを使うことができます。時刻は、区切り記号としてコロンを使って表示されます。内部的には、時刻は時間間隔を表す秒数として保存されています。たとえば、10:00:00 は、内部では 36000 として保存されます (60 秒 x 60 分 x 10 時間)。

変数ラベル

記述的な変数ラベルは 256 文字 (2 バイト言語では 128 文字) まで割り当てることができます。変数ラベルには、変数名に許されていない空白や予約語を含めることができます。

変数ラベルを指定するには

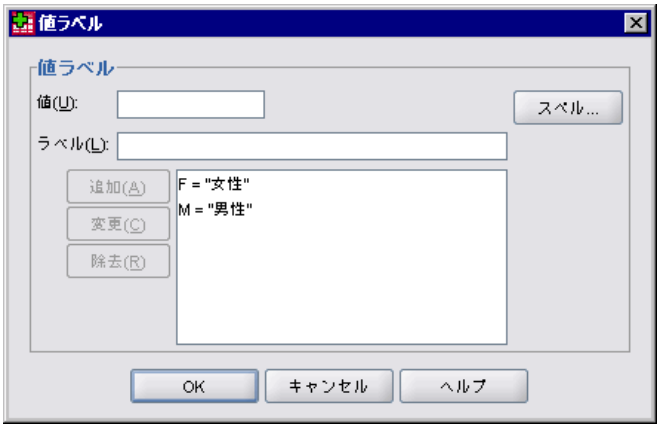
- ▶ データ エディタ ウィンドウをアクティブにします。
- ▶ [データ ビュー] の列の最上段にある変数名をダブルクリックするか、または [**変数ビュー**] タブをクリックします。
- ▶ 変数の [**ラベル**] セルに、記述的な変数ラベルを入力します。

値ラベル

変数の各値に記述的な値ラベルを割り当てることができます。このプロセスは、データ ファイルで数値コードを使用して、数値以外のカテゴリを表しているような場合に、特に有効です (たとえば、**男性**と **女性**を表すためのコードとして 1 と 2 を使っている場合)。

- 値ラベルの最大長さは、120 バイトです。

図 5-4
[**値ラベル**] ダイアログ ボックス



値ラベルを指定するには

- ▶ 定義する変数の [**値**] セル内のボタンをクリックします。
- ▶ それぞれの値に、値と値ラベルを入力します。
- ▶ 値ラベルを入力するには、[**追加**] をクリックします。
- ▶ [OK] をクリックします。

ラベルへの改行の挿入

変数ラベルと値ラベルは、ピボット テーブルおよび図表においてセルまたは領域が狭くて 1 行にラベル全体を表示できない場合に自動的に複数の行に折り返されます。ただし、ラベルを他の位置で折り返したい場合は、手動で改行を挿入することができます。さらに、**常に**特定の場所で折り返して複数の行に表示されるような変数ラベルおよび値ラベルを作成することもできます。

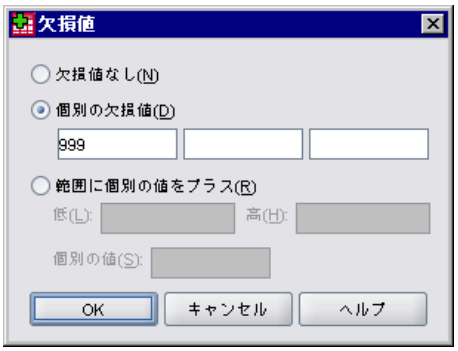
- 変数ラベルの場合は、データ エディタの [変数ビュー] で目的の変数の [ラベル] セルを選択します。
- 値ラベルの場合は、データ エディタの [変数ビュー] で目的の変数の [値] セルを選択し、セル内のボタンをクリックします。次に [値ラベル] ダイアログ ボックスで、変更するラベルを選択します。
- ラベルを折り返す位置に「¥n」と入力します。

  「¥n」は改行文字として解釈され、ピボット テーブルや図表には表示されません。

欠損値 (変数の定義)

欠損値は、指定したデータ値を **ユーザー欠損値**と定義します。たとえば、回答者が答えることを拒否したことによるデータ欠損と、質問が回答者に適用されなかったことによるデータ欠損を区別したい場合などです。ユーザー欠損値として指定されたデータ値は、特別に扱われるために区別され、ほとんどの計算から除外されます。

図 5-5
[欠損値] ダイアログ ボックス



- 最大で 3 つの個別の欠損値、一定範囲の欠損値、または範囲に個別の値 (1 つ) をプラスしたものを入力できます。
- 範囲は、数値変数だけに指定できます。
- NULL 値または空白値などのすべての文字列値は、欠損値として明示的に定義されなければ、有効な値とみなされます。

- [文字型変数の欠損値] は 8 バイトを超えることができません。(文字型変数の定義された列幅に制限はありませんが、ユーザー定義の欠損値は 8 バイトを超えることはできません。)
- 文字型変数に対して NULL 値または空白値を欠損値として定義するには、[**個別の欠損値**] にあるボックスの 1 つにスペースを 1 つ入力します。

欠損値を定義するには

- ▶ 定義する変数の [**欠損値**] セル内のボタンをクリックします。
- ▶ 欠損値に指定する値、または値の範囲を入力します。

列幅

列幅を、文字数として指定できます。列幅は、[データ ビュー] で列の境界線をクリックし、ドラッグして変更することもできます。

列書式は、データ エディタの値の表示にだけ影響します。列幅を変更しても変数に定義された幅は変化しません。

変数の配置

配置は、[データ ビュー] におけるデータ値または値ラベルの表示を制御します。デフォルトの配置は、数値変数については右、文字型変数については左です。この設定は、[データ ビュー] での表示にしか影響しません。

変数定義属性を複数の変数に適用する

ある変数に変数定義属性を定義した後、その属性をコピーして 1 つ以上の変数に適用できます。

通常のコピーと貼り付け操作により、変数定義属性を他の変数に適用できます。次のようにできます。

- 属性の 1 つ (たとえば、値ラベル) をコピーして、1 つ以上の変数の同じ属性セルに貼り付けます。
- ある変数のすべての属性をコピーした後、それを 1 つ以上の他の変数に貼り付けます。
- コピーした変数のすべての属性を持つ、複数の新しい変数を作成します。

変数定義属性を他の変数に適用する

定義済みの変数の個々の属性を適用するには

- ▶ [変数ビュー] で、他の変数に適用する属性セルを選択します。

▶ メニューから次の項目を選択します。

編集
コピー

▶ 属性を適用する属性セルを選択します。(対象となる変数は複数選択できます。)

▶ メニューから次の項目を選択します。

編集
貼り付け

属性を空白行に貼り付けた場合、選択されている属性以外のすべての属性をデフォルト属性とした新しい変数が作成されます。

定義済みの変数のすべての属性を適用するには

▶ [変数ビュー] で、適用したい属性を持つ変数の行番号を選択します。(その行全体が強調表示されます。)

▶ メニューから次の項目を選択します。

編集
コピー

▶ 属性を適用する変数の行番号を選択します。(対象となる変数は複数選択できます。)

▶ メニューから次の項目を選択します。

編集
貼り付け

同じ属性を持つ新しい変数を複数作成する

▶ [変数ビュー] で、新しい変数に使用したい属性を持つ変数の行番号をクリックします。(その行全体が強調表示されます。)

▶ メニューから次の項目を選択します。

編集
コピー

▶ データファイル内で最後に定義された変数の下にある、空白行の行番号をクリックします。

▶ メニューから次の項目を選択します。

編集
変数の貼り付け...

▶ [変数の貼り付け] ダイアログ ボックスで、作成する変数の数を入力します。

▶ 新しい変数の接頭辞と開始番号を入力します。

▶ [OK] をクリックします。

新しい変数名は、指定した接頭辞および指定した数字で始まる連続した番号によって構成されます。

ユーザー指定の変数属性

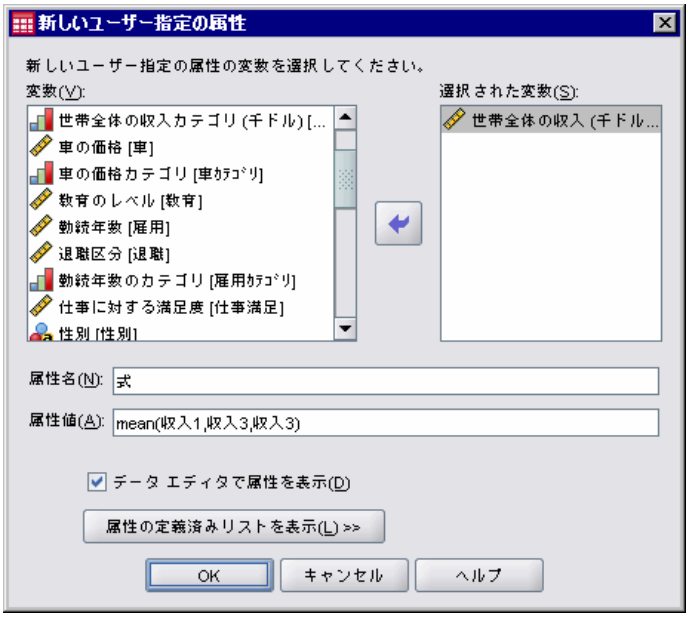
標準の変数属性（たとえば、値ラベル、欠損値、尺度）に加えて、ユーザー指定の変数属性を作成することができます。標準の変数属性と同様に、ユーザー指定の属性は SPSS 形式のデータ ファイルとともに保存されます。したがって、調査の質問（例えば、単一選択、複数選択、空欄記入）の回答の型を表す変数属性、または計算済み変数に使用された式を作成できます。

ユーザー指定の変数属性の作成

新しいユーザー指定の属性を作成するには

- ▶ [変数ビュー] で、メニューから次の項目を選択します。
  データ
  新しいユーザー指定の属性...
- ▶ 新しい属性を割り当てる変数を [選択された変数] のリストにドラック アンド ドロップします。
- ▶ 属性名を入力します。属性名は、変数名と同じ命名規則に従っている必要があります。 詳細は、85 ページの変数名を参照してください。
- ▶ 属性の選択値を入力します。複数の変数を選択している場合、入力した値は選択されたすべての変数に割り当てられます。何も入力せず空白にしておくこともできます。その場合、[変数ビュー] で変数ごとに値を入力できます。

図 5-6
[新しいユーザー指定の属性] ダイアログ ボックス



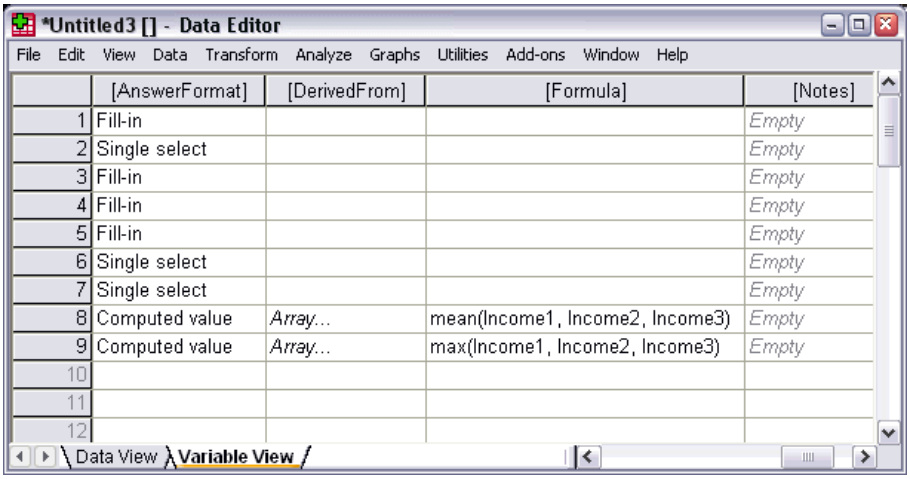
データ エディタで属性を表示。データ エディタの [変数ビュー] で、属性を表示します。ユーザー指定の属性の表示方法については、下記の ユーザー指定の変数属性の表示と編集を参照してください。

属性の定義済みリストを表示。データセットに対してすでに定義されているユーザー指定の属性のリストを表示します。ドル記号 ($) で始まる属性名は、予約属性であり変更できません。

ユーザー指定の変数属性の表示と編集

ユーザー指定の変数属性は、データ エディタの [変数ビュー] で表示および編集することができます。

図 5-7
[変数ビュー] に表示されたユーザー指定の変数属性



- ユーザー指定の変数属性の名前は、角かっこで囲まれて表示されます。
- ドル記号で始まる属性名は、予約済みのため変更できません。
- 空白のセルは、対象となる変数の属性が存在していないことを意味しています。セル内に **空**と表示されている場合は、対象となる変数の属性は存在していますが、属性値が割り当てられていません。セル内にテキストを入力すると、対象となる変数の属性が、入力した値で存在するようになります。
- セル内に **配列**...と表示されている場合、属性配列であることを意味しています。この属性には複数の値が含まれます。値のリストを表示するには、セル内のボタンをクリックします。

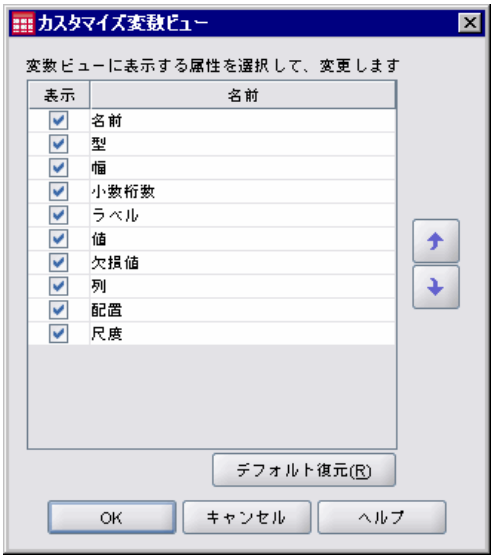
ユーザー指定の変数属性を表示および編集するには

▶ [変数ビュー] で、メニューから次の項目を選択します。

表示
変数ビューをカスタマイズ...

▶ 表示するユーザー指定の変数属性を選択 (チェック) します。(ユーザー指定の変数属性は、角かっこで囲まれて表示されます。)

図 5-8
変数ビューをカスタマイズ

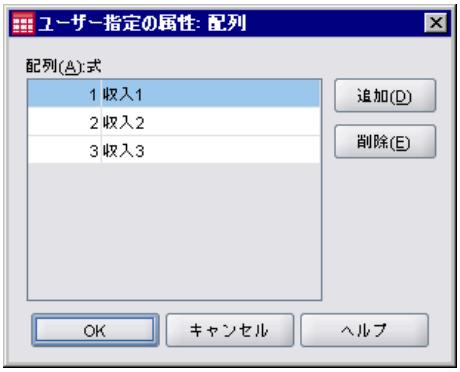


一度属性が [変数ビュー] で表示されると、データ エディタで直接編集することができます。

変数属性配列

[変数ビュー]、または [変数プロパティの定義] の [ユーザー指定の変数属性] ダイアログ ボックスでは、ユーザー指定の変数属性が複数の値を含む **属性配列**である場合、対応するセルに「**配列...**」というテキストが表示されます。たとえば、ある派生変数の計算に使用するすべてのソース変数を識別する属性配列を定義できます。値のリストを表示したり編集したりするには、セル内のボタンをクリックします。

図 5-9
[ユーザー指定の属性: 配列] ダイアログ ボックス



変数ビューをカスタマイズ

[変数ビュー] (例えば、名前、型、ラベル) に表示された属性および表示された順序を制御するには [変数ビューをカスタマイズ] を使用できます。

- (データセットに関連するユーザー指定の変数属性は、角かっこで囲まれて表示されます)。詳細は、94 ページのユーザー指定の変数属性の作成を参照してください。
- 表示設定のカスタマイズは、SPSS 形式のデータ ファイルとともに保存されます。
- [変数ビュー] の属性のデフォルト表示および順序を制御もできます。詳細は、503 ページの 45 章のデフォルト変数の変更を参照してください。

変数ビューをカスタマイズするには

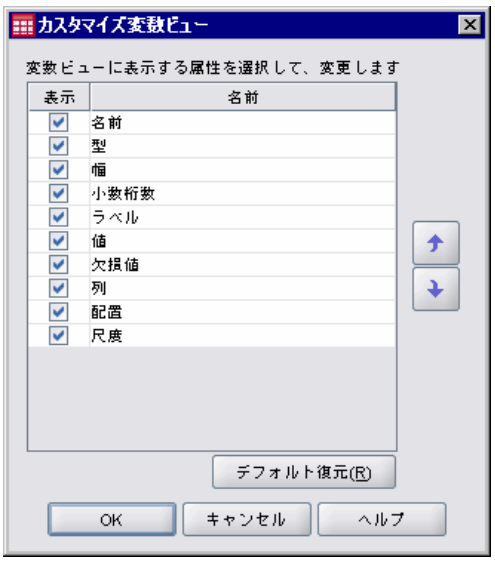
▶ [変数ビュー] で、メニューから次の項目を選択します。

表示
変数ビューをカスタマイズ...

▶ 表示する変数属性を選択 (チェック) します。

▶ 属性の表示順を変更するには上および下矢印を使用します。

図 5-10
[変数ビュー] ダイアログ ボックスをカスタマイズ...



復元デフォルト。デフォルトの表示および順序設定を適用します。

スペル チェック変数と値ラベル

変数ラベルと値ラベルのスペリング チェックをするには

▶ [変数ビュー] で、コンテキスト メニューの [**ラベル**] または [**値**] 列上を右クリックし、次の項目を選択します。
スペリング

または

▶ [変数ビュー] で、メニューから次の項目を選択します。
ユーティリティ
スペリング

または

▶ [値ラベル] ダイアログ ボックスの [**スペリング**] をクリックします。(こうして特定の変数の値ラベルのスペル チェッキングを制限します。)

スペル チェッキングはデータ エディタの [変数ビュー] の変数ラベルおよび値ラベルに制限されます。

データの入力

[データ ビュー] で、データ エディタにデータを直接入力することができます。入力の順序は自由です。データは、選択した領域または個別のセルに対して、ケース別または変数別に入力できます。

- アクティブ セルが強調表示されます。
- 変数名およびアクティブ セルの行番号は、データ エディタの左上端に表示されます。
- セルを選択してデータ値を入力すると、データ エディタの上部にあるセル エディタに値が表示されます。
- Enter キーを押すか別のセルを選択するまで、データ値は記録されません。
- 単純な数値データ以外のものを入力するには、最初に変数の型を定義する必要があります。

空の列に値を入力すると、データ エディタにより新しい変数が自動的に作成され、変数名が割り当てられます。

数値データを入力するには

▶ [データ ビュー] でセルを選択します。

▶ データ値を入力します。(この値は、データ エディタの上部にあるセル エディタに表示されます)。

▶ Enter キーを押すか、別のセルを選択して値を記録します。

非数値データを入力するには

- [データ ビュー] の列の最上段にある変数名をダブルクリックするか、または [**変数ビュー**] タブをクリックします。
- 変数の [**型**] セル内のボタンをクリックします。
- [変数の型] ダイアログ ボックスでデータの型を選択します。
- [OK] をクリックします。
- 行番号をダブルクリックするか、または [**データ ビュー**] タブをクリックします。
- 新しく定義した変数の列にデータを入力します。

データ入力に値ラベルを使用するには

- 値ラベルが現在 [データ ビュー] に表示されていない場合は、メニューから次の項目を選択します。

  表示
  値ラベル

- 値を入力するセルをクリックします。
- ドロップダウン リストから値ラベルを選択します。

  値が入力され、値ラベルがセルに表示されます。

  注: このプロセスは、変数に定義済みの値ラベルがある場合にだけ使用できます。

データ エディタでのデータ値の制限

定義済みの変数の型および幅によって、[データ ビュー] でセルに入力できる値の型が決まります。

- 定義された変数型で許可されていない文字を入力すると、文字は入力できません。
- 文字型変数では、定義された幅を超える文字は許可されません。
- 数値型変数では、定義された幅を超える整数値を入力することはできますが、データ エディタは、科学的表記または省略符号 (...) の付いた値の一部を表示して、その値が定義された幅を超えていることを示します。セルに値を表示するには、変数の定義された幅を変更します。

  注: 列幅を変更しても変数の幅は影響を受けません。

データの編集

データ エディタを使うと、[データ ビュー] 内のデータ値を修正することができます。次のようにできます。

- データ値の変更
- データ値の切り取り、コピー、および貼り付け
- ケースの追加と削除
- 変数の追加と削除
- 変数の順序変更

データ値の置換または修正

古い値を削除して新しい値を入力するには、次のようにします。

▶ [データ ビュー] 内のセルをダブルクリックします。(セルの値は、セル エディタに表示されます。)

▶ 値を直接セルで編集するか、またはセル エディタで編集します。

▶ Enter キーを押すか、別のセルを選択して新規の値を記録します。

データ値の切り取り、コピー、および貼り付け

データ エディタでは、個別のセルの値または値のグループの切り取り、コピー、および貼り付けを行うことができます。次のようにできます。

- 1 つのセルの値を別のセルに移動またはコピーする。
- 1 つのセルの値をセルのグループに移動またはコピーする。
- 1 つのケース (行) の値を複数のケースに移動またはコピーする。
- 1 つ変数 (列) の値を複数の変数に移動またはコピーする。
- セルの値のグループを別のグループに移動またはコピーする。

データ ビューで貼り付けられた値のデータ変換

コピー元セルと目標セルの定義済みの変数の型が異なる場合は、データ エディタにより値が変換されます。変換ができない場合は、目標セルにシステム欠損値が挿入されます。

数値または日付を文字型に変換。数値 (たとえば、数値、ドル、ドット、またはカンマ) 書式 および日付書式は、文字型変数セルに貼り付けると文字型に変換されます。セルで表示されていた数値がそのまま文字列値になります。たとえばドルの場合、表示されるドル記号は文字列値の一部になります。定義された文字型変数幅を超える値は切り捨てられます。

文字型を数値または日付に変換。目標セルの数値または日付書式として通用する文字を含む文字型の値は、そのまま数値または日付値に変換されます。たとえば、25/12/91 という文字列は、目標セルの書式型が day-month-year 書式の 1 つ

の場合には有効な日付に変換されますが、目標セルの書式型が month-day-year 書式の 1 つの場合にはシステム欠損値に変換されます。

日付を数値に変換。日付と時刻値は、目標セルが数値書式 (たとえば、数値、ドル、ドット、またはカンマ) の 1 つの場合、秒数に変換されます。日付は 1582 年 10 月 14 日以後の秒数として内部的に保存されているので、日付を数値に変換すると非常に大きな数になります。たとえば、10/29/91 という日付は、12,908,073,600 に変換されます。

数値を日付または時間に変換。数値は、有効な日付または時間となりうる秒数を表している場合、日付または時間に変換されます。日付では、86,400 未満の数値はシステム欠損値に変換されます。

新しいケースを挿入する

空白行のセルにデータを入力すると、データ エディタで新しいケースが自動的に作成されます。そのケースの他の変数には、すべてシステム欠損値が挿入されます。新しいケースと既存のケース間に空白行があると、その空白行は新しいケースになり、すべての変数にシステム欠損値が割り当てられます。既存のケース間に新しいケースを挿入することもできます。

既存のケース間に新しいケースを挿入するには

- [データ ビュー] で、新しいケースを挿入する位置の下にあるケース (行) 内のセルを選択します。
- メニューから次の項目を選択します。

  データ
  ケースの挿入

  ケースに新しい行が挿入され、すべての変数にシステム欠損値が割り当てられます。

新しい変数の挿入

[データ ビュー] 内の空白の列にデータを入力するか、または [変数ビュー] 内の空白の行にデータを入力すると、デフォルトの変数名 (接頭辞 var と連続した数字) およびデフォルトのデータ書式型 (数値) を持つ新しい変数が自動的に作成されます。データ エディタにより、新しい変数のすべてのケースにシステム欠損値が挿入されます。新しい変数と既存変数の間で [データ ビュー] に空白列がある場合や、[変数ビュー] に空白行がある場合も、すべてのケースがシステム欠損値を持つ新しい変数になります。既存の変数間に新しい変数を挿入することもできます。

既存の変数間に新しい変数を挿入するには

- 新しい変数を挿入する位置の右 ([データ ビュー] の場合) または下 ([変数ビュー] の場合) にある変数のセルを選択します。
- メニューから次の項目を選択します。

  データ
  変数の挿入

  新しい変数が挿入されて、すべてのケースに欠損値が割り当てられます。

変数を移動するには

- 変数を選択するには、[データ ビュー] の変数名、または [変数ビュー] の変数の行番号をクリックします。
- 新しい場所に変数をドラッグ アンド ドロップします。
- 既存の 2 つの変数の間に変数を配置する場合、[データ ビュー] では、変数を配置する場所の右にある変数の列上にドロップします。[変数ビュー] では、変数を配置する場所の下にある変数の行にドロップします。

データの型を変更するには

[変数ビュー] で [変数の型] ダイアログ ボックスを使用すれば、いつでも変数のデータの型を変更することができます。データ エディタにより既存の値は新しい型に変換されます。変換できない場合は、システム欠損値が割り当てられます。変換規則は、書式型が異なる変数にデータ値を貼り付ける場合と同じです。データ書式を変更すると欠損値の指定または値ラベルが失われる場合は、データ エディタが警告ボックスを表示して、変更を行うかまたはキャンセルするかを確認します。

ケースまたは変数について表示する

[指定のケース/変数への移動] ダイアログ ボックスを使用すると、データ エディタで指定したケース (行) または変数名へ移動することができます。

- メニューから次の項目を選択します。

  編集
  指定のケースへの移動

  または

  編集
  変数へ移動

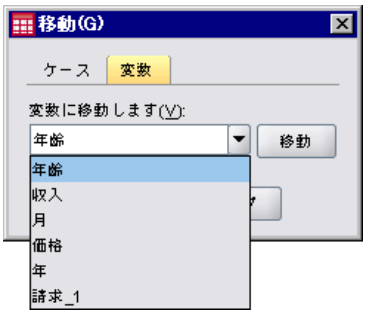
- ケースでは、[データ ビュー] の現在の行数を表す整数値を入力してください。

  注: 特定のケースの現在の行数は並び替えまたは他の処理で変更できます。

または

▶ 変数では、変数名を入力するか、ドロップダウン リストから変数を選択してください。

図 5-11
ダイアログ ボックスに移動



データおよび属性値の検索および置換

[データ ビュー] のデータ値または [変数ビュー] の属性値を検索および/または置換するには:

▶ 検索したい列のセルをクリックします。(値の検索および置換は単一列に制限されます。)

▶ メニューから次の項目を選択します。

編集
検索

または

編集
置換

データ ビュー

- [データ ビュー] の検索はできません。検索方向はいつも下です。
- 日付と時刻の場合、[データ ビュー] に表示された形式の値が検索されます。たとえば、10/28/2007 として表示された日付は 10-28-2007 の日付の検索では見つけることができません。
- その他の数値型変数の場合、**[含む]**、**[開始]**、および **[終了]** によりフォーマットされた値が検索されます。たとえば、**[開始]** オプションを使用する場合、ドル記号の変数に対する値 $123 の検索では $123.00 と $123.40 の両方が見つかりますが、$1,234 は見つかりません。**[すべてのセル]** オプションを使用する場合、検索する値はフォーマットされた値もフォーマットされていない値も使用できますが (単純な数値型 (F) 書式)、(データ エディタで表示される精度で) 正確に一致する数値のみが該当します。

- 値ラベルが選択変数列で表示されると、下線のデータ値ではなくラベル テキストが検索されラベル テキストは置き換えることはできません。

変数ビュー

- 検索は、[名前]、[ラベル]、[値]、[欠損値]、およびユーザー指定変数属性列でのみ可能です。
- 置換は、[ラベル]、[値]、およびユーザー指定属性列でのみ可能です。
- [値] (値ラベル) 列では、検索文字列データ値または値ラベルの両方と一致します。

  注: データ値を置換すると、その値に関連する前の値ラベルは削除されます。

データ エディタでのケースの選択状態

あるケースのグループが選択されていて、かつ選択されていないケースが削除されていない場合、選択されなかったケースには、データ エディタの行番号の上に斜めの線 (スラッシュ) が引かれます。

図 5-12
データ エディタでの分析から除外されたケース

データ エディタ表示オプション

[表示] メニューには、データ エディタで使用できる表示オプションがいくつかあります。

フォント。データ表示のフォントの指定をコントロールします。

格子線。格子線の表示、非表示を切り替えます。

値ラベル。実際のデータ値の表示と、ユーザーが定義した記述的な値ラベルの表示を切り替えます。このオプションは、[データ ビュー] でしか使用できません。

複数のビューの使用

[データ ビュー] で、水平スクロール バーの下と垂直スクロール バーの右にある区切り線を使用して、複数のビュー (ペイン) を作成できます。

また、[ウィンドウ] メニューを使用して、ペインの区切り線を挿入および削除できます。区切り線を挿入するには、次の手順を実行します。

▶ [データ ビュー] で、メニューから次の項目を選択します。

ウィンドウ
分割

選択されたセルの上と左に、区切り線が挿入されます。

- 最上部左のセルが選択されている場合は、現在のビューの縦と横をほぼ半分に分割する位置に区切り線が挿入されます。
- 最初の列の最上部以外のセルが選択されている場合、選択されたセルの上に水平ペイン区切り線が挿入されます。
- 最上行の最初のセル以外のセルが選択されている場合、選択されたセルの左側に垂直ペイン区切り線が挿入されます。

データ エディタの印刷

データ ファイルは、画面上に表示されるとおりに印刷されます。

- 現在表示されている [データ ビュー] または [変数ビュー] 内の情報が印刷されます。[データ ビュー] では、データが印刷されます。[変数ビュー] では、データ定義情報が印刷されます。
- 選択されたビュー内に現在格子線が表示されていれば、格子線が印刷されます。
- [データ ビュー] 内で値ラベルが表示されている状態であれば、値ラベルが印刷されます。それ以外の場合には、実際のデータ値が印刷されます。

データ エディタ ウィンドウの [表示] メニューを使うと、格子線を表示することも隠すこともでき、さらにデータ値と値ラベルの表示を切り替えることもできます。

データ エディタの内容を印刷するには

▶ データ エディタ ウィンドウをアクティブにします。

▶ 印刷するビューのタブをクリックします。

▶ メニューから次の項目を選択します。

ファイル
印刷...

章

6

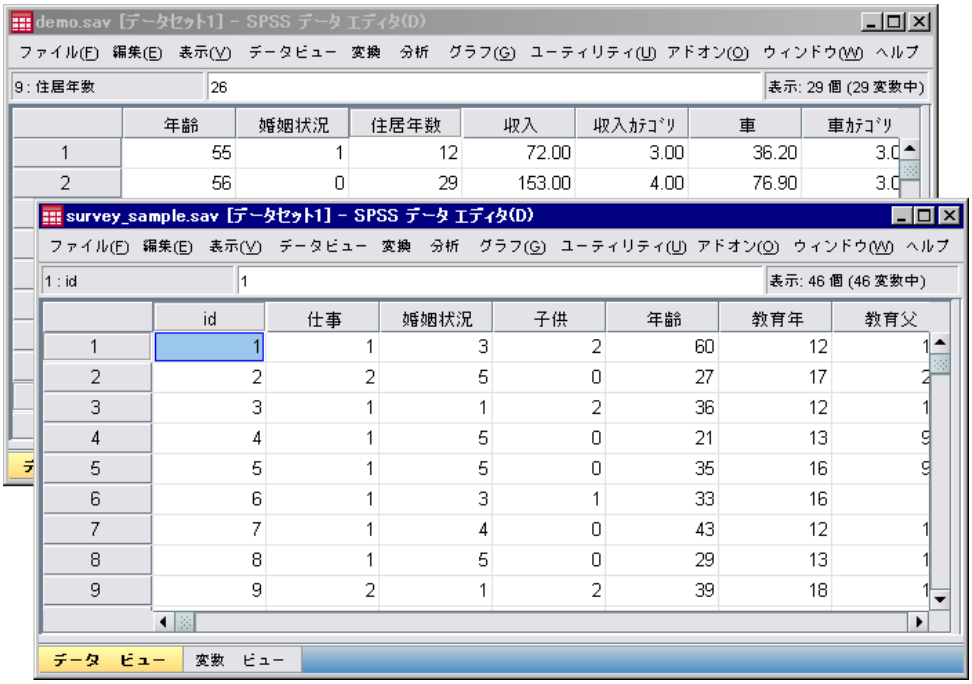
多重データ ソースの使用方法

バージョン 14.0 を開始すると、多重データ ソースを同時に開くことにより、以下のことが簡単に実行できます

- データ ソース間を前後に切り替え。
- 異なるデータ ソースの内容の比較。
- データ ソース間のデータのコピーと貼り付け。
- 分析用のケースまたは変数、あるいはその両方の複数のサブセットを作成。
- 最初に各データを保存することなく、様々なデータ形式 (たとえば、スプレッドシート、データベース、テキスト データ) から多重データ ソースを結合。

多重データ ソースの基本的な処理

図 6-1
同時に開いている 2 つのデータ ソース

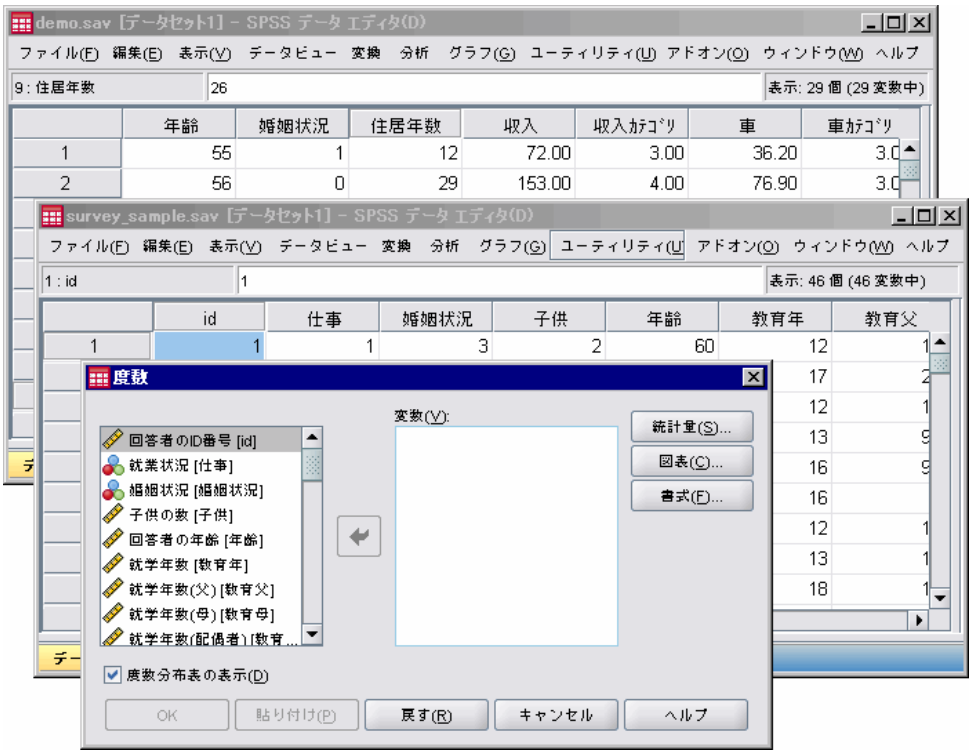


デフォルトでは、開いたデータ ソースがそれぞれ新しい [データ エディタ] ウィンドウに表示されます。(単一 [データ エディタ] ウィンドウの中で 1 個のデータセットを表示するためのデフォルトの動作の変更の情報は、[全般] オプションを参照してください。)

- すでに開いているデータ ソースは、開いたままさらに使用できます。
- データ ソースは、最初に開いた時点で、自動的に **アクティブなデータセット**に変わります。
- 使用するデータ ソースの [データ エディタ] ウィンドウの任意の場所をクリックするだけで、または [ウィンドウ] メニューからそのデータ ソースの [データ エディタ] ウィンドウを選択することにより 作業データセットを変更することができます。

- 作業データセットの変数のみが分析に使用できます。

図 6-2
作業データセットの変数を含む変数リスト



- データにアクセスするダイアログ ボックスが開いているときは、作業データセットを変更することはできません(変数リストを表示するすべてのダイアログ ボックスを含む)。
- セッション中には、少なくとも 1 つの [データ エディタ] ウィンドウが開いていなければなりません。開いている最後の [データ エディタ] ウィンドウを閉じると、SPSS は自動的にシャットダウンし、最初に変更の保存を要求するダイアログ ボックスが表示されます。

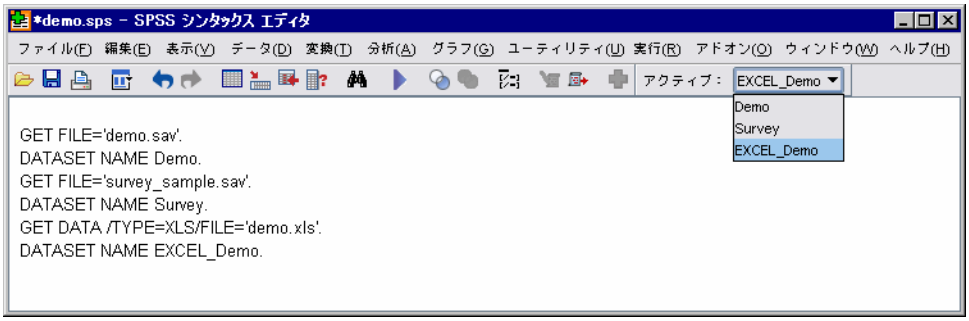
コマンド シンタックスの複合データセットを使用する作業

コマンド シンタックスを使ってデータ ソースを開く場合 (たとえば、GET FILE、GET DATA)、複数のデータ ソースを同時に開くためには各データセットに対して明示的に名前を指定する DATASET NAME を使用する必要があります。

コマンド シンタックスを使用すると、アクティブなデータセット名がシンタックス ウィンドウのツールバーに表示されます。次のアクションでアクティブ データセットが変更します。

- DATASET ACTIVATE コマンドの使用。
- データセットの [データ エディタ] のどこでもクリック。
- シンタックス ウィンドウのツールバーからデータセット名を選択。

図 6-3
シンタックス ウィンドウのツールバーに表示されたデータセットを開く



データセット間での情報のコピーと貼り付け

単独のデータ ファイル内で情報をコピーして貼り付けるのと基本的に同じ方法でデータと変数定義属性の両方を 1 つのデータセットから別のデータセットにコピーすることができます。

- 選択したデータ セルを [データ ビュー] にコピーして貼り付けると、データ値だけが貼り付けられ、変数定義属性は貼り付けられません。
- 列の 1 番上の変数の名前を選択して [データ ビュー] の変数全体をコピーして貼り付けると、すべてのデータおよびその変数のすべての変数定義属性が貼り付けられます。
- 変数定義属性または [変数ビュー] 内の変数全体をコピーして貼り付けると、選択された属性 (すなわち変数定義全体) は貼り付けられますが、データ値が貼り付けられることはありません。

データセットの名前の変更

メニューやダイアログ ボックスを使用してデータ ソースを開くと、それぞれのデータ ソースには自動的に DataSetn というデータセット名が割り当てられます。ここで、*n* は、連続する整数値 を表します。ただし、コマンド シンタックスを使ってデータ ソースを開くときは、DATASET NAME でデータセット名を明示的に指定しないと、データセット名は割り当てられません。よりわかりやすいデータセット名を付けるには、以下のようにします。

- 名前を変更するデータセットの [データ エディタ] ウィンドウのメニューから、以下を選択します。

  ファイル
  データセットの名前の変更...

- 変数命名規則に従った新しいデータセット名を入力します。詳細は、85 ページの 5 章の変数名を参照してください。

複合データセットを抑制

複合データセット機能を抑制し、一度に 1 個のデータセットだけを利用したいときは次のようにします。

- メニューから次の項目を選択します。

  編集
  オプション...

- [**全般**] タブをクリックします。

  [**一度に 1 個のデータセットを開く**] を選択 (クリック) します。

  詳細は、498 ページの 45 章の[全般] オプションを参照してください。

章
7

データの準備

データ エディタでデータ ファイルを開くかデータを入力した後は、追加の準備作業を行う必要なく、報告書や図表の作成、分析を開始することができます。作業を容易にするために、次に示すデータ準備機能が用意されています。

- データを記述し、特定の値をどのように扱うかを指定する変数プロパティを割り当てる。
- 重複した情報を含むケースを識別し、それらのケースを分析から除外したりデータ ファイルから削除する。
- 大量の値を持ち得る変数の値範囲を表すいくつかの異なるカテゴリを持つ新しい変数を作成する。

変数プロパティ

データ エディタの [データ ビュー] に入力したデータや、外部ファイル形式 (Excel スプレッドシートやテキスト データ ファイルなど) から SPSS に読み込んだデータは、特定の変数プロパティが設定されていません。そのようなプロパティには次のものがあり、非常に役立ちます。

- 数値コードに対する記述的な値ラベルの定義 (たとえば、0 = **男性**で、1 = **女性**)。
- 欠損値コードの識別 (たとえば、99 = **適用なし**)。
- 尺度 (名義、順序、スケール) の割り当て。

これら (およびその他) の変数プロパティはすべて、データ エディタの [変数ビュー] で割り当てることができます。この処理で役に立つユーティリティもいくつかあります。

- **変数プロパティの定義**は、記述的な値ラベルおよび欠損値の定義に役立ちます。カテゴリ値に数値コードを使用するカテゴリ データでは特に有効です。
- **データ プロパティのコピー**は、既存の SPSS 形式のデータ ファイルを、現在のデータ ファイル内にあるファイルおよび変数プロパティのテンプレートとして使用できます。これは、類似の内容を含む外部形式のデータ ファイル (Excel 形式の月次報告書など) を頻繁に使用する場合は特に有効です。

変数プロパティの定義

[変数プロパティの定義] は、カテゴリ (名義、順序) 変数の説明的な値ラベルを作成するのに役立つよう設計されています。[変数プロパティの定義] の機能は次のとおりです。

- 実際のデータ値をスキャンして、選択した各変数に対して一意的な全データ値をリストします。
- ラベルのない値を認識して、「自動ラベル」機能を提供します。
- 他の変数から選択した変数へ、あるいは選択した変数から複数の追加的な変数へ、定義済みの値ラベルをコピーできます。

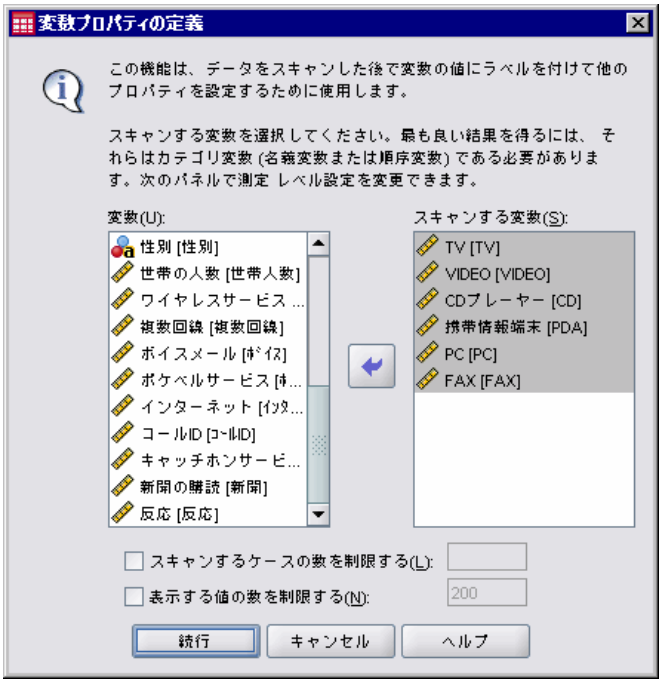
注: ケースをスキャンしないで [変数プロパティの定義] を使用する場合は、スキャンするケース数として 0 を入力します。

変数プロパティの定義

▶ メニューから次の項目を選択します。

データ
変数プロパティの定義...

図 7-1
定義する変数を選択するための最初のダイアログ ボックス



- 値ラベルを作成するために、または欠損値や説明的な変数ラベルなど、その他の変数プロパティの定義や変更を行うために、数値型変数または文字型変数を選択します。
- 一意的な値のリストを生成するには、スキャンするケース数を指定します。ケース数が非常に多いデータ ファイルの場合、データ ファイル全体のスキャンに時間がかかるので、このようなときに特に役立ちます。
- 表示する一意的な値の数に上限を指定します。尺度 (連続区間、比率) 変数の値のために、数百、数千、あるいは数百万の値をリストしないで済むようにする場合、これは特に役立ちます。
- [続行] をクリックすると、メインの [変数プロパティの定義] ダイアログ ボックスが表示されます。
- 値ラベルを作成するために、あるいはその他の変数プロパティの定義や変更を行うために、変数を選択します。
- [値ラベル] グリッドに表示されているラベルのない値に対して、ラベル テキストを入力します。
- 値ラベルを作成したい値が表示されていない場合は、スキャンした最後の値の下にある [値] 列に値を入力します。
- 値ラベルを作成したいリスト中の各変数に対して、この手順を繰り返します。
- [OK] をクリックすると、値ラベルや他の変数プロパティが適用されます。

値ラベルと他の変数プロパティの定義

図 7-2
[変数プロパティの定義] -> メイン ダイアログ ボックス

メインの [変数プロパティの定義] ダイアログ ボックスは、スキャンされた変数に関して次の情報を提供します。

スキャンされた変数のリスト。スキャンされた各変数について、値ラベルが割り当てられていない値が含まれている場合は、**[ラベルなし]** ([U]) 列にチェック マークが表示されます。

変数リストを並べ替えて、ラベルのない値がある変数をリストの上部にすべて表示するには、次の手順に従ってください。

▶ [スキャンされた変数のリスト] で **[ラベルなし]** という列見出しをクリックします。

また [スキャンされた変数リスト] で対応する列見出しをクリックすると、変数名または尺度で並べ替えを実行できます。

値ラベル グリッド

- **ラベル。**すでに定義した値ラベルを表示します。この列でラベルの追加または変更を実行できます。
- **値。**選択した各変数の一意的な値です。この一意的な値のリストは、スキャンされたケースの数に基づいています。たとえば、データ ファイルで最初の 100 ケースだけをスキャンした場合、このリストにはこれらのケースにある一意的な値だけが反映されます。値ラベルを割り当てたい変数でデータ ファイルをすでに並べ替えた場合、このリストにはデータに実際に存在しているよりもはるかに少ない一意的な値しか表示されていない可能性があります。
- **度数。**スキャンされたケースで各値が検出された件数です。
- **欠損値。**欠損データを表すと定義した値です。チェック ボックスをクリックすることで、カテゴリの欠損値の扱い方を変更できます。チェックは、カテゴリをユーザー欠損カテゴリとして定義したことを示します。ある範囲の値 (たとえば、90 ～ 99) がすでにユーザー欠損値として定義されている変数については、その欠損値カテゴリを [変数プロパティの定義] で追加したり削除したりすることはできません。欠損値の範囲が定義れている変数の欠損値カテゴリを変更するには、データ エディタの [変数ビュー] を使用します。 詳細は、91 ページの 5 章の欠損値 (変数の定義)を参照してください。
- **変更済。**値ラベルの追加または変更を行ったことを示します。

注: 最初のダイアログ ボックスでスキャンするケース数に 0 を指定した場合、[値ラベル] グリッドは、すでに存在する値ラベルおよび選択した変数に定義されている欠損値カテゴリを除いて何も表示されない状態になります。また、測定レベルに関する **[推奨]** ボタンが表示されます。

測定レベル。主に値ラベルはカテゴリ (名義と順序) 変数に役立ちますが、一部の手続きはカテゴリ変数とスケール変数を違った方法で処理するため、時として正しい測定レベルを割り当てることが重要になります。ただし、デフォルトでは、新しい全数値型変数に測定尺度水準が割り当てられます。したがって、実際のところ、多くのカテゴリ変数が最初に尺度として表示される場合があります。

変数に割り当てる測定レベルがわからない場合、[**推奨**] をクリックします。

プロパティをコピー。他の変数から現在選択されている変数へ、あるいは現在選択されている変数から他の複数の変数へ、変数ラベルおよびその他の変数プロパティをコピーできます。

ラベルなしの値。ラベルのない値に対してラベルを自動的に作成するには、[**自動ラベル**] をクリックします。

変数ラベルと表示書式

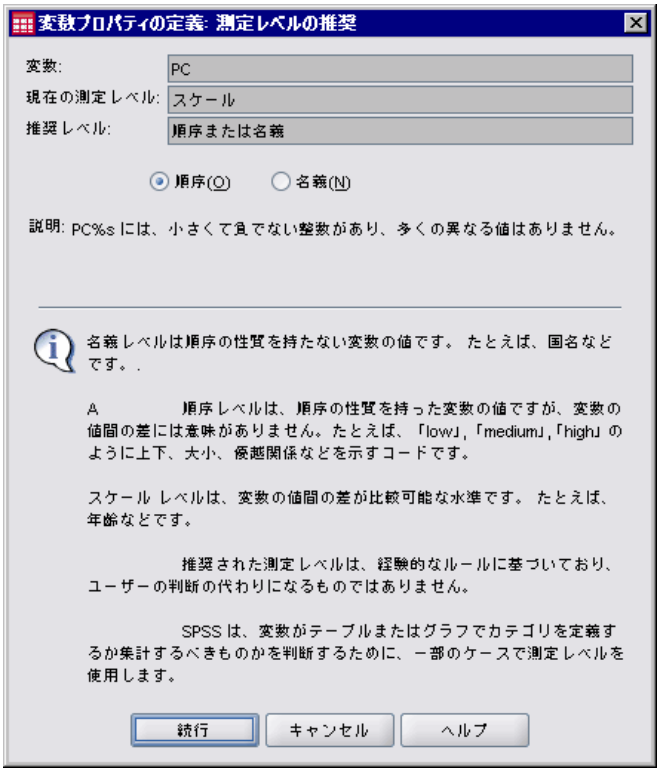
説明的な変数ラベルと表示書式は変更できます。

- 変数の基本的な型 (文字型または数値型) は変更できません。
- 文字型変数の場合、変更できるのは変数ラベルだけで、表示書式は変更できません。
- 数値型変数の場合、数値の型 (数値、日付、ドル記号、ユーザー指定の通貨など)、幅 (小数点または複数桁の区切り文字、あるいはこの両方などを含めた最大桁数)、小数桁数を変更できます。
- 数値型の日付書式では、特定の日付書式 (dd-mm-yyyy、mm/dd/yy、yyyyddd など) を選択できます。
- ユーザー指定の数値型書式では、ユーザー指定の 5 つの通貨書式 (CCA ～ CCE) から 1 つを選択できます。詳細は、503 ページの 45 章の[通貨] オプションを参照してください。
- 指定した幅が、スキャンされた値の幅よりも短い場合、あるいはすでに存在している定義済み値ラベルまたは欠損値カテゴリに対して表示される値よりも短い場合、[**値**] 列にはアスタリスクが表示されます。
- スキャンされた値、あるいはすでに存在している定義済み値ラベルまたは欠損値カテゴリに対して表示される値が、選択した表示書式で無効な場合は、ピリオド (.) が表示されます。たとえば、日付書式変数の場合、86,400 未満の内的な数値は無効とされます。

尺度の割り当て

メインの [変数プロパティの定義] ダイアログ ボックスで測定レベルに関する [**推奨**] をクリックすると、スキャンされたケースと定義済みの値ラベルに基づいて現在の変数が評価され、推奨される測定レベルが [測定レベルの推奨] ダイアログ ボックスに表示されます。[説明] 領域では、尺度の示唆に使用された条件が簡単に説明されます。

図 7-3
[測定レベルの推奨] ダイアログ ボックス



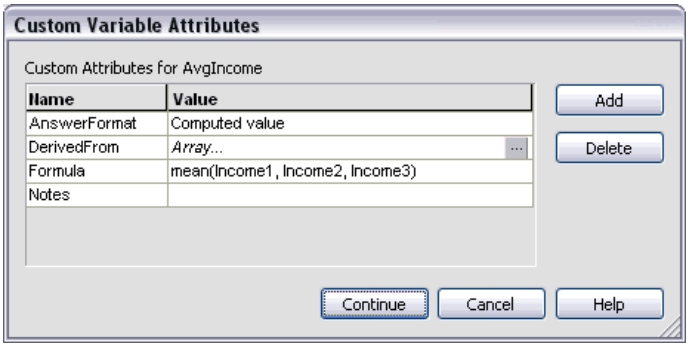
注: 欠損値として定義された値は、測定レベルの評価に含まれません。たとえば、尺度の示唆に関する説明では、この示唆が「変数には負の値を入れないと指定してありますが、実際には負の値がある場合もあります。これは、これらの値がすでに欠損値として定義されているからです」という状況に基づいていると説明されます。

▶ 推奨される測定レベルを使用する場合は、[**続行**] をクリックします。また、現在の測定レベルを変更しない場合は、[**キャンセル**] をクリックします。

ユーザー指定の変数属性

[変数プロパティの定義] の [**属性**] ボタンをクリックすると、[ユーザー指定の変数属性] ダイアログ ボックスが表示されます。値ラベル、欠損値、尺度など標準の変数属性に加えて、ユーザー指定の変数属性を作成することができます。標準の変数属性と同様に、ユーザー指定の属性は SPSS 形式のデータ ファイルとともに保存されます。

図 7-4
ユーザー指定の変数属性



名前。属性名は、変数名と同じ命名規則に従っている必要があります。詳細は、85 ページの 5 章の変数名を参照してください。

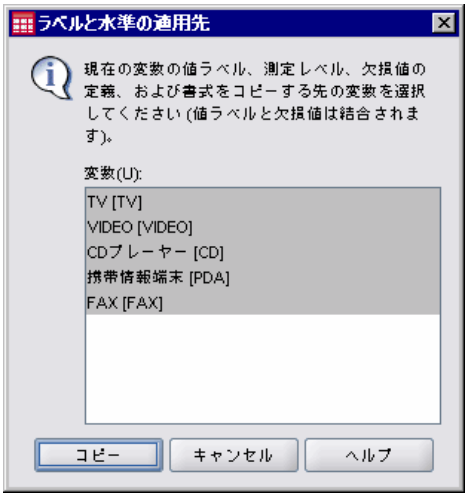
値。選択した変数の属性に割り当てられた値。

- ドル記号で始まる属性名は、予約済みのため変更できません。
- [値] セル内に「**配列...**」と表示されている場合、**属性配列**であることを意味しています。この属性には複数の値が含まれます。値のリストを表示するには、セル内のボタンをクリックします。

変数プロパティのコピー

メインの [変数プロパティの定義] ダイアログ ボックスで **[別の変数から]** または **[他の変数へ]** をクリックすると、[ラベルと水準の適用] ダイアログ ボックスが表示されます。スキャンされた変数の中で、現在の変数の型 (数値型または文字型) に一致したものがすべて表示されます。文字型変数の場合、設定した幅も一致する必要があります。

図 7-5
[ラベルと水準の適用先] ダイアログ ボックス



- 値ラベルやその他の変数プロパティ (変数ラベル以外) をコピーするために、コピー元とする 1 つの変数を選択します。

  または

- 値ラベルやその他の変数プロパティをコピーするために、コピー先とする 1 つ以上の変数を選択します。

- 値ラベルや測定レベルをコピーするには、**[コピー]** をクリックします。
  - コピー先の変数にある既存の値ラベルや欠損値カテゴリは置換されません。
  - コピー先の変数にある特定の値に、値ラベルや欠損値カテゴリが定義されていない場合、コピー先の変数に指定された値ラベルや欠損値カテゴリにこれらが追加されます。
  - コピー先の変数の尺度は、常に置換されます。
  - コピー元またはコピー先の変数のいずれかに欠損値の定義済みの範囲がある場合、欠損値の定義はコピーされません。

多重回答グループ (多重回答)

カスタム テーブルと図表ビルダーは、**多重回答グループ**という特殊な「変数」をサポートします。多重回答グループは、実際には、一般的な意味の「変数」ではありません。多重回答グループはデータ エディタには表示されず、その他の手続きにも認識されません。多重回答グループは、複数の変数を使用して、回答者が複数の回答をする調査アンケートの回答を記録します。多重回答グループはカテゴリ変数として扱われます。カテゴリ変数で実行できるほとんどの処理は、多重回答グループでも実行できます。

多重回答グループは、データ ファイルにある複数の変数で作成されます。多重回答グループは、SPSS 形式のデータ ファイルにある特別な構成要素です。SPSS 形式のデータ ファイルでは多重回答グループの定義と保存が可能ですが、多重回答グループのインポートまたはエクスポートを、他のファイル形式のファイルとの間で行うことはできません。[データ エディタ] ウィンドウの [データ] メニューにある [データ プロパティのコピー] を使用すれば、他の SPSS データ ファイルから多重回答グループをコピーできます。詳細は、124 ページのデータ プロパティのコピーを参照してください。

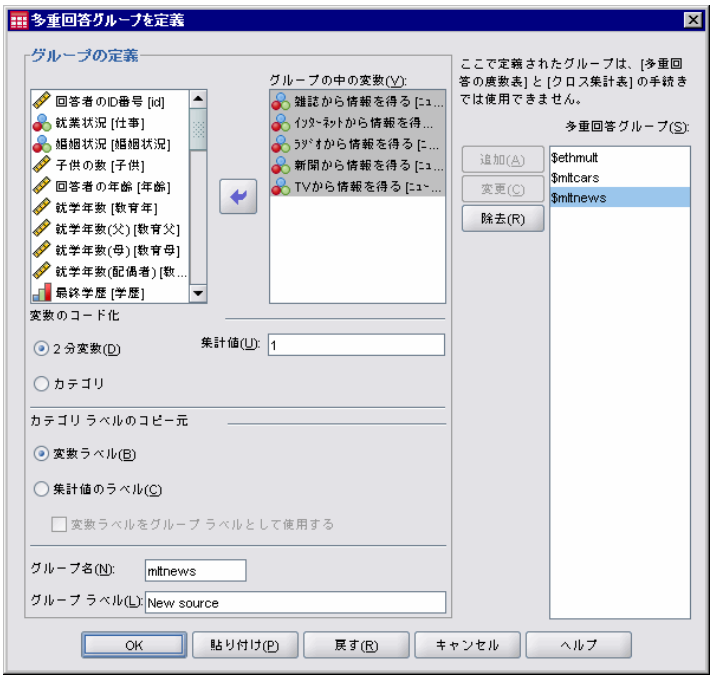
多重回答グループを定義

多重回答グループを定義するには

▶ メニューから次の項目を選択します。

分析
多重回答グループを定義...

図 7-6
[多重回答グループを定義] ダイアログ ボックス



▶ 2 つ以上の変数を選択します。変数が 2 分変数としてコード化されている場合には、数えておきたい値を指定します。

▶ 各多重回答グループに対して、一意の名前を入力します。この名前の長さは最大で 64 バイトです。グループ名の先頭には、ドル記号が自動的に付加されます。

- ▶ グループを説明するラベルを入力します。(これはオプションです)。
- ▶ 定義したグループのリストに多重回答グループを追加するには、[**追加**] をクリックします。

2 分変数 (多重回答グループの定義)

多重 2 分変数グループは一般的に複数の 2 分変数、つまり、はい/いいえ、有/無、チェック入り/チェック解除など、2 つの値しかない変数で構成されています。これらの変数は厳密に 2 分変数でない場合もありますが、このグループにあるすべての変数は同じ方法でコード化され、[集計値] によって正負/有無/チェックなどの状態が表現されます。

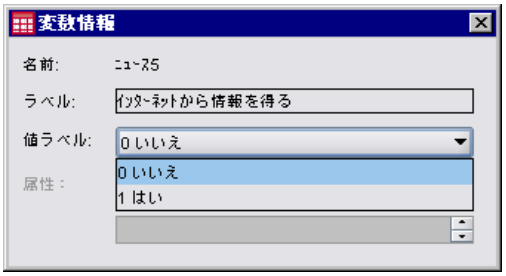
たとえば、ある調査で「ニュースを知る上でよく利用する情報源はどれですか」という質問をして、5 つの選択肢を提示します。回答者は、各選択肢の隣りにあるボックスにチェックを入れ、複数の選択肢を選択できます。この 5 つの回答はデータ ファイル内で 5 つの変数となり、**いいえ** (チェック解除) の場合は 0 にコード化され、**はい** (チェック入り) の場合は 1 にコード化されます。多重 2 分グループでは、[集計値] は 1 となります。

サンプル データ ファイルの survey_sample.sav には、すでに定義した多重回答グループが 3 つあります。$mltnews は多重 2 分グループです。

- ▶ [多重回答グループ] リストで $mltnews を選択 (クリック) します。回答グループが一覧表示されます。

  ここには、多重回答グループの定義に使用する変数と設定が表示されます。
  - [変数] リストには、多重回答グループの作成に使用する 5 つの変数が表示されます。
  - [変数のコード化] グループには、これらの変数が 2 分変数であることが示されます。
  - [カウントされる値] は 1 です。
- ▶ [変数] リストで変数の 1 つを選択 (クリック) します。
- ▶ 変数を右クリックし、ポップアップ コンテキスト メニューから [**変数情報**] を選択します。
- ▶ [変数情報] ウィンドウで [値ラベル] ドロップダウン リストの矢印をクリックし、定義済みの値ラベルのリストをすべて表示します。

図 7-7
多重 2 分ソース変数の変数情報



値ラベルは、その変数が 0 または 1 の値を持つ 2 分変数であることを示します。0 は **いいえ**、1 は **はい**を表します。このリストにある 5 つの変数はすべて同じ方法でコード化されます。値 1 (**はい**のコード) は、多重 2 分グループの集計値です。

カテゴリ

多重カテゴリ グループは複数の変数で構成されます。すべての変数が同じ方法でコード化され、ほとんどの場合、回答カテゴリでコード化されます。たとえば、「あなたの民族性を説明するのに最も適した国を 3 つまで挙げてください」という調査項目があるとします。数百もの回答が予想されますが、コード化を行うために、最も一般的な 40 か国だけをリストします。それ以外はすべて「その他」のカテゴリにまとめます。データ ファイルでは、3 つの選択肢が 3 つの変数になり、それぞれに 41 のカテゴリ (コード化した 40 の国と「その他」という 1 カテゴリ) が使用されます。

サンプル データ ファイルでは、**$ethmult** と **$mltcars** が多重カテゴリ グループです。

カテゴリ ラベル ソース

多重 2 分変数の場合、グループへのラベル付けについてさまざまな設定を行うことができます。

- **変数ラベル。**グループ カテゴリのラベルとして、定義済みの変数ラベル (定義済みの変数ラベルがない変数の場合は変数名) が使用されます。たとえば、グループ内のすべての変数において、集計値 (**はい**など) に同じ値ラベル (または定義済みの値ラベル) が設定されている場合、グループ カテゴリのラベルには変数ラベルを使用する必要があります。
- **集計値のラベル。**グループ カテゴリのラベルとして、集計値の定義済み値ラベルが使用されます。このオプションを選択するのは、すべての変数において、集計値に定義済みの値ラベルが設定されている場合、かつ集計値に対する値ラベルが変数ごとに異なる場合に限ります。
- **変数ラベルをグループ ラベルとして使用する。**[**集計値のラベル**] を選択している場合は、グループ ラベルとして、定義済みの変数ラベルを持つグループ内の最初の変数に対する変数ラベルを使用することもできます。定義済みの変数

ラベルを持つ変数がグループ内に存在しない場合は、グループ ラベルとして、グループ内の最初の変数に対する名前が使用されます。

データ プロパティのコピー

[データ プロパティのコピー] ウィザードでは、外部の SPSS データ ファイルを使用することにより、アクティブなデータセット内でファイルや変数のプロパティを定義することができます。またアクティブなデータセットにある変数は、同じデータセットにある別の変数のテンプレートとしても利用できます。次のようにできます。

- 外部データ ファイルまたは開いているデータセットから、選択したファイルのプロパティをアクティブなデータセットにコピー。ファイルのプロパティには、文書、ファイル ラベル、多重回答グループ、変数グループ、および重み付けがあります。
- 外部データ ファイルまたは開いているデータセットから選択した変数のプロパティを、アクティブなデータセット内の一致する変数にコピー。変数のプロパティには、値ラベル、欠損値、尺度、変数ラベル、印刷および書き込み形式、位置合わせ、列幅 (データ エディタ内) があります。
- 外部データ ファイル、開いているデータセット、またはアクティブなデータセット内の 1 変数から選択した変数プロパティを、アクティブなデータセット内の多数の変数にコピー。
- 外部データ ファイルまたは開いているデータセット内で選択した変数を基に、アクティブなデータセット内に変数を新規作成。

データ プロパティをコピーする場合は、次の一般的な規則が適用されます。

- 外部データ ファイルをソース データ ファイルとして使用する場合は、SPSS 形式のデータ ファイルでなければなりません。
- アクティブなデータセットをソース データ ファイルとして使用する場合は、そのファイルに、1 つ以上の変数が含まれている必要があります。アクティブな空のデータセットをソース データ ファイルとして使用することはできません。
- ソース データセット内の未定義の (空の) プロパティで、アクティブなデータセット内の定義済みプロパティを上書きすることはできません。
- 変数プロパティは、ソース変数から、一致する型の目標変数、つまり、文字型 (英数字) または数値型 (数字、日付、通貨など) にだけコピーされます。

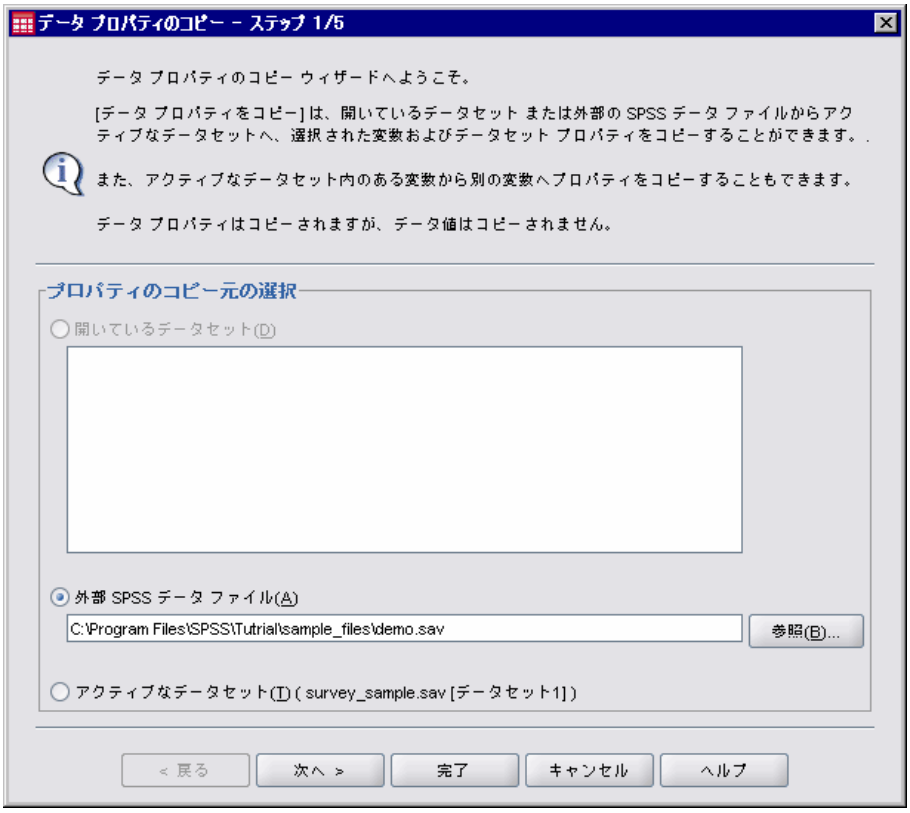
注: [データ プロパティのコピー] は、これまで、[ファイル] メニューからアクセスしていた [データ辞書の適用] に代わる機能です。

データ プロパティをコピーするには

▶ [データ エディタ] ウィンドウのメニューから、次の項目を選択します。

データ
データ プロパティのコピー

図 7-8
[データ プロパティのコピー] ウィザード: ステップ 1

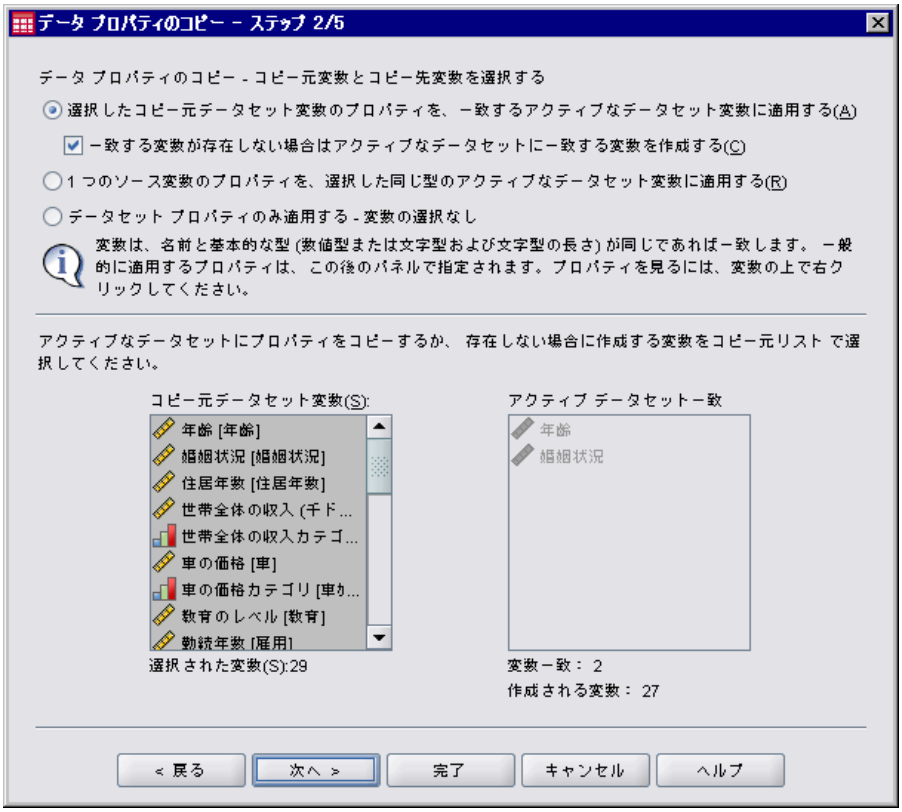


- コピーするファイルまたは変数プロパティを含むデータ ファイルを選択します。これには、現在開いているデータセット、SPSS 形式の外部データ ファイル、またはアクティブなデータセットが利用できます。
- [データ プロパティのコピー] ウィザードの段階的な順番に従います。

ソース変数と目標変数の選択

このステップでは、コピーする変数プロパティを含むソース変数と、その変数プロパティを受け取る目標変数を指定できます。

図 7-9
[データ プロパティのコピー] ウィザード: ステップ 2



選択したコピー元データセット変数のプロパティを、一致するアクティブなデータセット変数に適用する。変数プロパティは、1 つ以上のソース変数から、アクティブなデータセット内の一致する変数にコピーされます。変数は、変数名と型 (文字型または数値型) が同じ場合に限り「一致」します。文字型変数の場合は、定義された長さも同じでなければなりません。デフォルトでは、一致する変数だけが 2 つの変数リストに表示されます。

- **一致する変数が存在しない場合はアクティブなデータセットに一致する変数を作成する。**これにより、コピー元リストが更新されて、すべての変数がコピー元データ ファイルに表示されます。(変数名を基に) アクティブなデータセットに存在しないコピー元変数を選択した場合は、コピー元データ ファイルの変数名およびプロパティを持つアクティブなデータセットに新しい変数が作成されます。

アクティブなデータセットに変数が含まれていない場合 (新規作成された空のデータ ファイル) は、コピー元データ ファイル内のすべての変数が表示され、選択したコピー元変数を基にした新しい変数が、アクティブなデータセットに自動的に作成されます。

1 つのコピー元変数のプロパティを、選択した同じ型のアクティブなデータセット変数に適用する。コピー元リストで選択した 1 つの変数の変数プロパティは、アクティブ データセット リストで選択した 1 つ以上の変数に適用できます。コピー元リストで選択した変数と同じ型 (数値型または文字型) の変数だけが、アクティブ データセット リストに表示されます。文字型変数の場合は、コピー元変数と同じ定義長の文字だけが表示されます。アクティブなデータセットに変数が含まれていない場合は、このオプションを選択できません。

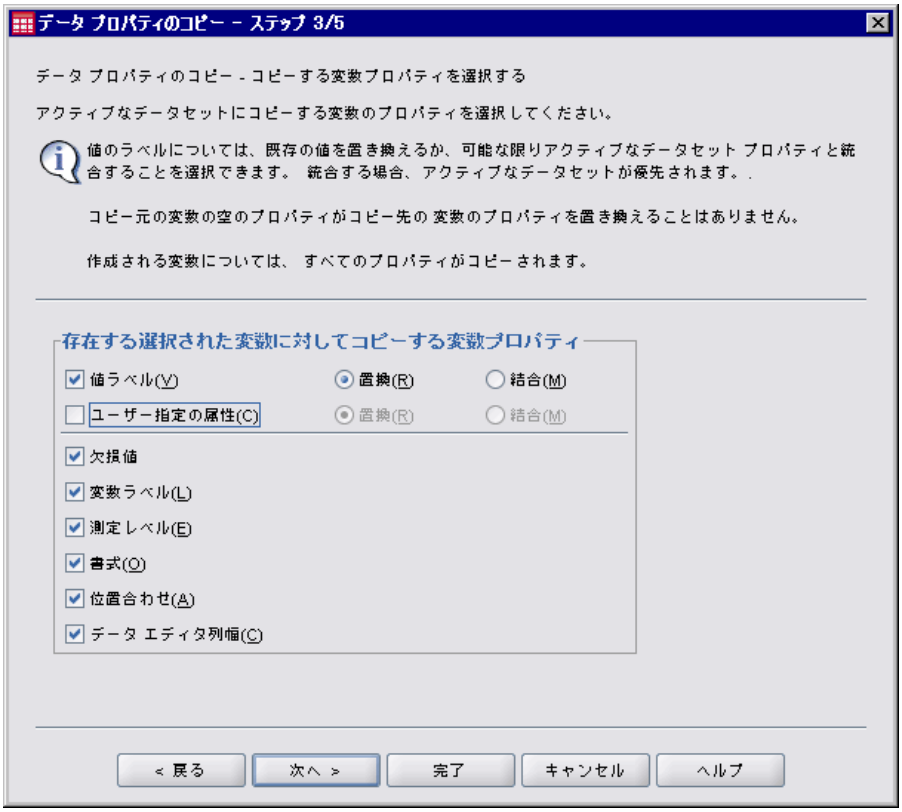
注: このオプションを使用して、アクティブなデータセットに変数を新規作成することはできません。

データセット プロパティのみ適用し、変数の選択なし。ファイルのプロパティ (文書、ファイル ラベル、重みなど) だけがアクティブなデータセットに適用されます。変数のプロパティは適用されません。アクティブなデータセットがソース データ ファイルでもある場合は、このオプションを選択できません。

コピーする変数プロパティの選択

選択したソース変数のプロパティを目標変数にコピーできます。ソース変数内の未定義の (空の) プロパティで、目標変数内の定義済みプロパティを上書きすることはできません。

図 7-10
[データ プロパティのコピー] ウィザード: ステップ 3



値ラベル。値ラベルは、データ値と関連付けられた記述的なラベルです。値ラベルは、数値データ値を使用して数値以外のカテゴリを表している場合に利用できます (たとえば、**男性**と **女性**を表すためのコードとして 1 と 2 を使用している場合)。目標変数内の値ラベルを置換または結合できます。

- [**置換**] は、目標変数の定義済み値ラベルを削除して、ソース変数の値ラベルに置き換えます。
- [**結合**] は、ソース変数の定義済み値ラベルを、目標変数が持つ既存の定義済み値ラベルと結合します。同じ値が、ソース変数と目標変数の両方に定義済み値ラベルを持っている場合、目標変数の値ラベルは変わりません。

ユーザー指定の属性。ユーザー指定の変数属性。詳細は、94 ページの 5 章のユーザー指定の変数属性を参照してください。

- [**置換**] は、目標変数のユーザー指定の属性を削除して、ソース変数の定義済み属性に置き換えます。
- [**結合**] は、ソース変数の定義済み属性を、目標変数が持つ既存の定義済み属性と結合します。

欠損値。欠損値は、欠損データを表すものとして識別される値です (たとえば、98 は **わからない**、99 は **該当しない**を表します)。通常、これらの値は、欠損値コードが表す内容を示す定義済みの値ラベルを持っています。目標変数の既存の定義済み欠損値は削除され、ソース変数の定義済み欠損値に置き換えられます。

変数ラベル。記述的な変数ラベルには、変数名に許されていない空白や予約語を含めることができます。1 つのソース変数の変数プロパティを複数の目標変数にコピーする場合は、このオプションを選択する前に再考してください。

測定レベル。測定レベルは、名義、順序、またはスケールのいずれかです。異なる尺度を区別する手続きでは、名義と順序は、どちらも **カテゴリ**とみなされます。

書式。数値型変数の場合は、ここで、数値型 (数値、日付、通貨など)、幅 (接頭、接尾文字や少数点などの表示文字の総数)、および表示される小数桁数を管理します。このオプションは、文字型変数では無視されます。

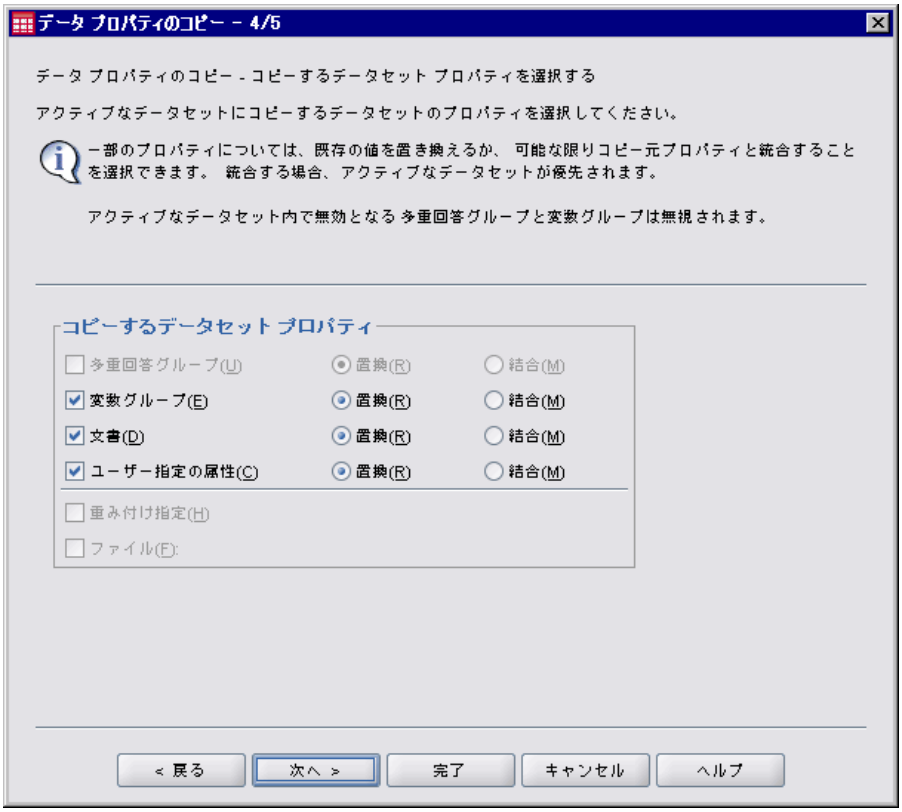
位置合わせ。データ エディタの [データ ビュー] 内の配置 (左、右、中央) だけに作用します。

データ エディタ列幅。データ エディタの [データ ビュー] 内の列幅だけに作用します。

データセット (ファイル) プロパティのコピー

選択したソース データ ファイルのグローバル データセット プロパティを、アクティブなデータセットに適用できます。(アクティブなデータセットがソース データ ファイルである場合は、このオプションを選択できません)。

図 7-11
[データ プロパティのコピー] ウィザード: ステップ 4



多重回答グループ。ソース データ ファイルの多重回答グループ定義をアクティブなデータセットに適用します (。(**注**: 現在は、チャート ビルダおよびテーブルの追加成分だけが多重回答グループを使用しています)。

- 変数が、[データ プロパティのコピー] ウィザードのステップ 2 (ソース変数と目標変数の選択) の指定に基づいて作成されたものでない場合、アクティブなデータセットに存在しない変数を含むソース データ ファイルの多重回答グループは無視されます。
- [**置換**] は、アクティブなデータセット内の多重回答グループを削除して、ソース データ ファイル内の多重回答グループに置き換えます。
- [**結合**] は、ソース データ ファイル内の多重回答グループをアクティブなデータセット内の多重回答グループのコレクションに追加します。両方のファイルに同じ名前のグループが存在しても、アクティブなデータセット内の既存のグループは変更されません。

変数グループ。[変数グループ] は、ダイアログ ボックスに表示される変数リストの管理に使用します。変数グループは、[ユーティリティ] メニューから [グループの定義] を選択して定義します。

- 変数が、[データ プロパティのコピー] ウィザードのステップ 2 (ソース変数と目標変数の選択) の指定に基づいて作成されたものでない場合、アクティブなデータセットに存在しない変数を含むソース データ ファイルのグループは無視されます。
- [置換] は、アクティブなデータセット内の既存の変数グループを削除して、ソース データ ファイル内の変数グループに置き換えます。
- [結合] は、ソース データ ファイル内の変数グループをアクティブなデータセット内の変数グループのコレクションに追加します。両方のファイルに同じ名前のグループが存在しても、アクティブなデータセット内の既存のグループは変更されません。

文書。DOCUMENT コマンドにより、データ ファイルに添付された記録。

- [置換] は、アクティブなデータセット内の既存の文書を削除して、ソース データ ファイル内の文書に置き換えます。
- [結合] は、ソース データ ファイルとアクティブなデータセットの文書を結合します。アクティブなデータセットに存在しない、ソース データ ファイルに固有の文書がアクティブなデータセットに追加されます。次に、すべての文書が日付順に並べ替えられます。

ユーザー指定の属性。ユーザー指定のデータ ファイル属性で、通常はコマンド シンタックス の DATAFILE ATTRIBUTE コマンドで作成されます。

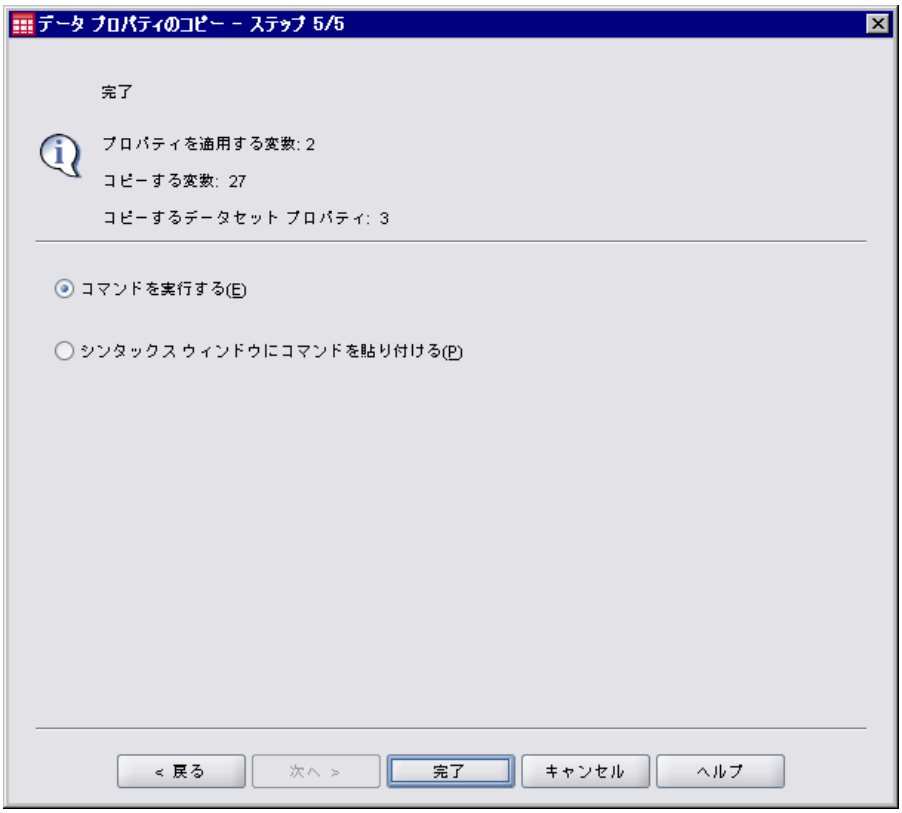
- [置換] は、アクティブなデータセット内の既存のユーザー指定データ ファイル属性を削除して、ソース データ ファイル内のデータ ファイル属性に置き換えます。
- [結合] は、ソース データ ファイルとアクティブなデータセットのデータ ファイル属性を結合します。アクティブなデータセットに存在しない、ソース データ ファイルに固有の属性名がアクティブなデータセットに追加されます。両方のデータ ファイルに同じ属性名が存在しても、アクティブなデータセット内の同名属性は変更されません。

重み付け指定。アクティブなデータセット内に一致する変数が存在する場合は、ソース データ ファイル内の現在の重み付き変数によるケースの重み付け。これは、アクティブなデータセット内で現在有効な重みを無効にします。

ファイル ラベル。FILE LABEL コマンドにより、データ ファイルに適用される記述的なラベル。

結果

図 7-12
[データ プロパティのコピー] ウィザード: ステップ 5



[データ プロパティのコピー] ウィザードの最終ステップでは、プロパティがソースデータ ファイルからコピーされた変数の数、新規作成される変数の数、およびコピーされるデータセット (ファイル) プロパティの数に関する情報が表示されます。

また、生成されたコマンド シンタックスをシンタックス ウィンドウに貼り付け、後で使用できるように保存することもできます。

重複ケースの特定

データ内で「ケースの重複」が発生する場合、次のような理由が考えられます。

- データ入力エラーによって、同じケースが誤って 2 度以上入力されている場合。

- 複数のケースが、共通するプライマリ ID と異なるセカンダリ ID を持つ場合。たとえば、家族全員が 1 つの家に住んでいる場合などです。
- 同じケースを表す複数のケースで、ケースの識別に使用する変数以外で、変数の値が異なっている場合。たとえば、同じ人物または企業が異なる製品を購入した場合や、同じ製品を異なる時期に購入した場合などです。

[重複ケースの特定] を使用すると、「**重複**」を自由に定義し、重複ケースとプライマリ ケースが自動的に決定されるよう、いくつかの設定を行うことができます。

重複ケースを特定し、フラグを付けるには

▶ メニューから次の項目を選択します。
データ
重複ケースの特定...

▶ 重複ケースを識別する変数を 1 つ以上選択します。

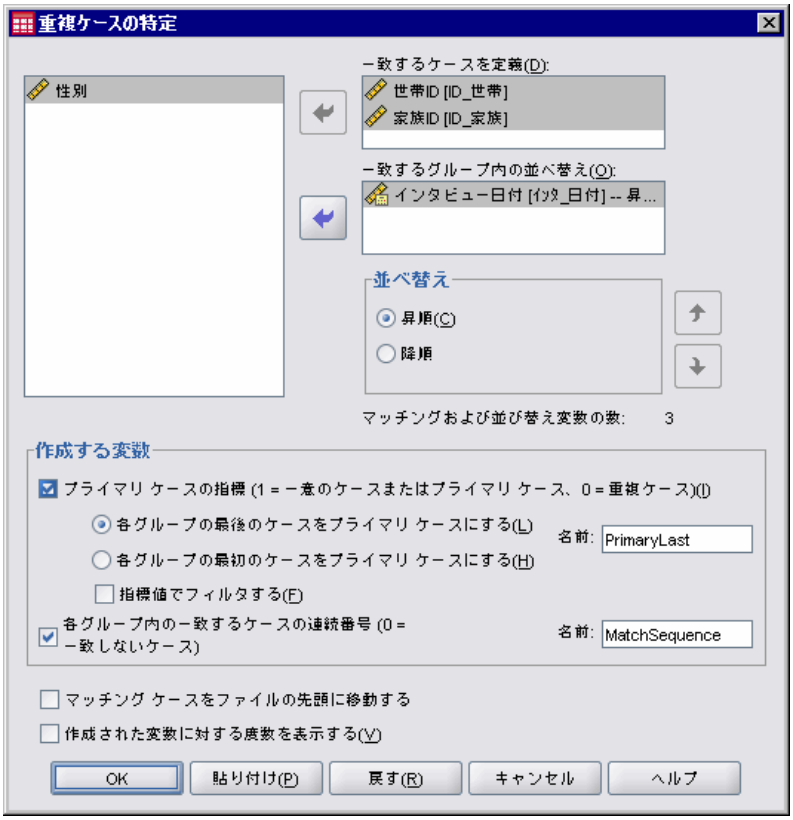
▶ [作成する変数] グループのオプションを 1 つ以上選択します。

オプションとして、次の選択が可能です。

▶ 1 つ以上の変数を選択して、ケースをソートします。ソートするケースは、一致するケースを定義する変数によって定義されたグループ内にあります。これらの変数によって定義されるソート順によって、各グループ内の最初のケースと最後のケースが決定します。ソート順を定義しない場合、元のファイル順が使用されます。

▶ レポート、図表、統計量の計算に重複ケースが含まれないように、自動的に重複ケースが除外されます。

図 7-13
[重複ケースの特定] ダイアログ ボックス



一致するケースを定義。選択した**すべて**の変数について値が一致した場合に、重複ケースとみなします。内容が完全に一致するケースを重複ケースとする場合は、すべての変数を選択します。

一致するグループ内の並べ替え。一致するケースを定義する変数の順で、ケースが自動的にソートされます。一致した各グループ内の順序を決定するために、追加の並べ替え変数を選択できます。

- 各並べ替え変数を使用して、昇順または降順でソートできます。
- 複数の並べ替え変数を選択すると、ケースはボックスのリスト直前の変数のカテゴリ内にある各変数別に並べ替えられます。たとえば、最初の並べ替え変数として **日付**を選択し、2 番目の並べ替え変数として **金額**を選択すると、ケースは日付順に並べ替えられ、さらに各日付の中で金額順に並べ替えられます。
- 変数のソート順は、リストの右にある上下方向の矢印ボタンで変更できます。
- ソート順によって決定する、一致する各グループ内の最初のケースと最後のケースによって、オプションで使用するプライマリ指示変数の値が決定します。たとえば、一致するグループ内の最新のケース以外を除外する場合

は、グループ内のケースを日付変数の昇順でソートして、最新の日付をグループ内の最後の日付にします。

プライマリ ケースの指標。すべての一意なケースと、一致するケースのグループ内のプライマリ ケースとされるケースに値 1、重複ケースのうちプライマリでないケースに値 0、がそれぞれ付いた変数が作成されます。

- 一致したグループ内のソート順によって並び替えられた最初のケースまたは最後のケースのどちらでも、プライマリ ケースに指定することができます。並べ替え変数を指定していない場合は、グループ内のケースは元のファイル順によってソートされます。
- 指示変数を**フィルタ変数**として使用すると、プライマリでないケースをレポートおよび分析で除外することができます (データ ファイルからは削除されません)。

各グループ内の一致するケースの連続番号。一致した各グループ内のケースの順序に応じて、1 ～ *n* の値が付いた変数が作成されます。順序は、元のファイル順または並べ替え変数によって決定した、各グループ内のケースの現在の順序に基づいたものになります。

一致するケースをファイルの先頭に移動する。一致するケースのグループがデータファイルの先頭になるように、データファイルをソートします。これによって、データエディタでの一致するケースの視覚的な検査が容易になります。

作成された変数に対する度数を表示する。作成された変数の各値に関する度数分布表が表示されます。たとえばプライマリ指示変数に関しては、重複の数に相当する、値 0 を持つケース数と、一意なケースとプライマリ ケースの数に相当する、値 1 を持つケース数が表示されます。

欠損値。数値型変数では、システム欠損値は他の値と同様に処理されます。識別変数にシステム欠損値を持つケースは、数値型変数に対して一致する値を持つものとして処理されます。文字型変数では、識別変数に値がないケースは、数値型変数に対して一致する値を持つものとして処理されます。

連続変数のカテゴリ化

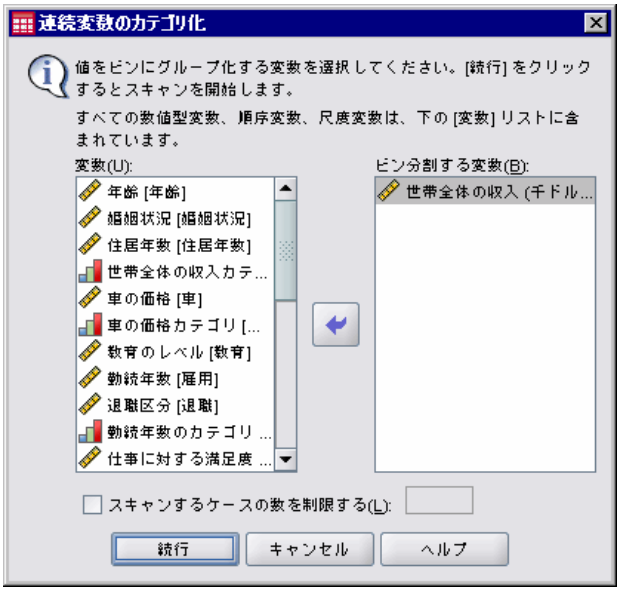
連続変数のカテゴリ化は、既存の変数の連続値を少数の個別カテゴリにグループ化することで、新しい変数を作成する手順を支援します。連続変数のカテゴリ化を使用すると、次のことができます。

- 連続するスケール変数からカテゴリ変数を作成する。たとえば、「収入」スケール変数を使用して、収入の範囲を含む新しいカテゴリ変数を作成できます。
- 大量の順序カテゴリを小さなカテゴリのセットに集約する。たとえば、9 段階の評価尺度を低、中、高を表す 3 つのカテゴリに集約できます。

最初の手順として、次のことを行います。

▶ 新しいカテゴリ (ビン分割) 変数を作成する数値尺度または順序変数を選択します。

図 7-14
ビン分割する変数を選択するための最初のダイアログ ボックス



オプションとして、スキャンするケースの数を制限できます。大量の数のケースがあるデータ ファイルについてはスキャンするケースの数を制限することで時間を短縮できますが、後で連続変数のカテゴリ化で行う計算に使用される値の分布が影響を受けるため、このような操作はできる限り避けてください。

注: 文字型変数と名義数値型変数は、ソース変数リストに表示されません。連続変数のカテゴリ化では、データ値は、値を有意な方法でグループ化するために使用できる論理的順序を表すと想定されるので、尺度レベルまたは順序レベルで測定された数値型変数が必要となります。データ エディタの [変数ビュー] では、定義されている変数の尺度を変更できます。 詳細は、87 ページの 5 章の変数尺度を参照してください。

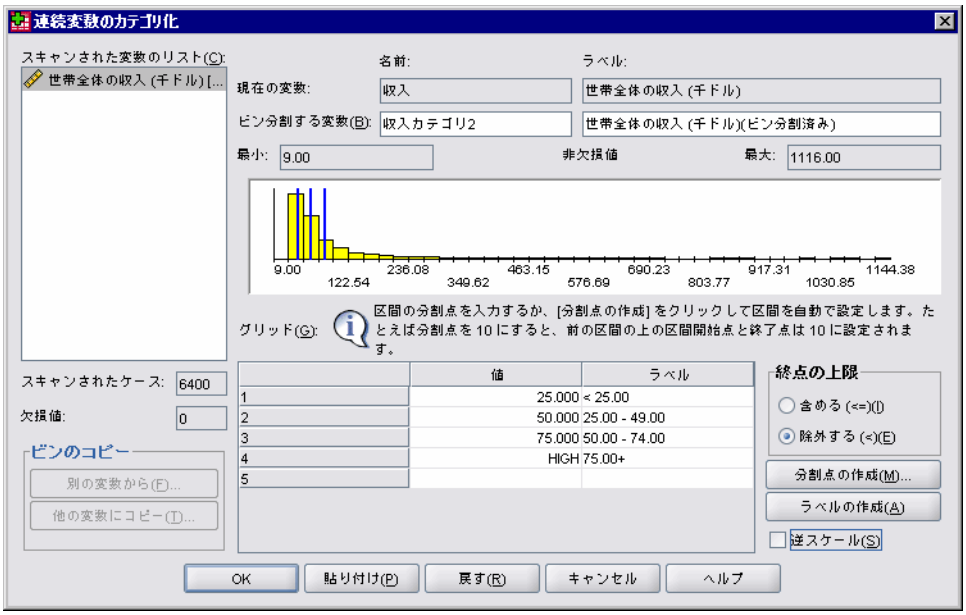
変数をビン分割するには

- [データ エディタ] ウィンドウのメニューから、次の項目を選択します。
  変換
  連続変数のカテゴリ化...
- 新しいカテゴリ (ビン分割) 変数を作成する数値尺度または順序変数を選択します。
- [スキャンされた変数のリスト] で変数を選択します。
- 新しいビン分割変数の名前を入力します。変数名は一意でなければならず、変数のネーミングルール (命名規則) に従っている必要があります。 詳細は、85 ページの 5 章の変数名を参照してください。

- ▶ 新しい変数のビン分割条件を定義します。詳細は、137 ページの変数のビン分割を参照してください。
- ▶ [OK] をクリックします。

変数のビン分割

図 7-15
連続変数のカテゴリ化、メイン ダイアログ ボックス



連続変数のカテゴリ化のメイン ダイアログ ボックスには、スキャンされた変数に関して次の情報が表示されます。

スキャンされた変数のリスト。最初のダイアログ ボックスで選択した変数が表示されます。このリストは、列見出しをクリックすることで、尺度 (尺度または順序) または変数のラベルまたは名前でソートできます。

スキャンされたケース。スキャンされたケースの数が表示されます。スキャンされたケースのうち、選択されている変数に関するユーザー欠損値やシステム欠損値が存在しないケースはすべて、連続変数のカテゴリ化の計算 (メイン ダイアログ ボックスに表示されるヒストグラム、およびパーセンタイルや標準偏差単位に基づく分割点など) で必要となる値の分布を生成するために使用されます。

欠損値。スキャンされたケースの中でユーザー欠損値またはシステム欠損値が存在するケースの数が表示されます。欠損値は、ビン分割されたカテゴリのいずれにも含まれません。詳細は、143 ページの連続変数のカテゴリ化でのユーザー欠損値を参照してください。

現在の変数。新しいビン分割の基準として使用される、現在選択されている変数の名前と変数ラベル (存在する場合)。

ビン分割する変数。新しくビン分割されたする変数の名前とオプションの変数ラベル。

- **名前。**新しい変数の名前を入力します。変数名は一意でなければならず、変数のネーミングルール (命名規則) に従っている必要があります。 詳細は、85 ページの 5 章の変数名を参照してください。
- **ラベル。**半角で最大 255 文字の変数の内容を説明する変数ラベルを入力できます。デフォルトの変数ラベルは、ソース変数の変数ラベル (存在する場合) または変数名の末尾に「**(ビン分割済み)**」という文字列が追加されたラベルになります。

最小/最大。スキャンされたケースに基づいた、現在選択されている変数の最小値と最大値。ユーザー欠損値として定義された値は含まれません。

非欠損値。スキャンされたケースに基づいて、現在選択されている変数の欠損値以外の分布がヒストグラムに表示されます。

- 新しい変数のビンを定義すると、ビンを定義する分割点を示す垂直線がヒストグラムに表示されます。
- 分割点を示す線をクリックしたままヒストグラム上の別の位置にドラッグすると、ビンの範囲を変更できます。
- 分割点を示す線をヒストグラムの外側にドラッグすると、ビンを削除できます。

注: (非欠損値を示す) ヒストグラム、最小値、および最大値は、スキャンされた値に基づいています。スキャンにすべてのケースを含めていない場合は、真の分布が正確に反映されないことがあります。選択した変数でデータファイルがソートされている場合は特にこれに該当します。ケースのスキャンを行わなかった場合、値の分布に関する情報は表示されません。

グリッド。それぞれのビンおよびそのオプションの値ラベルの終点の上限を定義する値が表示されます。

- **値。**各ビンの終点の上限を定義する値。値を入力するか、または [**分割点の作成**] を使用し、選択した条件に基づいてビンを自動的に作成します。デフォルトでは、値 HIGH を持つ分割点が自動的に含められます。このビンには、他の分割点を超えるすべての非欠損値が含まれます。最も小さな分割点によって定義されるビンには、その値以下のすべての非欠損値 (または、終点の上限の定義によってはその値未満の非欠損値) が含められます。
- **ラベル。**新しいビン分割の値を説明しているラベル (オプション)。新しい変数の値は単純に 1～ *n* までの連続する整数となるので、値が何を表すかを説明するラベルを付けておくと便利です。ラベルを入力するか、または [**ラベルの作成**] を使用して自動的に値ラベルを作成します。

グリッドからビンを削除するには

▶ グリッドで表示されたビンの [**値**] または [**ラベル**] セル上で右クリックします。

▶ ポップアップ コンテキスト メニューから、[**行を削除**] を選択します。

注:HIGH のビンを削除すると、最後に指定された分割点の値よりも大きな値を持つすべてのケースに新しい変数のシステム欠損値が割り当てられます。

ラベルまたは定義済みのビンをすべて削除するには

▶ グリッド内の任意の場所を右クリックします。

▶ ポップアップ コンテキスト メニューから、[**すべてのラベルを削除**] または [**すべての分割点を削除**] を選択します。

終点の上限。グリッドの [**値**] 列に入力された終点の上限値の処理方法を制御します。

- **含める** (<=)。[**値**] セルに指定された値を持つケースは、ビン分割されたカテゴリに含められます。たとえば、25、50、および 75 という値を指定した場合、値 25 を持つケースは 1 つ目のビンに含められます。これは、この設定では、25 以下の値を持つすべてのケースが対象となるためです。
- **除外する** (<)。[**値**] セルに指定された値を持つケースは、ビン分割されたカテゴリに含められません。その代わりに、次のビンに含められます。たとえば、25、50、および 75 という値を指定した場合、値 25 を持つケースは 1 つ目のビンではなく 2 つ目のビンに含められます。これは、この設定では、1 つ目のビンには 25 未満の値を持つケースだけが含められるためです。

分割点の作成。同じ幅の区間、同じケース数、または標準偏差に基づく区間となるように、ビン分割されたカテゴリを自動的に生成します。ケースをスキャンしていない場合は、このボタンを使用できません。 詳細は、140 ページのビン分割されたカテゴリの自動生成を参照してください。

ラベルの作成。グリッドの値および終点の上限に関して指定された処理方法 ([含める] または [除外する]) に基づいて、新しくビン分割される変数の連続する整数値に対して説明的なラベルを生成します。

逆スケール。デフォルトでは、新しくビン分割された変数の値は、1～ *n* の昇順に連続する整数となります。逆スケールを指定した場合、値は *n* ～ 1 の降順に連続する整数となります。

ビンのコピー。他の変数から現在選択されている変数へ、あるいは選択されている変数から他の複数の変数へ、ビン分割指定をコピーできます。 詳細は、142 ページのビン分割されたカテゴリのコピーを参照してください。

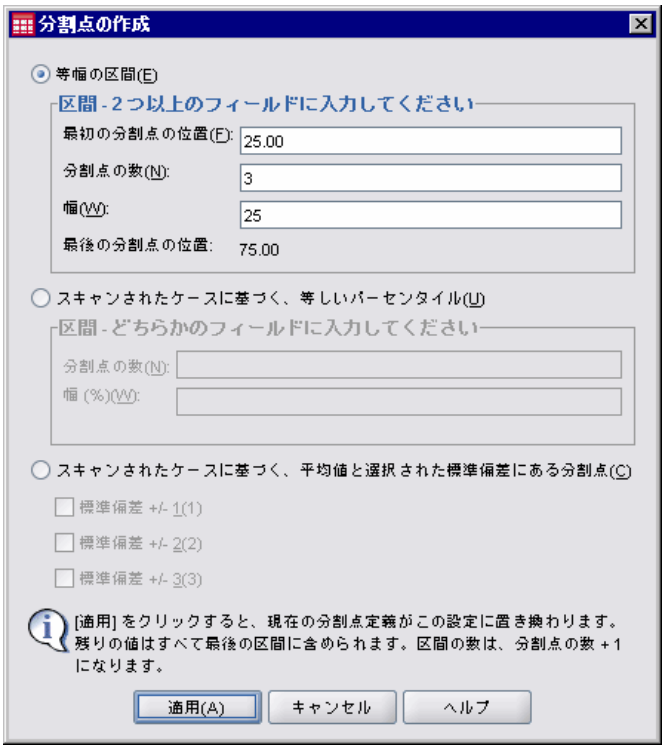
ビン分割されたカテゴリの自動生成

[分割点の作成] ダイアログ ボックスでは、選択した条件に基づいて、ビン分割されたカテゴリを自動的に生成できます。

[分割点の作成] ダイアログ ボックスを使用するには

- ▶ [スキャンされた変数のリスト] で変数を選択 (クリック) します。
- ▶ **[分割点の作成]** をクリックします。
- ▶ 分割点を作成するための条件を選択して、ビン分割されるカテゴリを定義します。
- ▶ **[適用]** をクリックします。

図 7-16
[分割点の作成] ダイアログ ボックス



注: ケースをスキャンしていない場合は、[分割点の作成] ダイアログ ボックスを使用できません。

等幅の区間。次の 3 つの条件のうちの 2 つに基づいて、等幅の (たとえば、1 ～ 10、11 ～ 20、21 ～ 30) ビン分割されたカテゴリを生成します。

- **最初の分割点の位置。**最も小さなビン分割されたカテゴリの上限を定義する値 (たとえば、値 10 は、10 までのすべての値を含む範囲を示します)。

- **分割点の数。**分割点の数に **1 を加えた数**が、ビン分割されたカテゴリの数となります。たとえば、9 個の分割点を指定した場合、10 個のビン分割されたカテゴリが生成されます。
- **幅。**各区間の幅。たとえば、10 を指定した場合、**年齢**は 10 才ごとの値にビン分割されます。

スキャンされたケースに基づく、等しいパーセンタイル。次のどちらかの条件に基づき、(パーセンタイルに **経験的アルゴリズム**を使用して) 各ビンに等しい数のケースを含むビン分割されたカテゴリを生成します。

- **分割点の数。**分割点の数に **1 を加えた数**が、ビン分割されたカテゴリの数となります。たとえば、3 つの分割点を指定した場合、それぞれがケースの 25% を含む 4 つのパーセンタイル ビン (4 分位) が生成されます。
- **幅 (%)。**ケースの合計数のパーセントで表される、各区間の幅。たとえば、値 33.3 を指定した場合、それぞれがケースの 33.3% を含む 3 つのビン分割されたカテゴリ (2 つの分割点) が生成されます。

ソース変数に比較的少数の他と異なる値が含まれている場合、または多数の同じ値を持つケースが含まれている場合は、要求した数よりも少ないビンが生成されることがあります。分割点に同じ値が複数ある場合、これらの値はすべて同じ区間に入るため、実際のパーセントは、必ずしも完全に同じになるとは限りません。

スキャンされたケースに基づく、平均値と選択された標準偏差にある分割点。変数の分布の平均および標準偏差の値に基づいて、ビン分割されたカテゴリを生成します。

- 標準偏差区間を指定しなかった場合は、ビンを分割する分割点として平均値が使用された、2 つのビン分割されたカテゴリが作成されます。
- 平均値 ± 1 標準偏差、平均値 ± 2 標準偏差、および平均値 ± 3 標準偏差の中から、標準偏差の区間を任意に組み合わせて選択できます。たとえば、3 つのオプションすべてを選択した場合は、8 つのビン分割されたカテゴリが生成されます (平均値 ± 1 標準偏差区間で 6 つのビン、平均値の上下で平均値 ± 3 標準偏差を超えるケースに対して 2 つのビン)。

正規分布では、68% のケースが平均値 ± 1 標準偏差の中に含まれ、95% のケースが平均値 ± 2 標準偏差に、99% のケースが平均値 ± 3 標準偏差に含まれます。ビン分割されたカテゴリを標準偏差に基づいて作成すると、実際のデータ範囲の外部や可能なデータ値の範囲の外部にビンが定義されることがあります (たとえば、給与の範囲が負になることがあります)。

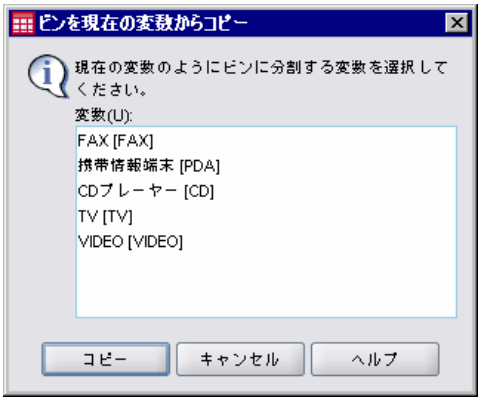
注: パーセンタイルおよび標準偏差の計算は、スキャンされたケースに基づいて行われます。スキャンされるケースの数を制限すると、データ ファイルがソース変数でソートされている場合は特に、生成されたビンにおいてケースの割合が期待どおりにならないことがあります。たとえば、1000 個のケースを含むデータ ファイルの最初の 100 ケースにスキャンを制限し、データ ファイルがケースのそれぞれ 25% を含んでいる 4 つのパーセンタイル年齢ビンではなく、回答者の年齢で昇順

にソートされている場合、最初の 3 つのビンにはケースの約 3.3% しか含まれず、最後のビンにケースの 90% が含まれる結果になります。

ビン分割されたカテゴリのコピー

1 つまたは複数の変数に対してビン分割されたカテゴリを作成する場合、他の変数から現在選択されている変数へ、あるいは選択されている変数から他の複数の変数へ、ビン指定をコピーできます。

図 7-17
現在の変数を対象としたビンのコピー



ビン指定をコピーするには

- ビン分割するカテゴリを 1 つ以上の変数に対して定義します。ただし、[OK] または [**貼り付け**] をクリック**しないで**ください。
- [スキャンされた変数のリスト] で、ビン分割されたカテゴリが定義されている変数を選択 (クリック) します。
- [**他の変数へ**] をクリックします。
- 同じビン分割されたカテゴリを使用して新しい変数を作成するための変数を選択します。
- [**コピー**] をクリックします。

または

- [スキャンされた変数のリスト] で、定義済みのビン分割されたカテゴリのコピー先となる変数を選択 (クリック) します。
- [**別の変数から**] をクリックします。
- コピー元となる、定義済みのビン分割されたカテゴリを含む変数を選択します。
- [**コピー**] をクリックします。

ビン指定のコピー元の変数に対して値ラベルが指定されている場合は、それらのラベルもコピーされます。

注: いったん連続変数のカテゴリ化のメイン ダイアログ ボックスで [OK] をクリックしてビン分割を新しく作成すると（またはダイアログ ボックスを他の方法で閉じると）、連続変数のカテゴリ化を使用してこれらのビン分割されたカテゴリを他の変数にコピーできなくなります。

連続変数のカテゴリ化でのユーザー欠損値

ソース変数のユーザー欠損値（欠損データのコードとして識別された値）として定義されている値は、新しい変数のビン分割されたカテゴリには含まれません。ソース変数のユーザー欠損値は新しい変数のユーザー欠損値として定義され、欠損値コードに対して定義されている値ラベルもすべてコピーされます。

欠損値コードが、新しい変数のビン分割されたカテゴリの値のいずれかと競合する場合、新しい変数の欠損値コードは、最も大きなビン分割されたカテゴリ値に 100 を加えることで、競合しない値に再割り当てされます。たとえば、値 1 がソース変数のユーザー欠損値として定義されていて、新しい変数が 6 つのビン分割されたカテゴリを持つ場合、ソース変数に対して値 1 を持つすべてのケースは新しい変数に対して値 106 を持ち、106 がユーザー欠損値として定義されます。ソース変数のユーザー欠損値に定義済みの値ラベルがある場合、このラベルは、新しい変数の再割り当て値の値ラベルとして保持されます。

注: ソース変数に *LO–n* 形式（*n* は正の値）でユーザー欠損値の範囲が定義されている場合、対応する新しい変数のユーザー欠損値は負の値となります。

章

8

データの変換

理想的な状況では、生データは希望の分析に完全に適合しており、どの変数も、整った線形関係か直交関係にあります。ところが、このような状況はめったに存在しません。予備分析を行うことで、不便なコード方法やコーディング エラーが明らかになります。あるいは、変数間の真の関係を明らかにするために、データ変換が必要となることもあります。

SPSS を使用すると、分析に適したカテゴリの分類などの単純なものから、複雑な方程式や条件文に基づく新しい変数の作成など、さまざまなデータ変換を行うことができます。

変数の計算

[計算] ダイアログ ボックスを使用して、他の変数の数値変換に基づいて変数の値を計算します。

- 数値型変数または文字型変数の値を計算できます。
- 新しい変数を作成したり、変数の値を置換したりできます。新しい変数を作成する場合は、変数型やラベルを指定することもできます。
- 論理条件に基づいて、任意選択的にデータのサブセットの値を計算できます。
- 算術関数、統計関数、分布関数、および文字型関数など、70 種類以上の組み込み関数を使用できます。

図 8-1
[変数の計算] ダイアログ ボックス

変数を計算するには

▶ メニューから次の項目を選択します。

変換
変数の計算...

▶ [目標変数] ボックスに目標変数の名前を 1 つ入力します。これは、既存の変数でも、アクティブなデータ セットに追加する新しい変数でも構いません。

▶ 式を作成するには、[数式] ボックスに成分を貼り付けるか、直接入力します。

- [関数グループ] リストからグループを選択し、[関数と特殊変数] リストで関数または変数をダブルクリックする (または、関数や変数を選択し、[関数グループ] リストの隣にある矢印をクリックする) ことで、関数や通常使用するシステム変数を貼り付けることができます。次に、クエスチョン マークのところにすべてのパラメータを入力します (この操作は関数のみに適用されます)。**[すべて]** というラベルの付いた関数グループには、使用可能な関数およびシステム変数がすべてリスト表示されます。現在選択している関数または変数の簡単な説明が、ダイアログ ボックスの予約領域に表示されます。
- 文字定数は、引用符またはアポストロフィで囲みます。

- 値に小数が含まれる場合、小数点には必ずピリオド (.) を使用してください。
- 新しい文字型変数の場合は、[**型とラベル**] をクリックしてデータ型を指定する必要があります。

変数の計算: IF 条件

[IF 条件] ダイアログ ボックスでは、条件式を使用して、選択したケースのサブセットだけにデータの変換を適用できます。条件式は、ケースごとに **真**の値、**偽**の値、または **欠損値**のいずれかを返します。

図 8-2
[変数の計算: IF 条件] ダイアログ ボックス

- 条件式の結果が **真**の場合、ケースは、選択したサブグループに含まれます。
- 条件式の結果が **偽**または **欠損**の場合、ケースは、選択したサブグループには含まれません。
- 条件式はそのほとんどが、計算パッド上の 6 つの関係演算子 (<、>、<=、>=、=、~=) を使用して表されます。
- 条件式には、変数名、定数、算術演算子、数値、および 関数、論理変数、関係演算子を使用できます。

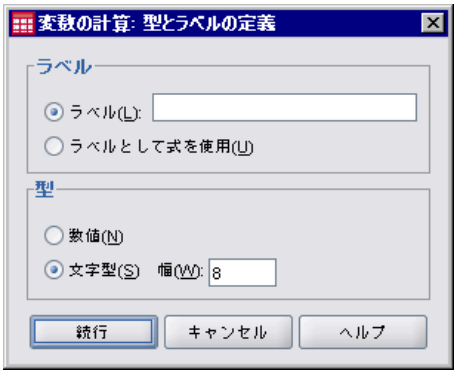
変数の計算: 型とラベルの定義

デフォルトでは、計算された新規の変数は数値型です。新規の文字型変数を計算するには、そのデータ型および幅を指定する必要があります。

ラベル。オプションの記述的な変数ラベルは最大 255 バイトまでです。ラベルを入力するか、計算式の最初の 110 文字をラベルとして使用できます。

型。計算された変数は数値型または文字型です。文字型変数は、計算に使用できません。

図 8-3
[型とラベルの定義] ダイアログ ボックス



関数

SPSS では、次に示すように、多数の関数型をサポートしています。

- 算術関数
- 統計関数
- 文字型関数
- 日付と時刻の関数
- 分布関数
- 乱数変数関数
- 欠損値関数
- スコア関数 (SPSS Server 限定)

各関数についての詳しい説明を参照する場合は、ヘルプ システムの [キーワード] タブに「**関数**」と入力してください。

関数の欠損値

関数と単純な算術式では、欠損値の取り扱い方が異なります。次の式

```
(var1+var2+var3)/3
```

では、3 つの変数のいずれかに欠損値が含まれているケースの場合、結果は欠損となります。

次の式

```
MEAN(var1, var2, var3)
```

では、3 つの変数のすべてに欠損値が含まれているケースだけ、結果が欠損となります。

統計関数では、非欠損値を含む引数の最小値を指定できます。引数の最小値を指定するには、次のように、関数名の後にピリオドと最小値を入力します。

```
MEAN.2(var1, var2, var3)
```

乱数ジェネレータ

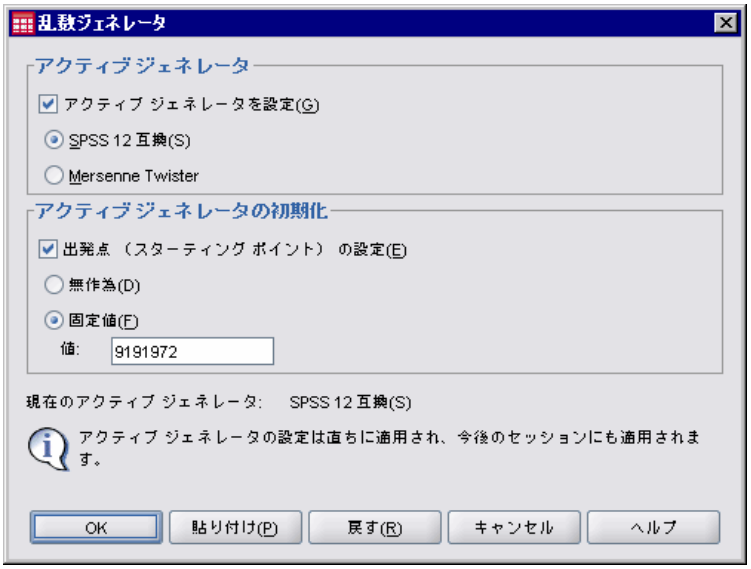
[乱数ジェネレータ] ダイアログ ボックスでは、乱数の系列を再現できるように、乱数ジェネレータを選択して、開始系列値を設定できます。

アクティブ ジェネレータ。2 種類の異なる乱数ジェネレータを使用できます。

- **バージョン 12 互換。**バージョン 12 と以前のリリースで使用された乱数ジェネレータです。指定されたシード値に基づいて以前のリリースで作成されたランダム化された結果を再生する必要がある場合は、この乱数ジェネレータを使用してください。
- **Mersenne Twister。**シミュレーションを行う場合に利用できる、より信頼性の高い新しい乱数ジェネレータです。バージョン 12 またはそれ以前にランダム化された結果の再生ができる場合は、この乱数ジェネレータを使用します。

アクティブ ジェネレータの初期化。乱数のシードは、変換 (乱数分布関数など)、無作為抽出、またはケースの重み付けに使用されるため、乱数が生成されるたびに変わります。乱数の系列を複製するには、乱数を使用する各分析の前に、初期化の始点の値を設定します。値は正の整数にする必要があります。

図 8-4
[乱数ジェネレータ] ダイアログ ボックス



乱数ジェネレータの選択または初期化の値の設定、あるいはその両方を行うには

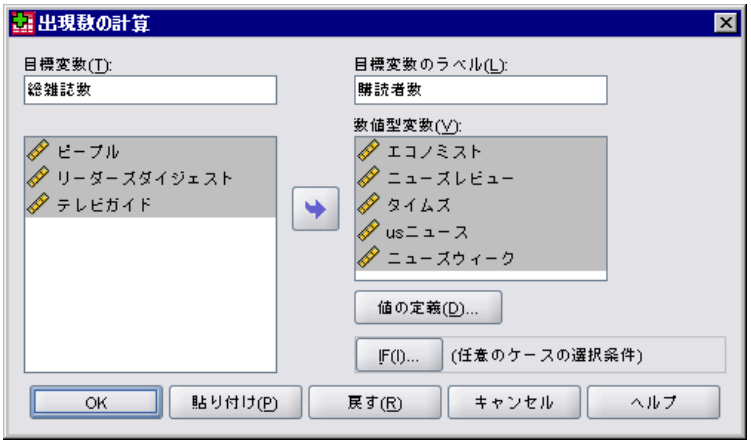
▶ メニューから次の項目を選択します。

変換
乱数ジェネレータ

ケース内での値の出現数の計算

[出現数の計算] ダイアログ ボックスでは、各ケースごとの変数リストにおける同一値の出現数を計算する変数が作成されます。たとえばある調査で、各回答者が普段読んでいる雑誌を調べるために「**はい**」または「**いいえ**」のチェック欄を設けた雑誌のリストを使用するとします。このとき、各回答者が「**はい**」と回答した雑誌の数を集計すれば、読まれている雑誌の総数を表す変数を新たに作成することができます。

図 8-5
[出現数の計算] ダイアログ ボックス



ケース内での値の出現数を計算するには

- ▶ メニューから次の項目を選択します。
  変換
  出現数の計算...
- ▶ 目標変数名を入力します。
- ▶ 変数の一覧から同じ型の変数 (数値型か文字型のどちらか) を 2 つ以上選択して、[変数] ボックスに移動します。
- ▶ **[値の定義]** をクリックして、集計の対象にする値を指定します。

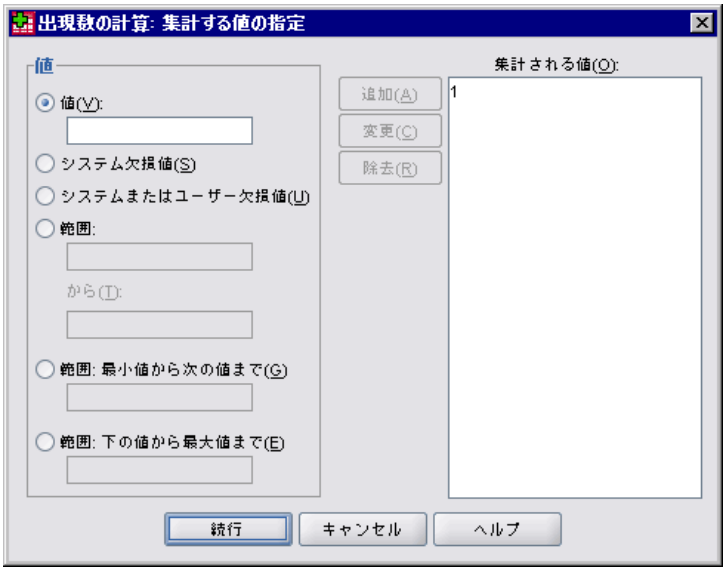
  オプションとして、指定の値を集計するケースのサブセットを定義できます。

出現数の計算: 集計する値の指定

[出現数の計算] ダイアログ ボックスの [目標変数] ボックスの値は、選択された変数のうちの 1 つが集計される値リストにある指定と合致するたびに 1 ずつ増分されます。ケースが任意の変数の複数の指定と合致する場合、目標変数は、その変数に対して合致する回数だけ増分されます。

値の指定には、個々の値、欠損値またはシステム欠損値、範囲を含めることができます。範囲には、その終点およびその範囲内にある任意のユーザー欠損値を含めることができます。

図 8-6
[集計する値の指定] ダイアログ ボックス



出現数の計算: IF条件

[IF 条件] ダイアログ ボックスでは、条件式を使用して、選択されたケースのサブグループの出現数を計算できます。条件式は、ケースごとに **真**の値、**偽**の値、または **欠損値**のいずれかを返します。

図 8-7
[出現数の計算: IF 条件] ダイアログ ボックス

[IF 条件] ダイアログ ボックスを使用する際の一般的な考慮事項は、146 ページの変数の計算: IF 条件を参照してください。

値の再割り当て

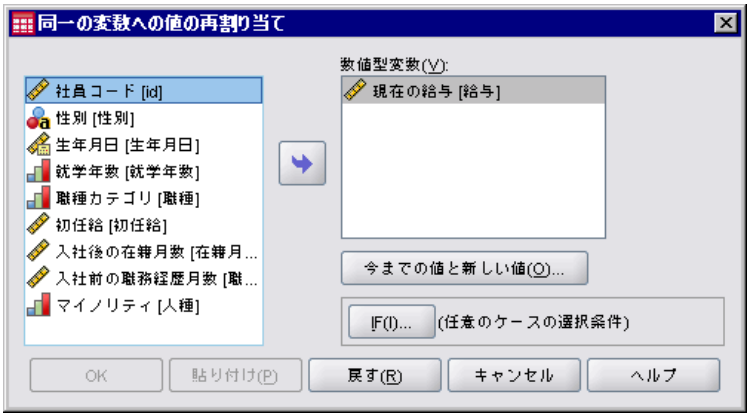
値の再割り当てによりデータ値を修正できます。これは、カテゴリの集約または結合の際に特に有効です。既存の変数内の値を再割り当てしたり、既存の変数の再割り当てされた値に基づいて新規変数を作成したりできます。

同一変数への値の再割り当て

[同一変数への値の再割り当て] ダイアログ ボックスでは、既存の変数の値を再割り当てしたり、または既存の値の範囲を新たな値に集約したりできます。たとえば、ある給料群を給料範囲のグループに集約することなどができます。

数値型変数および文字型変数の値の再割り当てを行うことができます。複数の変数を選択する場合は、同じ型の変数でなければなりません。数値型変数と文字型変数の値の再割り当てを同時に行うことはできません。

図 8-8
[同一変数への値の再割り当て] ダイアログ ボックス



変数の値を再割り当てするには

▶ メニューから次の項目を選択します。

変換
同一変数への値の再割り当て...

▶ 変数の一覧から再割り当てしたい変数を選択して、[変数] ボックスに移動します。複数の変数を選択する場合は、同じ型 (数値型か文字型のどちらか) でなければなりません。

▶ [**今までの値と新しい値**] をクリックして、値の再割り当ての方法を指定します。

オプションとして、再割り当てするケースのサブセットを定義できます。この操作に使用する [IF 条件] ダイアログ ボックスは、出現数の計算で説明したものと同じです。

同一の変数への値の再割り当て: 今までの値と新しい値

このダイアログ ボックスで、再割り当てする値を定義できます。値定義はすべて、メイン ダイアログ ボックスで選択された変数と同一のデータ型 (数値型か文字型のどちらか) でなければなりません。

今までの値。再割り当てする値。単一の値、値の範囲、および欠損値を再割り当てできます。システム欠損値および範囲は、いずれの概念も文字型変数に適用しないため、文字型変数に選択できません。範囲には、その終点およびその範囲内にある任意のユーザー欠損値を含めることができます。

- **値 (ピボットテーブル ノンパラメトリック検定)。**新しい値に再割り当てする個別の今までの値。値は、再割り当てする変数と同じデータ型 (数値型か文字型のどちらか) である必要があります。
- **システム欠損値。**指定した書式に基づいてデータの値が定義されない場合、数値が空白になっている場合、または変換コマンドの結果の値が定義されない場合に、プログラムによって割り当てられる値。数値型変数のシステム欠損

値はピリオドとして示されます。文字型変数では、すべての文字が有効であるため、システム欠損値はありません。

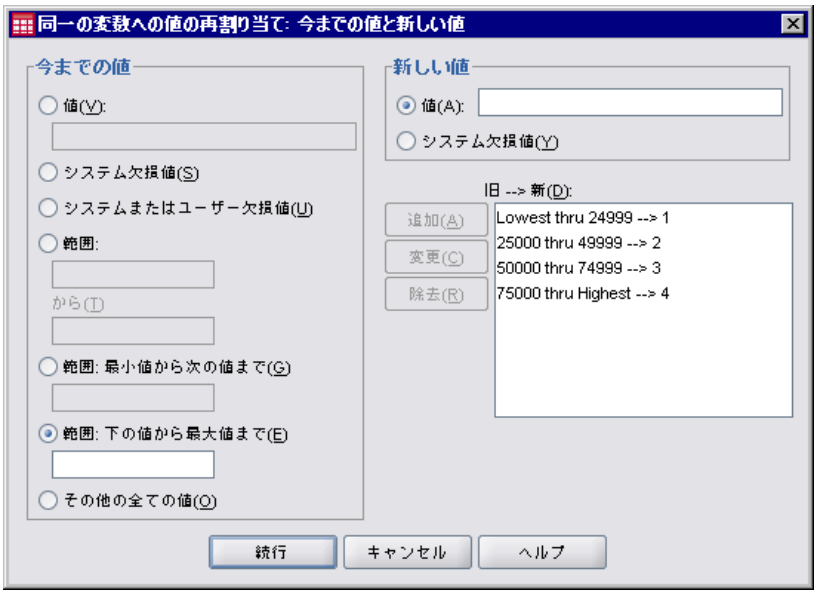
- **システムまたはユーザー欠損値。**ユーザー欠損値として定義されている値、または不明でシステム欠損値として割り当てられている値を持つ観測値。システム欠損値はピリオド (.) で示されます。
- **OK (ファイルオープン時のオプション)。**値の範囲。文字型変数には利用できません。範囲内のユーザー欠損値はすべて含まれます。
- **その他の全ての値。**[旧 - 新] ボックス内の指定に含まれていない残りの値。[旧 - 新] では、ELSE として表示されます。

新しい値。今までの各値または値の範囲が再割り当てされる単一の値。任意の値を入力したり、システム欠損値を割り当てることができます。

- **値 (ピボットテーブル ノンパラメトリック検定)。**今までの 1 つまたは複数の値に再割り当てする値。値は、今までの値と同じデータ型 (数値型か文字型のどちらか) である必要があります。
- **システム欠損値。**指定された今までの値をシステム欠損値に再割り当てします。計算ではシステム欠損値は使用されません。システム欠損値のあるケースは多くの手続きで除外されます。文字型変数には利用できません。

旧 -> 新。変数の再割り当てに使用される指定のリスト。リストの追加、変更および削除ができます。リストは、今までの値の指定に基づいて自動的に並べ替えられますが、その際に、単一の値、欠損値、範囲、他のすべての値という順序に従います。リストにある再割り当ての指定を変更すると、SPSS は必要に応じて自動的にリストの再並べ替えを行い、この順序は変わりません。

図 8-9
[今までの値と新しい値] ダイアログ ボックス

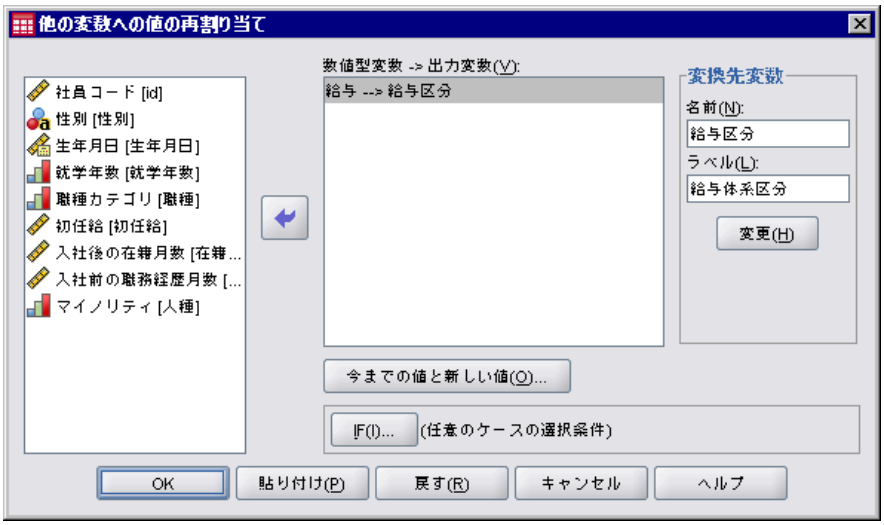


他の変数への値の再割り当て

[他の変数への値の再割り当て] ダイアログ ボックスでは、新しい変数について、既存の変数の値を再割り当てしたり、既存の値の範囲を新たな値に集約したりできます。たとえば、ある給料群を給料範囲グループを含む新しい変数に集約することなどができます。

- 数値型変数および文字型変数の値の再割り当てを行うことができます。
- 数字型変数を文字型変数に、またその逆に再割り当てできます。
- 複数の変数を選択する場合は、同じ型の変数でなければなりません。数値型変数と文字型変数の値の再割り当てを同時に行うことはできません。

図 8-10
[他の変数への値の再割り当て] ダイアログ ボックス



変数の値を新しい変数に再割り当てするには

▶ メニューから次の項目を選択します。

変換
他の変数への値の再割り当て...

▶ 変数の一覧から再割り当てしたい変数を選択して、[変数] ボックスに移動します。複数の変数を選択する場合は、同じ型 (数値型か文字型のどちらか) でなければなりません。

▶ それぞれの新しい変数の出力変数名 (新しい名前) を入力し、**[変更]** をクリックします。

▶ **[今までの値と新しい値]** をクリックして、値の再割り当ての方法を指定します。

オプションとして、再割り当てするケースのサブセットを定義できます。この操作に使用する [IF 条件] ダイアログ ボックスは、出現数の計算で説明したものと同じです。

他の変数への値の再割り当て: 今までの値と新しい値

このダイアログ ボックスで、再割り当てする値を定義できます。

今までの値。再割り当てする値。単一の値、値の範囲、および欠損値を再割り当てできます。システム欠損値および範囲は、いずれの概念も文字型変数に適用しないため、文字型変数に選択できません。今までの値は元の変数と同じデータ型(数値型か文字型のどちらか) でなければなりません。範囲には、その終点およびその範囲内にある任意のユーザー欠損値を含めることができます。

- **値 (ピボットテーブル ノンパラメトリック検定)。**新しい値に再割り当てする個別の今までの値。値は、再割り当てする変数と同じデータ型 (数値型か文字型のどちらか) である必要があります。
- **システム欠損値。**指定した書式に基づいてデータの値が定義されない場合、数値が空白になっている場合、または変換コマンドの結果の値が定義されない場合に、プログラムによって割り当てられる値。数値型変数のシステム欠損値はピリオドとして示されます。文字型変数では、すべての文字が有効であるため、システム欠損値はありません。
- **システムまたはユーザー欠損値。**ユーザー欠損値として定義されている値、または不明でシステム欠損値として割り当てられている値を持つ観測値。システム欠損値はピリオド (.) で示されます。
- **OK (ファイルオープン時のオプション)。**値の範囲。文字型変数には利用できません。範囲内のユーザー欠損値はすべて含まれます。
- **その他の全ての値。**[旧 - 新] ボックス内の指定に含まれていない残りの値。[旧 - 新] では、ELSE として表示されます。

新しい値。今までの各値または値の範囲が再割り当てされる単一の値。新しい値は、数値型または文字型です。

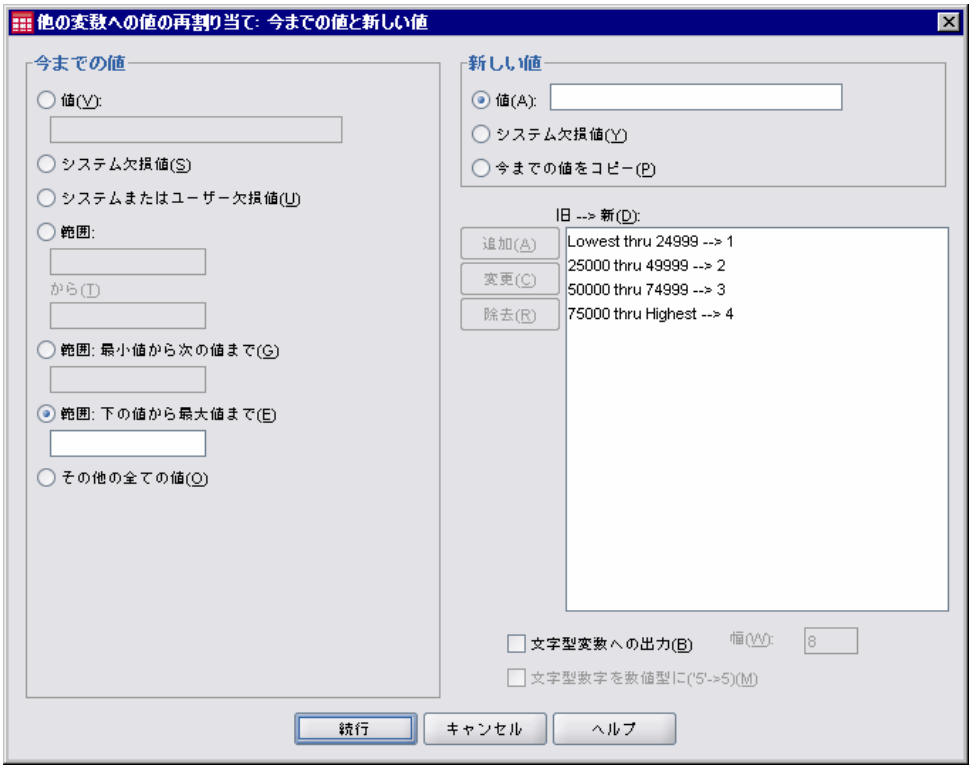
- **値 (ピボットテーブル ノンパラメトリック検定)。**今までの 1 つまたは複数の値に再割り当てする値。値は、今までの値と同じデータ型 (数値型か文字型のどちらか) である必要があります。
- **システム欠損値。**指定された今までの値をシステム欠損値に再割り当てします。計算ではシステム欠損値は使用されません。システム欠損値のあるケースは多くの手続きで除外されます。文字型変数には利用できません。
- **今までの値をコピー。**今までの値を保持します。いくつかの値が再割り当てを必要としない場合は、このオプションを使用して今までの値をそのままコピーします。指定されていない今までの値はいずれも新しい変数には含まれず、こうした値を持つケースには、新しい変数に対してシステム欠損値が割り当てられます。

文字型変数への出力。新しく再割り当てした変数を文字 (英数字) 変数として定義します。古い変数は、数値型と文字型のどちらでもかまいません。

文字型数字を数値型に。数値を含む文字型の値を数値に変換します。数字および任意の符号 (+ または -) 以外のものを含む文字列にはシステム欠損値が割り当てられます。

旧 -> 新。変数の再割り当てに使用される指定のリスト。リストの追加、変更および削除ができます。リストは、今までの値の指定に基づいて自動的に並べ替えられますが、その際に、単一の値、欠損値、範囲、他のすべての値という順序に従います。リストにある再割り当ての指定を変更すると、SPSS は必要に応じて自動的にリストの再並べ替えを行い、この順序は変わりません。

図 8-11
[今までの値と新しい値] ダイアログ ボックス



ケースのランク付け

[ケースのランク付け] ダイアログ ボックスでは、ランクを含む新しい変数、正規スコアとサベージ スコア、および数値型変数のパーセンタイル値を作成できます。

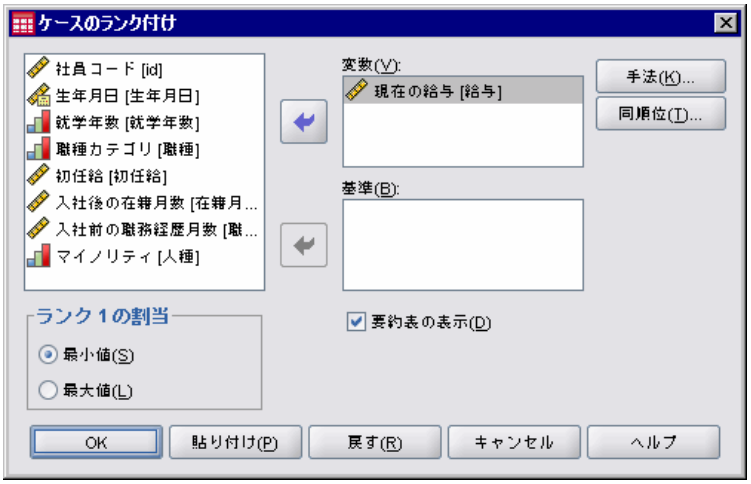
元の変数名および選択された測度に基づいて、新しい変数名および記述変数ラベルが SPSS により自動的に生成されます。要約表には、元の変数、新しい変数および変数ラベルがリストアップされます。

オプションとして、次の選択が可能です。

- ケースを昇順または降順でランク付けする。
- 変数の一覧から 1 つ以上のグルーピング変数を選択して、[グループ] ボックスに移動することにより、ランクをサブグループに編成する。ランクは各グループ内で計算されます。グループは、グループ化変数の値の組み合わせにより

定義されます。たとえば、GENDER と MINORITY をグループ化変数として選択すると、GENDER と MINORITY の組み合せにごとにランクが計算されます。

図 8-12
[ケースのランク付け] ダイアログ ボックス



ケースをランク付けするには

▶ メニューから次の項目を選択します。

変換
ケースのランク付け...

▶ 変数の一覧からランク付けを行う 1 つ以上の変数を選択して、[変数] ボックスに移動します。ランク付けできるのは数値型変数だけです。

オプションとして、ケースを [ランク 1 の割当] にランク付けしたり、ランクをサブグループに編成したりできます。

ケースのランク付け: 手法の選択

複数のランク付けの方法を選択できます。各方法に対して別々のランク付け変数が作成されます。ランク付けの方法には、[順位]、[サベージ スコア]、[小数点つき順位]、[小数点つき順位パーセント]、[ケースの重み付けの合計]、[百分位] の各チェック ボックスがあります。また、[その他] をクリックすると、比率推定値および正規スコアに基づいてランク付けを生成できます。

順位 (ランク)。単純な順位。新しい変数の値は、その順位と同じです。

サベージ スコア。新しい変数は、指数分布に基づくサベージ スコアを含みます。

小数点つき順位。新しい変数の値は、非欠損ケースの重みの合計でランクを割ったものになります。

小数点つき順位パーセント。各ランクを、有効な値を持つケース数で割り、100 倍します。

ケースの重み付けの合計。新しい変数の値は、そのケースの重み付けの合計と同じです。新しい変数の値は、ケースの重みの合計と等しくなります。

百分位。ランクは、各グループがほぼ同じ数のケースを含むパーセンタイル グループに基づきます。たとえば、4 分位では、ランク 1 は 25 パーセンタイルより低いケースに割り当てられ、ランク 2 は 25 パーセンタイルと 50 パーセンタイルの間のケースに、ランク 3 は 50 パーセンタイルと 75 パーセンタイルの間のケースに、ランク 4 は 75 パーセンタイルより上のケースに割り当てられます。

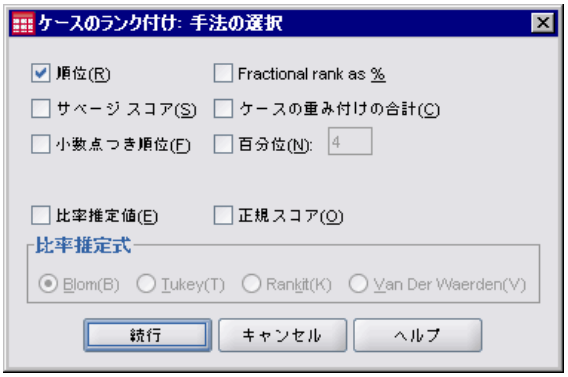
比率推定値。比率推定とは、特定のランクに対応する分布の累積比率を推定します。

正規スコア。正規スコアとは、推定された累積比率に対応する Z 得点のことです。

比率推定式。比率推定値および正規スコアの場合は、比率推定値の公式として [Blom]、[Tukey]、[Rankit]、または [Van Der Waerden] を選択できます。

- Blom。公式 (r–3/8) / (w+1/4) を使って、比率推定値に基づく新しい順位変数を作ります。この式で、r はランク、w はケースの重みの合計を示します。
- Tukey。公式 (r–1/3) / (w+1/3) を使用します。r はランクを、w はケースの重みの合計を表します。
- Rankit。公式 (r–1/2) / w を使用します。w は観測値の数、r はランクを表します。r は 1 から w までの範囲の値をとります。
- Van der Waerden。Van der Waerden 変換は公式 r/(w+1) で定義されます。w はケースの重みの合計を表し、r はランクで、1 から w までの範囲をとります。

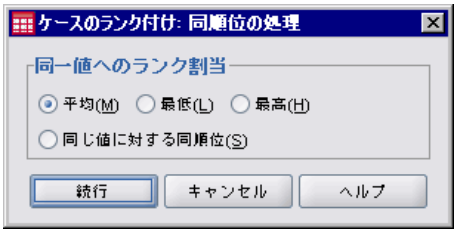
図 8–13
[ケースのランク付け: 手法の選択] ダイアログ ボックス



ケースのランク付け: 同順位の処理

このダイアログ ボックスでは、元の変数に同じ値を持つケースのランク付けを割り当てる方法を管理します。

図 8-14
[ケースのランク付け: 同順位の処理] ダイアログ ボックス



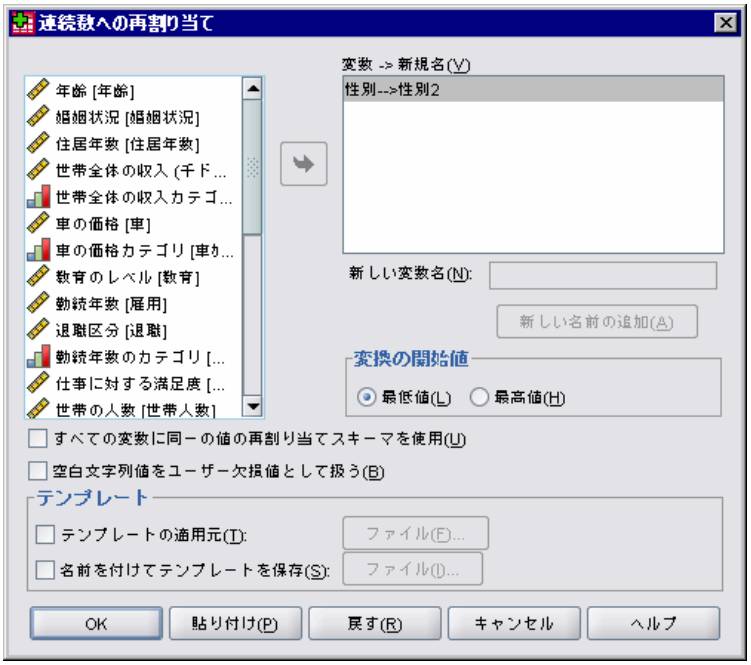
ここには、同順位の値にランクを割り当てるさまざまな方法が示されています。

値 (ピボットテーブル ノンパラメトリック検定)	平均	低	高	逐次 (MANOVA)
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

連続数への再割り当て

[連続数への再割り当て] ダイアログ ボックスでは、文字型値と数値型値を連続整数に変換できます。カテゴリ コードが連続していない場合は、実行後の空白セルにより動作が遅くなり、多くの SPSS 手続きのメモリー要件が増えます。さらに、手続きのなかには文字型変数が使用できないものもあり、また、因子の水準に連続した整数値が要求されることもあります。

図 8-15
[連続数への再割り当て] ダイアログ ボックス



- [連続数への再割り当て] ダイアログ ボックスにより作成される新しい変数には、今までの変数からの定義された変数および値ラベルがすべて保持されます。値ラベルの定義のない値の場合には、元の値がその再割り当て値のラベルとして使用されます。表には、今までの値ならびに新しい値および値ラベルが示されています。
- 文字型値はアルファベット順に再割り当てされ、大文字のものが小文字のものより先行します。
- 欠損値はいずれの非欠損値よりも高い値に再割り当てされますが、順序は保持されます。たとえば、元の変数が 10 個の非欠損値を持つとき、最小欠損値は 11 に再割り当てされ、値 11 は新しい変数の欠損値となります。

すべての変数に同一の値の再割り当てスキーマを使用。このオプションにより、1 つの自動再割り当てスキーマを、選択した変数すべてに適用できます。したがって、新しい変数すべてにおいて、一貫した割り当てスキーマが使用されます。

このオプションを選択した場合、次のルールと制限が適用されます。

- すべての変数は同じ型 (数値型または文字型) である必要があります。

- 選択した変数のすべての観測値を使用して、順序を並べ替えた値を作成し、連続する整数に再割り当てします。
- 新しい変数のユーザー欠損値は、定義されたユーザー欠損値を含むリストの最初の変数に基づいています。システム欠損値以外の、他の元の変数からのその他すべての値は、有効な値として処理されます。

空白の文字型値をユーザー欠損値として扱う。文字型変数で、空白値または NULL 値が **システム欠損値**として処理されません。このオプションを選択すると、空白文字列が自動的に、最も高い非欠損値よりも高い **ユーザー欠損**値に再割り当てされます。

テンプレート（テンプレート）

自動再割り当てスキーマをテンプレート ファイルに保存し、他の変数とデータファイルに割り当てることができます。

たとえば、ユーザーが毎月整数に自動再割り当てする、多数の英数字の製品コードがあるとします。ただし、新しい製品コードが追加されて、元の自動再割り当てスキーマを変更する月もあるとします。元のスキーマをテンプレートに保存し、一連の新しいコードを含む新しいデータに適用した場合、そのデータの新しいコードは、テンプレートの最後の値よりも高い値に自動的に再割り当てされます。これにより、元の製品コードに用意されている本来の自動再割り当てのスキーマが保存されます。

名前を付けてテンプレートを保存。選択された変数の自動再割り当てスキーマが、外部のテンプレート ファイルに保存されます。

- テンプレートには、元の欠損値以外の値を再割り当てされた値にマップする情報が含まれます。
- 欠損値以外の値の情報だけが、テンプレートに保存されます。ユーザー欠損値の情報は保存されません。
- 再割り当てする複数の変数を選択しても、同じ自動再割り当てスキーマをすべての変数には使用しないように選択した場合、または自動再割り当ての一環として既存のテンプレートを適用しない場合は、テンプレートはリストの最初の変数に基づくことになります。
- 再割り当てする変数を複数選択し、[**すべての変数に同一の値の再割り当てスキーマを使用**] または [**テンプレートを適用**] の一方、あるいは両方を選択した場合は、すべての変数について結合された自動再割り当てスキーマがテンプレートに含まれます。

テンプレートの適用元。以前保存された自動再割り当てテンプレートを、再割り当て用に選択された変数に割り当てます。その際、変数で見つかった追加の値をスキーマの末尾に追加し、保存されたスキーマに格納されている元の値と自動再割り当てされた値の関係を保持します。

- 自動再割り当て用に選択されたすべての変数は、同じ型（数値型または文字型）である必要があり、その型はテンプレートで定義された型に一致する必要があります。
- テンプレートには、ユーザー欠損値の情報は含まれません。対象となる変数のユーザー欠損値は、定義されたユーザー欠損値を含む元の変数リストの最初の変数に基づいています。システム欠損値以外の、他の元の変数からのその他すべての値は、有効な値として処理されます。
- テンプレートからの値のマッピングは、最初に適用されます。残りのすべての値は、テンプレートの最後の値よりも高い値に再割り当てされます。その際、ユーザー欠損値は、定義済みのユーザー欠損値を含むリストの最初の変数に基づいて、最後の有効な値よりも高い値に再割り当てされます。
- 自動再割り当て用に複数の変数を選択した場合、最初にテンプレートが適用され、次に、選択された変数で見つかった追加対象のすべての値に対し、共通の自動再割り当てを行います。その結果、選択されたすべての変数に、単一かつ共通の自動再割り当てスキーマが使用されることになります。

文字型値または数値型値を連続整数に再割り当てするには

▶ メニューから次の項目を選択します。

変換
連続数への再割り当て...

▶ 変数の一覧から再割り当てを行う 1 つ以上の変数を選択します。

▶ 選択した変数ごとに、新しい変数の名前を入力し、**[新しい変数名]** をクリックします。

日付と時刻ウィザード

日付と時刻ウィザードにより、日付変数や時刻変数に関連する多くの共通タスクが単純になります。

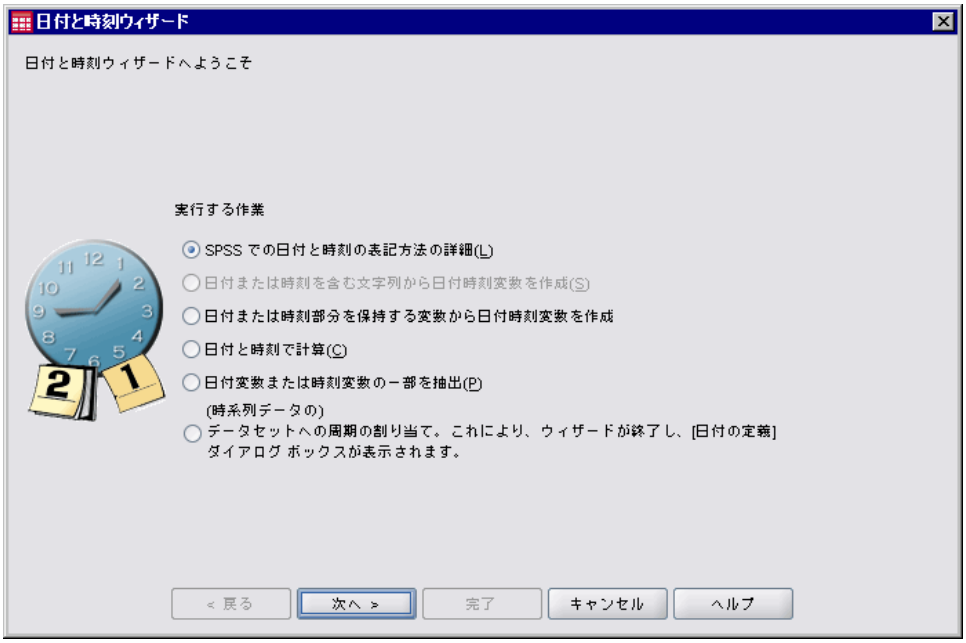
日付と時刻ウィザードを使うには

▶ メニューから次の項目を選択します。

変換
日付と時刻ウィザード...

▶ 実行するタスクを選択して、タスクを定義する手順に従います。

図 8-16
日付と時刻ウィザードの開始画面



- **日付と時刻の表記方法の詳細。**これを選択すると、SPSS での日付時刻変数の簡単な概要が画面に表示されます。[**ヘルプ**] をクリックすると、さらに詳しい情報を表示することもできます。
- **日付または時刻を含む文字列から日付時刻変数を作成。**このオプションを使用して、文字列変数から日付時刻変数を作成します。たとえば、mm/dd/yyyy 形式で日付を表す文字列変数があり、この変数から日付時刻変数を作成する場合などです。
- **日付または時刻の部分を保持する変数から日付時刻変数を作成。**これを選択すると、既存の変数のグループから日付時刻変数を作成できます。たとえば、月 (整数) を表す変数、その月の日を表す変数、年を表す変数があるとします。これら 3 つの変数を 1 つの日付時刻変数に組み合わせることができます。
- **日付と時刻で計算。**このオプションを使用して、日付時刻変数を加算または減算します。たとえば、プロセスの終了時刻を表す別の変数から、プロセスの開始時刻を表す変数を減算することによって、プロセスの実行時間を計算できます。
- **日付変数または時刻変数の一部を抽出。**このオプションにより、日付変数または時刻変数の一部を抽出できます。たとえば、mm/dd/yyyy という形式で表される日付時刻変数から日の部分を取り出すことができます。
- **データセットへの周期の割り当て。**これを選択すると、[日付の定義] ダイアログ ボックスが表示されます。このダイアログ ボックスを使用して、一連の連続日付で構成される日付時刻変数を作成します。通常、日付を時系列データに関連付けるためにこの機能が使用されます。

注: 実行に必要な変数型がデータセットに存在しないタスクは無効になります。たとえば、データセットに文字列変数が含まれていない場合は、文字列から日付時刻変数を作成するタスクは適用されず、無効になります。

SPSS での日付と時刻

SPSS では、日付と時刻を表す変数は、変数の型が数値型で、特定の日時書式に対応する書式で表示されます。通常、このような変数は日付時刻変数と呼ばれます。実際に日付を表す日付時刻変数と、20 時間 10 分 15 秒のように、日付とは無関係に時間の長さを表す変数は区別されます。後者は期間変数と呼ばれ、前者は日付変数または日付時刻変数と呼ばれます。すべての表示書式を列挙したリストについては、『Command Syntax Reference』の「Universals」にある「Date and Time」を参照してください。

日付変数と日付時刻変数。日付変数では、mm/dd/yyyy のように日付を表す書式が保持されます。日付時刻変数では、dd-mmm-yyyy hh:mm:ss のように日付と時刻を表す書式が保持されます。内部的に、ほとんどの日付書式変数は 1582 年 10 月 14 日からの秒数として保管されます。日付変数と日付時刻変数は、日付書式変数と呼ばれることもあります。

- 年は 2 桁で指定しても、4 桁で指定しても認識されます。デフォルトでは、年を表す 2 桁は、現在の日付の 69 年前から 30 年後までの範囲を表します。この範囲は、オプションの設定によって決まり、自由に変更することができます ([編集] メニューの **[オプション]** を選択し、**[データ]** タブをクリック)。
- day-month-year 書式では、ダッシュ、ピリオド、カンマ、スラッシュ、スペースを区切り記号として使用できます。
- 月は、数字、ローマ数字、3 文字の略語で示すこともでき、さらに略さずに綴ることもできます。3 文字の略語を使用する場合も、略さずに綴る場合も、月の名前は英語で表記する必要があります。他の言語の月名は認識されません。

期間変数。期間変数では、hh:mm のように時間の長さを表す書式が保持されます。期間変数は、内部的には特定の日付を参照しない秒数で格納されます。

- 時間指定 (日付時刻変数と期間変数に当てはまります) では、時、分、秒の区切り記号としてコロンを使用できます。時と分は必須ですが、秒は省略できます。小数部を含む秒数で、秒と小数部を区切るにはピリオドを使用します。時の指定には上限はありませんが、分の最大値は 59、秒の最大値は 59.999... になります。

現在の日付と時刻。システム変数 **$TIME** に現在の日付と時刻が保持されます。このシステム変数では、システム変数を使用する変換コマンドが実行される日時が、1582 年 10 月 14 日からの通算秒数で表されます。

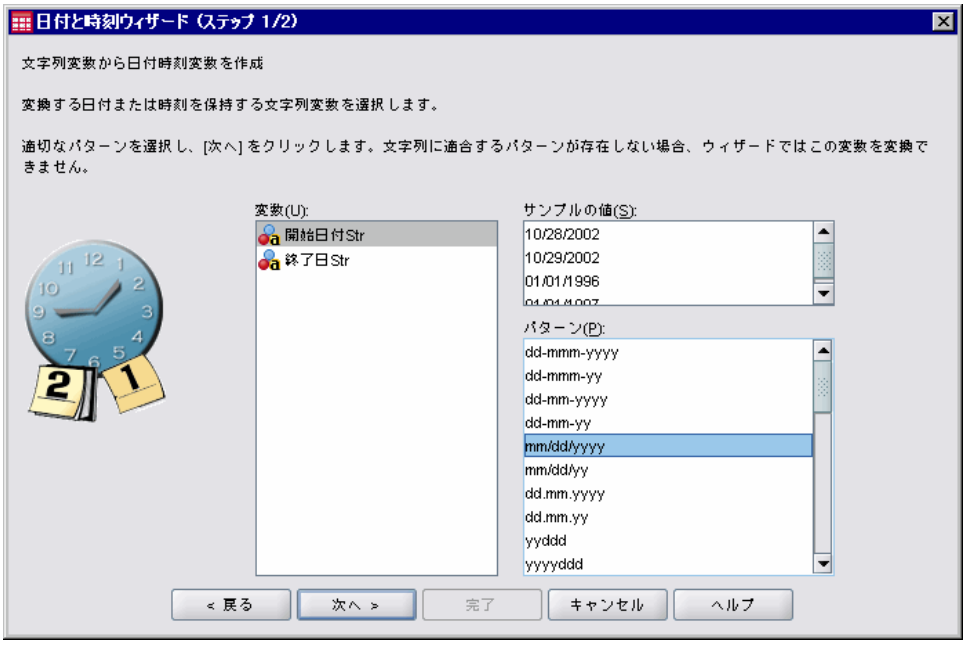
文字列からの日付時刻変数の作成

文字列変数から日付時刻変数を作成するには、次の手順を実行します。

- 日付と時刻ウィザードの開始画面で、[**日付または時刻を含む文字列から日付時刻変数を作成**] を選択します。

日付時刻変数に変換する文字列変数の選択

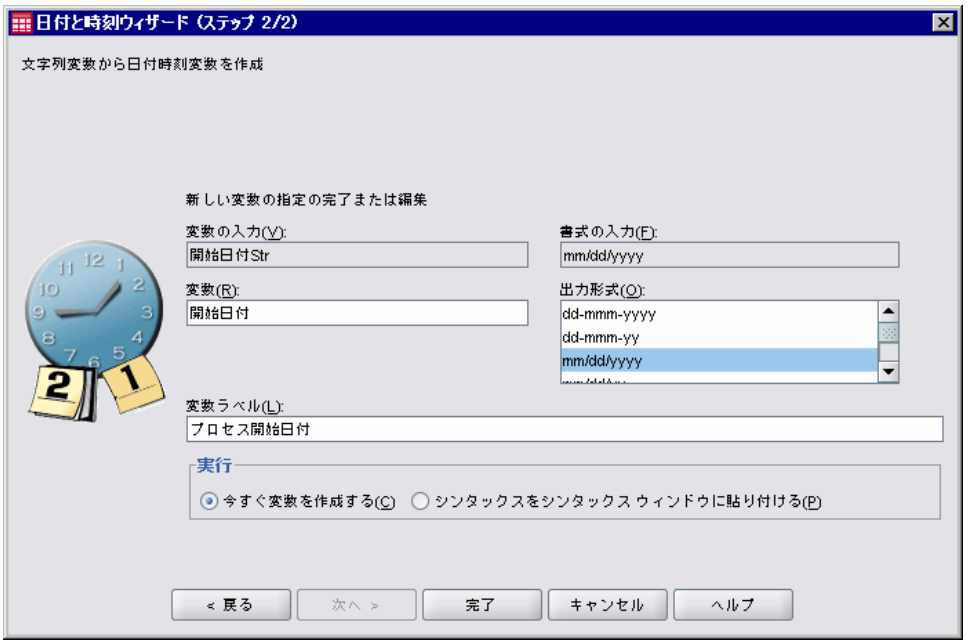
図 8-17
文字列からの日付時刻変数の作成 - ステップ 1



- [変数] リストで変換元の文字列変数を選択します。リストには文字列変数だけが表示されます。
- 文字列変数で日付が表されている方法と一致するパターンを [パターン] リストから選択します。[サンプルの値] リストには、データ ファイルで選択した変数の実際の値が表示されます。文字列変数の値が選択したパターンに適合しないと、新しい変数の値はシステム欠損値になります。

文字列変数から日付時刻変数への変換結果の指定

図 8-18
文字列からの日付時刻変数の作成 - ステップ 2



- ▶ [変数] に名前を入力します。この名前を既存の変数名にすることはできません。

  オプションとして、次の選択が可能です。
  - ■ [出力形式] リストから新しい変数の日時の書式を選択する。
  - ■ 新しい変数にわかりやすい変数ラベルを割り当てる。

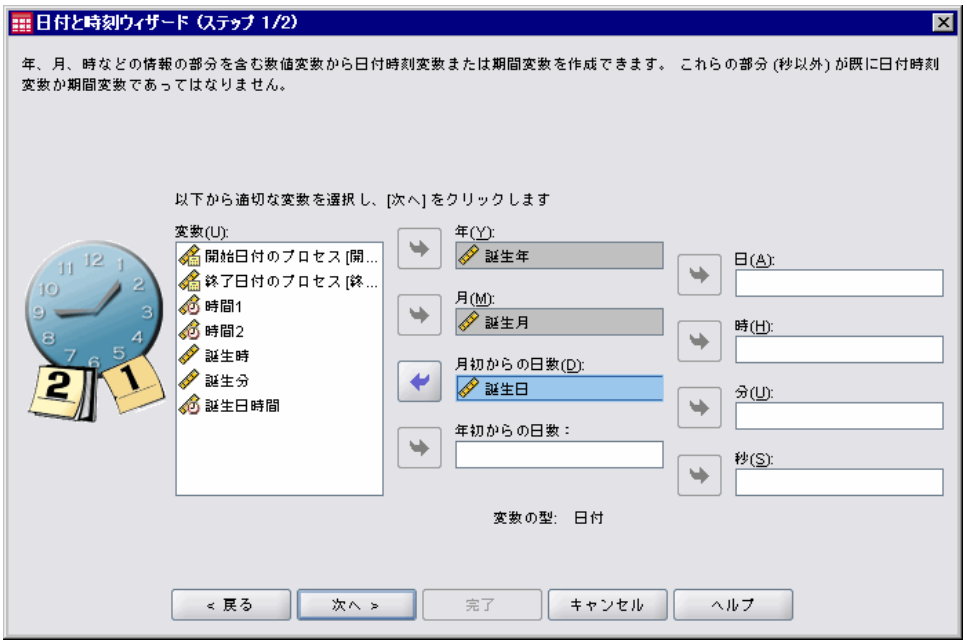
変数のグループからの日付時刻変数の作成

既存の変数のグループを 1 つの日付時刻変数に結合するには、次の手順を実行します。

- ▶ 日付と時刻ウィザードの開始画面で、**[日付または時刻部分を保持する変数から日付時刻変数を作成]** を選択します。

1 つの日付時刻変数に結合する変数の選択

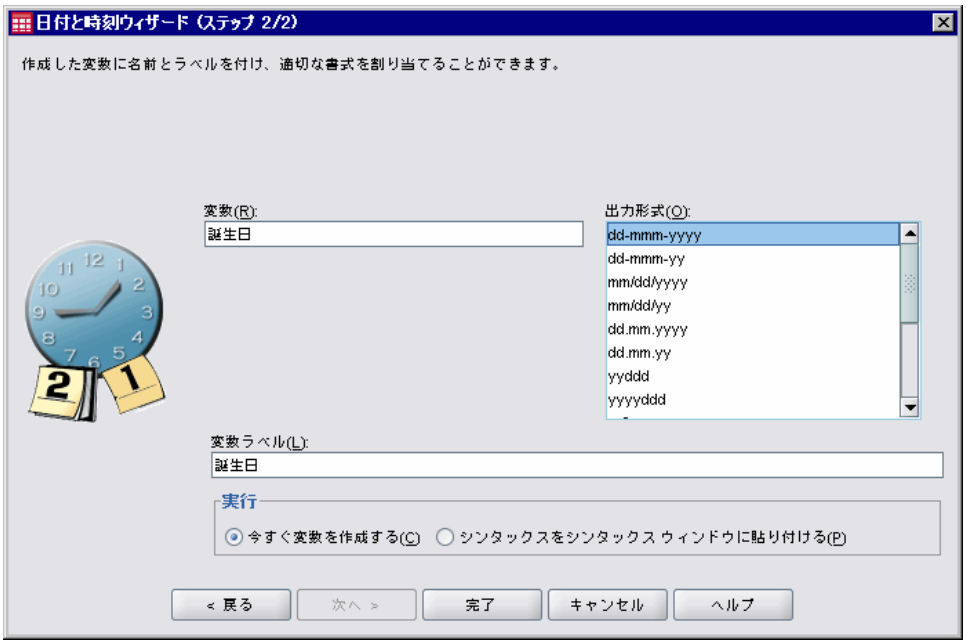
図 8-19
変数グループからの日付時刻変数の作成 - ステップ 1



- 日時のさまざまな部分を表す複数の変数を選択します。
  - 選択する変数として許可されない組み合わせもあります。たとえば、年と日から日付時刻変数を作成しようとすると、年を選択することによって日付を構成する部分がすべて必要になるので、このような組み合わせは無効になります。
  - 最終的に作成される日付時刻変数の部分の 1 つとして、既存の日付時刻変数を使用することはできません。新しい日付時刻変数の部分を構成する変数は、整数である必要があります。例外として、新しい変数の [秒] の部分に既存の日付時刻変数を使用することが許可されています。小数部分を含む秒が許可されているので、[秒] に使用する変数は整数である必要はありません。
  - 新しい変数の一部になる変数の値が許容範囲内に収まっていないと、新しい変数の値はシステム欠損値になります。たとえば、間違って [月] に日を表す変数を使用したとします。SPSS では月の有効範囲は 1 ～ 13 なので、日の値が 14 ～ 31 の場合は、新しい変数にシステム欠損値が割り当てられることになります。

複数の変数を結合して作成する日付時刻変数の指定

図 8-20
変数グループからの日付時刻変数の作成 – ステップ 2



- ▶ [変数] に名前を入力します。この名前を既存の変数名にすることはできません。
- ▶ [出力形式] リストから日時の書式を選択します。

  オプションとして、次の選択が可能です。
  - ■ 新しい変数にわかりやすい変数ラベルを割り当てる。

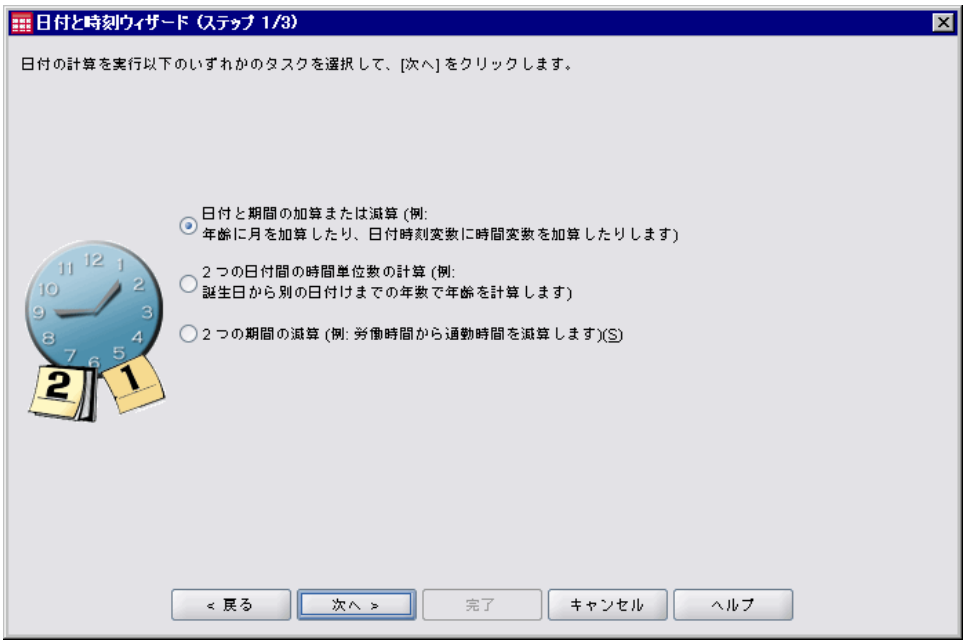
日付時刻変数の加算または減算

日付時刻変数を加算または減算するには、次の手順を実行します。

- ▶ 日付と時刻ウィザードの開始画面で、**[日付と時刻で計算]** を選択します。

日付時刻変数を使って実行する計算の種類の選択

図 8-21
日付時刻変数と値の加算または減算 – ステップ 1



- **日付と期間の加算または減算。**このオプションを使用して、日付書式変数との加算または減算を行います。10 日のような固定値の期間、または年を表す変数のような数値変数の値を加算または減算できます。
- **2 つの日付間の時間単位数の計算。**このオプションを使用して、2 つの日付の差を、選択した単位で計算して取得します。たとえば、2 つの日付の差を年数でも、日数でも取得できます。
- **2 つの期間の減算。**このオプションを使用して、hh:mm や hh:mm:ss のような期間の書式を持つ 2 つの変数の差を取得します。

注: 実行に必要な変数型がデータセットに存在しないタスクは無効になります。たとえば、データセットに期間の書式を持つ 2 つの変数が含まれていない場合は、2 つの期間を減算するタスクは適用されず、無効になります。

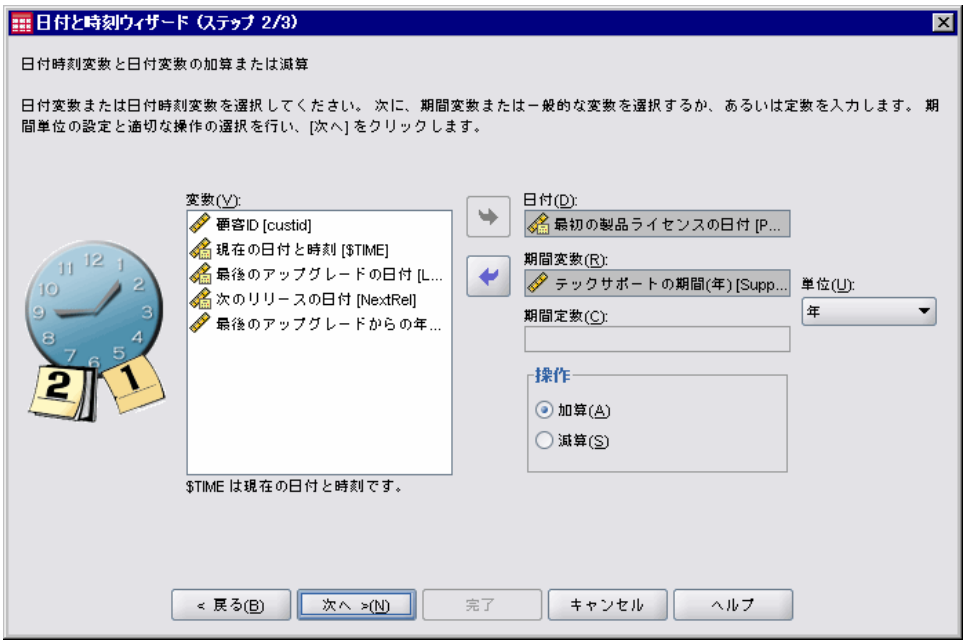
日付と期間の加算または減算

日付書式変数と期間を加算または減算するには、次の手順を実行します。

▶ 日付と時刻ウィザードの [**日付の計算を実行**]画面で、[**日付と期間の加算または減算**] を選択します。

加算または減算する日付時刻変数と期間の選択

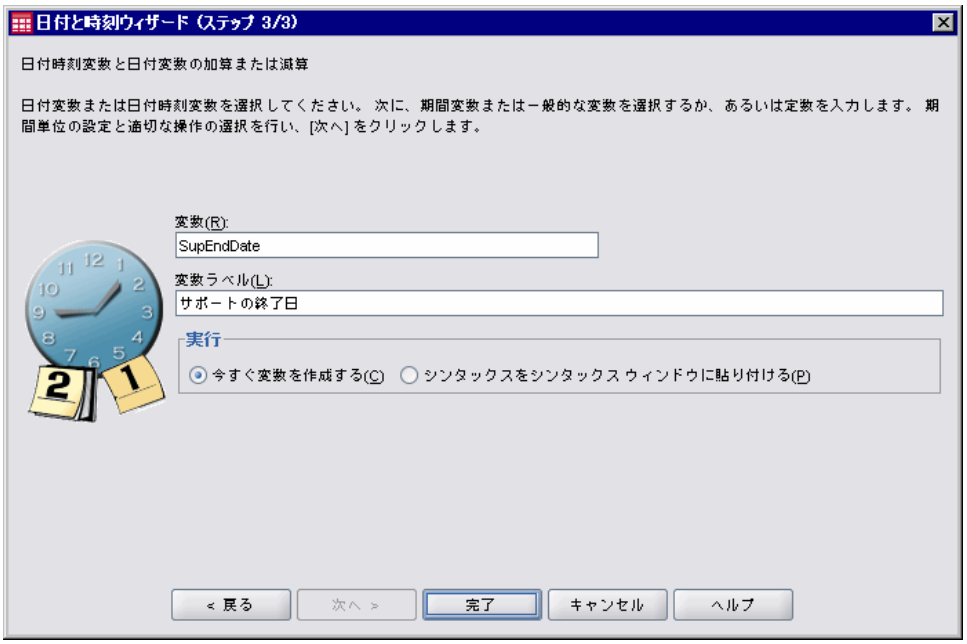
図 8-22
期間の加算または減算 - ステップ 2



- 日付 (または時刻) 変数を選択します。
- 期間変数を選択するか、[期間定数] に値を入力します。期間に使用する変数を日付変数または日付時刻変数にすることはできません。期間に使用する変数は、期間変数か単純な数値型変数にできます。
- ドロップダウン リストからその期間が表す単位を選択します。hh:mm や hh:mm:ss など期間を表す形式の変数を使用する場合は、**[期間]** を選択します。

日付時刻変数と期間の加算結果または減算結果の指定

図 8-23
期間の加算または減算 - ステップ 3



▶ [変数] に名前を入力します。この名前を既存の変数名にすることはできません。

オプションとして、次の選択が可能です。

- 新しい変数にわかりやすい変数ラベルを割り当てる。

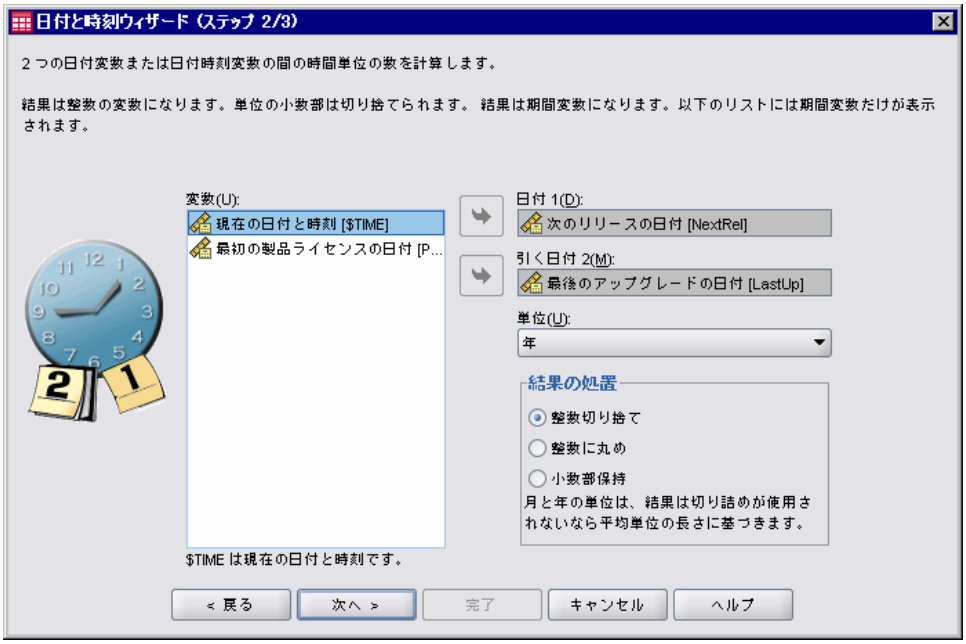
日付書式変数の減算

2 つの日付書式変数を減算するには、次の手順を実行します。

▶ 日付と時刻ウィザードの [**日付の計算を実行**] 画面で、[**2 つの日付間の時間単位数の計算**] を選択します。

減算する日付書式変数の選択

図 8-24
日付の減算 - ステップ 2



- ▶ 減算する変数を選択します。
- ▶ ドロップダウン リストから減算結果の単位を選択します。
- ▶ 結果が計算される方法を選択します (結果処置)。

結果の処置

結果の計算では次のオプションが利用できます。

- **整数に切り捨て。**小数点以下は無視します。たとえば、10/21/2007 から 10/28/2006 を引くと結果は年数は 0 で、月数は 11 です。
- **整数に丸め。**結果は最も近い整数に丸めます。たとえば、10/21/2007 から 10/28/2006 を引くと結果は年数は 1 で、月数は 12 です。
- **小数点以下を保持。**完全な値が保持され、切り捨てまたは丸めは適用されません。たとえば、10/21/2007 から 10/28/2006 を引くと結果は年数は 0.98 で、月数は 11.76 です。

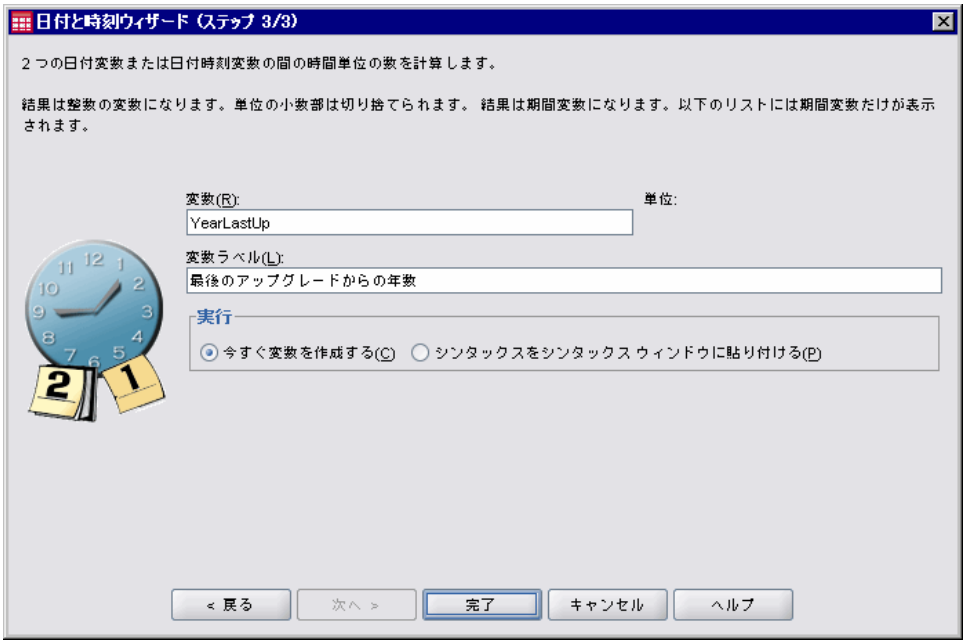
小数点の保持と丸めに関しては、年数の結果は (365.25) という 1 年の平均日数に準拠し、月数の結果は (30.4375) という 1 月の平均日数に準拠します。たとえば、(m/d/y 形式) の 3/1/2007 から 2/1/2007 を差し引くと、0.92 月が結果で、2/1/2007 から 3/1/2007 を差し引くと、結果は 1.02 月になります。このことは、うるう年を含む時間スパンでの計算された変数に影響を与えます。たとえ

ば、3/1/2008 から 2/1/2008 を差し引くと差は 0.95 月になり、うるう年でない同じ時間スパンでは 0.92 になります。

日付 1	日付 2	年			月		
		切り捨て	丸め	割合	切り捨て	丸め	割合
10/21/2006	10/28/2007	1	1	1.02	12	12	12.22
10/28/2006	10/21/2007	0	1	.98	11	12	11.76
2/1/2007	3/1/2007	0	0	.08	1	1	.92
2/1/2008	3/1/2008	0	0	.08	1	1	.95
3/1/2007	4/1/2007	0	0	.08	1	1	1.02
4/1/2007	5/1/2007	0	0	.08	1	1	.99

2 つの日付書式変数の減算結果の指定

図 8-25
日付の減算 - ステップ 3



▶ [変数] に名前を入力します。この名前を既存の変数名にすることはできません。

オプションとして、次の選択が可能です。

■ 新しい変数にわかりやすい変数ラベルを割り当てる。

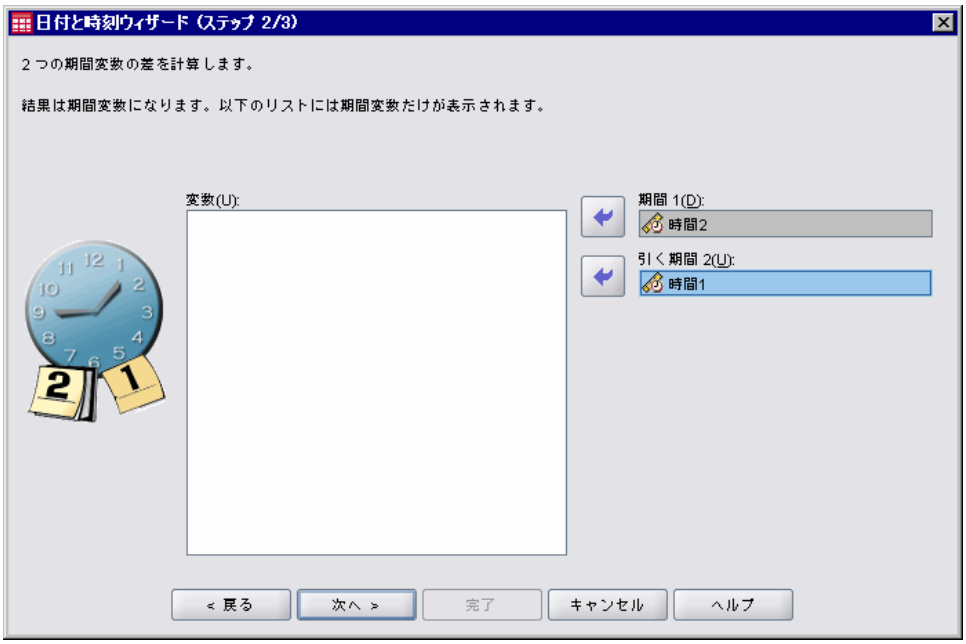
期間変数の減算

2 つの期間変数を減算するには、次の手順を実行します。

▶ 日付と時刻ウィザードの [**日付の計算を実行**] 画面で、[**2 つの期間の減算**] を選択します。

減算する期間変数の選択

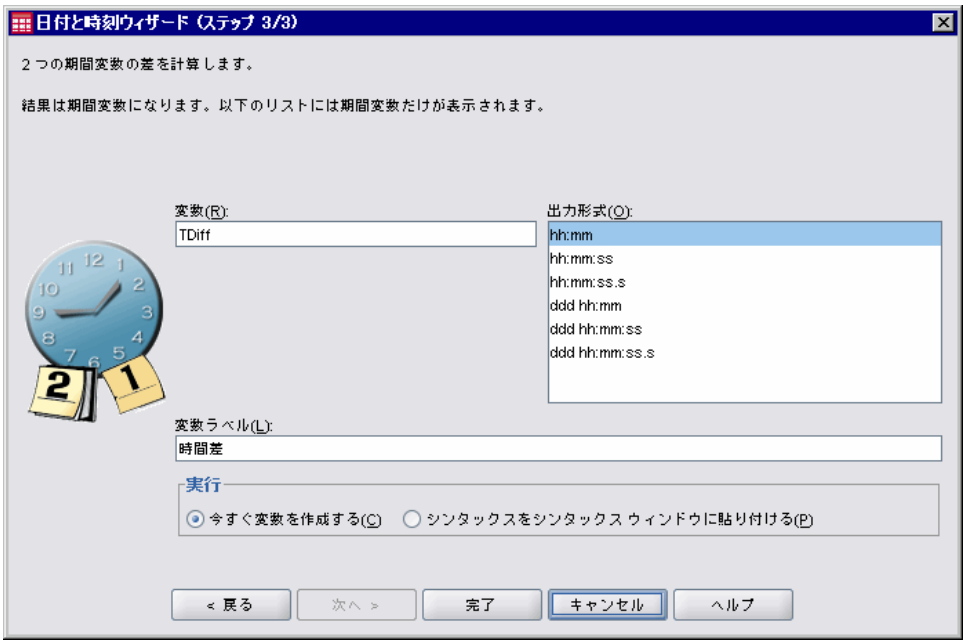
図 8-26
2 つの期間の減算 – ステップ 2



▶ 減算する変数を選択します。

2 つの期間変数の減算結果の指定

図 8-27
3 つの期間の減算 – ステップ 2



- ▶ [変数] に名前を入力します。この名前を既存の変数名にすることはできません。
- ▶ [出力形式] リストから期間の書式を選択します。

  オプションとして、次の選択が可能です。
  - ■ 新しい変数にわかりやすい変数ラベルを割り当てる。

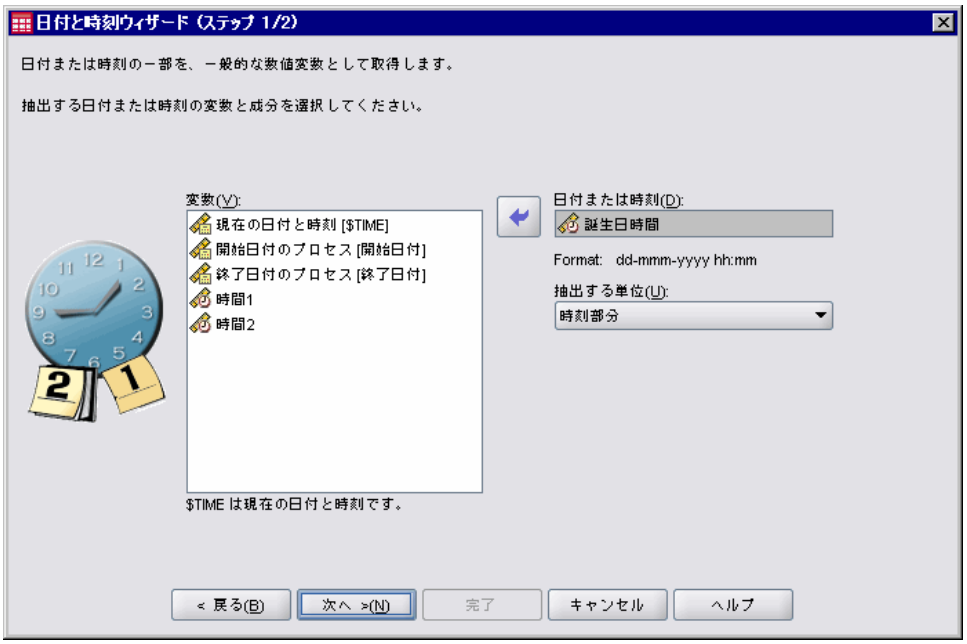
日付時刻変数の一部の抽出

日付時刻変数から年などの成分を抽出するには、次の手順を実行します。

- ▶ 日付と時刻ウィザードの開始画面で、**[日付変数または時刻変数の一部を抽出]** を選択します。

日付時刻変数から抽出する成分の選択

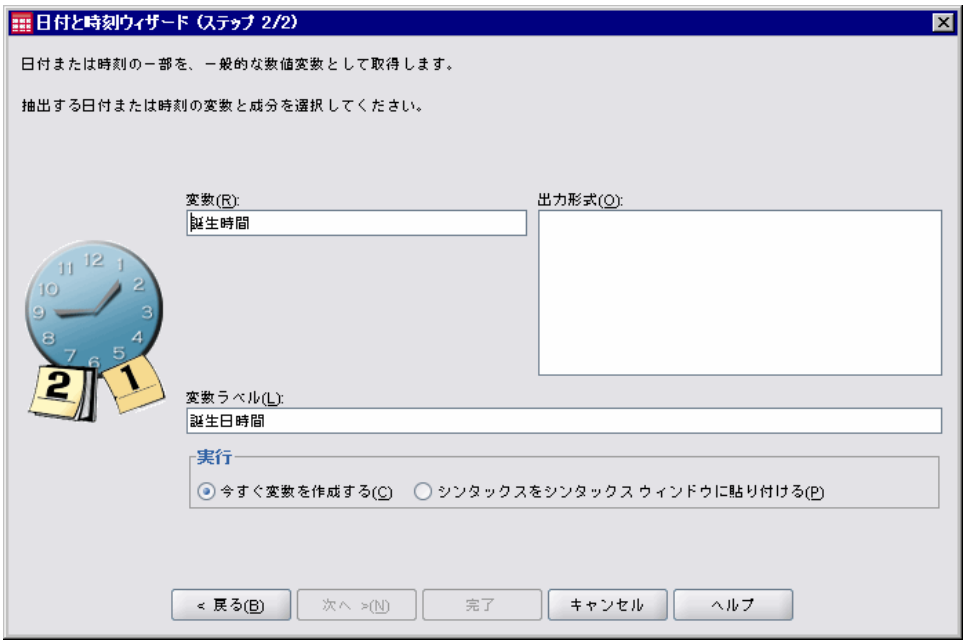
図 8-28
日付時刻変数の一部を取得 - ステップ 1



- 抽出する日付部分または時刻部分を含む変数を選択します。
- ドロップダウン リストから、抽出する変数の一部を選択します。曜日のように、表示日付の明確な一部ではない情報を日付から抽出できます。

日付時刻変数から抽出する成分の結果の指定

図 8-29
日付時刻変数の一部を取得 - ステップ 2



- ▶ [変数] に名前を入力します。この名前を既存の変数名にすることはできません。
- ▶ 日付時刻変数の日付部分または時刻部分を抽出している場合は、[出力形式] リストから書式を選択する必要があります。出力の書式が必要ない状況では、[出力形式] リストは無効になります。

  オプションとして、次の選択が可能です。
  - ■ 新しい変数にわかりやすい変数ラベルを割り当てる。

時系列データの変換

SPSS は、時系列分析に有用なデータ変換を複数提供しています。

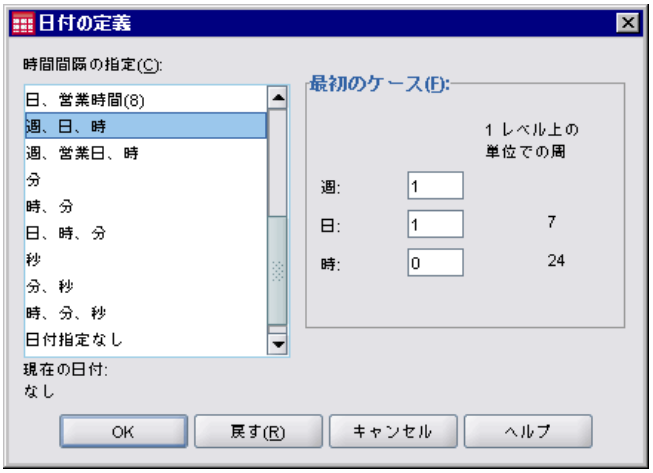
- ■ データ変数を生成して周期を確立し、履歴期間、検証期間、および予測期間を区別します。
- ■ 新しい時系列変数を既存の時系列変数の関数として作成します。
- ■ システム欠損値およびユーザー欠損値を、複数ある方法の 1 つに基づいて推定値と置換します。

ある期間を通して定期的に変数（または変数の組）を測定していくことにより、1 つの **時系列**が得られます。時系列データ変換はデータ ファイル構造を仮定しており、その構造においてそれぞれのケース（行）が他の時間における一組の観測値を表し、またケース間の時間長が一様となっています。

日付の定義

［日付の定義］ダイアログ ボックスでは、**時系列**の周期の指定や、時系列分析の結果に基づくラベル付けに使用できる日付変数を設定できます。

図 8-30
［日付の定義］ダイアログ ボックス



時間間隔の指定。日付を生成するときに使用する時間間隔を定義します。

- ［**日付指定なし**］を選択すると、以前に定義された日付変数がすべて削除されます。year_、quarter_、month_、week_、day_、hour_、minute_、second_、date_ のいずれかの名前を持つ変数はすべて削除されます。
- ［**ユーザー指定**］は、コマンド シンタックスで作成されたユーザー指定の日付変数（実働週 4 日など）が存在することを示します。この項目は、単にアクティブなデータセットの現在の状態を反映しているだけです。リストからそれを選択しても効果はありません。

最初のケース。最初のケースに割り当てられる開始日付を定義します。その後のケースには、時間間隔に基づいた連続値が割り当てられます。

1 レベル上の単位での周期。1 年の月数や週の日数などの、反復周期の変化を示します。表示される値は、入力できる最大値を示します。

日付の定義に使用される各成分に対して、SPSS は新しい数値型変数を作成します。新しい変数名は最後に下線が付きます。記述文字型変数 date_ は、成分からも作成されます。たとえば、**週、日、時間**を選択すると、week_、day_、hour_、date_ という 4 つの変数が新たに作成されます。

日付変数の定義がすでに行われている場合は、既存の日付変数と同じ名前の新しい日付変数を定義すると、置換されます。

時系列データに日付を定義するには

- メニューから次の項目を選択します。
  データ
  日付の定義...
- [時間間隔の指定] ボックスの一覧から時間間隔を選択します。
- [最初のケース] に開始日付を定義する値を入力します。これにより、最初のケースに割り当てられる日付が決まります。

日付変数と日付書式変数

[日付の定義] ダイアログ ボックスで作成した日付変数は、データ エディタの [変数ビュー] で定義される日付書式変数と混同しないようにしてください。日付変数は、時系列データの周期の設定に使用されます。日付書式変数は、各種の日付や時間書式で表示される日付や時間を表します。日付変数は、ユーザー指定の開始点からの日数、週数、時間数などを表す単一整数です。内部的に、ほとんどの日付書式変数は 1582 年 10 月 14 日からの秒数として保管されます。

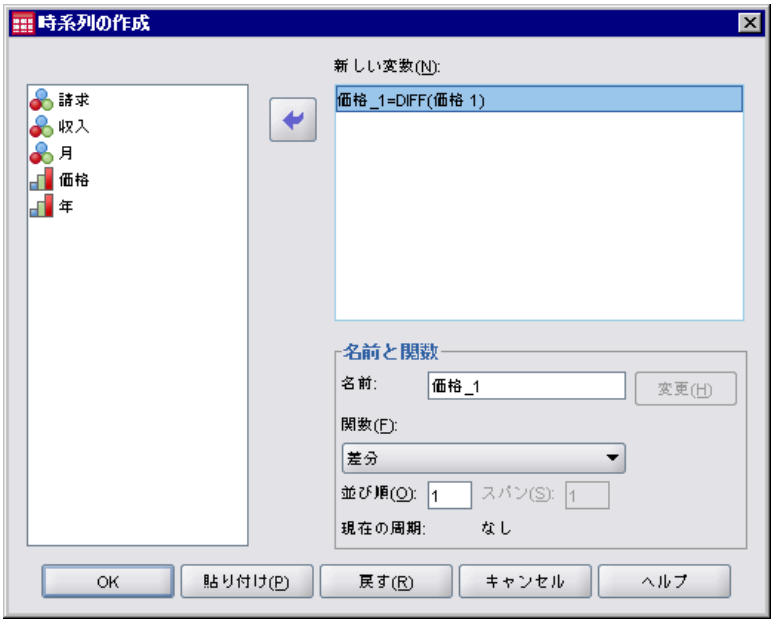
時系列の作成

[時系列の作成] ダイアログ ボックスでは、既存の数値型「**時系列**」変数の関数に基づいて新しい変数を作成できます。これらの値の変換は、多くの時系列分析を行う場合に便利です。

デフォルトの新規変数名は、その作成に使用されている既存の変数の最初の 6 文字で、下線および順番が後に付いています。たとえば、変数「**価格**」に対する新しい変数名は「**価格_1**」になります。新しい変数は、元の変数から定義された値ラベルを保持します。

時系列変数の作成には、差分、移動平均、移動中央値、ラグ、およびリード関数が利用できます。

図 8-31
[時系列の作成] ダイアログ ボックス



新しい時系列変数を作成するには

▶ メニューから次の項目を選択します。

変換
時系列の作成...

▶ 元の変数の変換に使用する時系列関数を選択します。

▶ 新しく作成する時系列変数の元となる変数を選択します。数値型変数だけが使用できます。

オプションとして、次の選択が可能です。

- デフォルトの新規変数名を上書きする変数名を入力する。
- 選択された変数の関数を変更する。

時系列変換関数

差分。系列内の連続した値間の非季節差分。次数は、差分の計算に使用された、前の値の数です。差分の次数のそれぞれに対して観測値が 1 つ失われるので、システム欠損値は系列の先頭に現れます。たとえば、差分次数が 2 の場合は、最初の 2 つのケースで、新しい変数にシステム欠損値が見られます。

季節差分。定数がスパンする系列値間の差分。スパンは、現在定義されている周期に基づいています。季節差分を計算するには、周期成分 (年間の月数など) を含む定義済みの日付変数 (データ メニュー、日付の定義) が必要となります。次

数は、差分の計算に使用される季節期間の数です。系列の先頭にあるシステム欠損値のあるケースの数は、周期値に次数を掛けたものと等しくなります。たとえば、現在の周期値が 12 で、次数が 2 の場合は、最初の 24 個のケースが新しい変数に対するシステム欠損値を持つことになります。

中心化移動平均。現在の値を含み、現在の値を囲んだ系列値のスパンの平均値。スパンは、平均値を計算するときに使用する系列値の数です。スパンが偶数のとき、移動平均は、各ペアの非中心化平均値を平均することにより計算されます。スパン *n* の系列の先頭および末尾にシステム欠損値を持つケース数は、スパン値が偶数の場合は *n* を 2 で割った、奇数の場合は(*n*–1) を 2 で割ったを商に等しくなります。たとえば、スパンが 5 の場合、その系列の先頭および最後にあるシステム欠損値のあるケース数は 2 です。

先行移動平均。現在の値に先行する系列値のスパンの平均値。スパンは、平均の計算に使用される先行系列値の数です。系列の先頭にあるシステム欠損値のあるケース数は、スパン値に等しくなります。

移動中央値。現在の値を含み、現在の値を囲んだ系列値のスパンの中央値。スパンは、中央値の計算に使用される系列値の数です。スパンが偶数のとき、中央値は、各ペアの非中心化中央値を平均することにより計算されます。スパン *n* の系列の先頭および末尾にシステム欠損値を持つケース数は、スパン値が偶数の場合は *n* を 2 で割った、奇数の場合は(*n*–1) を 2 で割ったを商に等しくなります。たとえば、スパンが 5 の場合、その系列の先頭および最後にあるシステム欠損値のあるケース数は 2 です。

累積集計。現在の値までの系列値の累積合計。

ラグ。指定したラグ次数に基づいた前のケースの値。次数は、値の入手元である現在のケースの前のケース数です。系列の先頭にあるシステム欠損値を持つケースの数は、次数値と等しくなります。

リード。指定したリード次数に基づいたその後のケースの値。次数は、値の入手元である現在のケースの後のケースの数です。系列の最後にあるシステム欠損値のあるケース数は、次数値と等しくなります。

平滑化。複合データ平滑法に基づく新しい規則手系列値。平滑法は、移動中央値 2 により中央化される移動中央値 4 で始まります。次に移動中央値 5、移動中央値 3、および移動重み付き平均値を適用することにより、これらの値を改めて平滑化します。残差は、平滑化系列を元の系列から差し引いて計算します。次に、計算結果の残差に対してこのプロセス全体を反復します。最後に、平滑化残差は、上記のプロセスを通して最初に得られた平滑化値を差し引いて計算します。この方法は **T4253H 平滑法**とも呼ばれます。

欠損値の置き換え

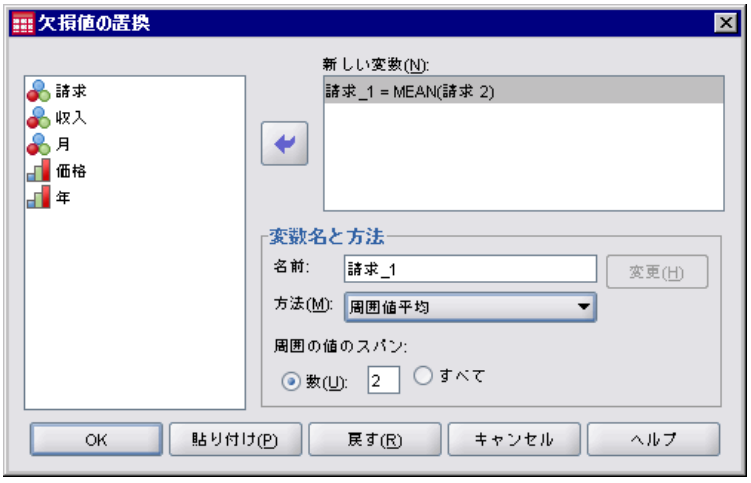
欠損観測値は分析上の問題児であり、時系列測定によっては、系列に欠損値があるときに計算できない場合があります。単純に、特定の観測値がわからない場合があります。さらに、欠損値が次のいずれかの結果になる場合があります。

- 差分の次数ごとに、系列の長さが 1 ずつ短くなる。
- 季節差分の次数ごとに、系列の長さが 1 季節ずつ短くなる。
- [保存] ボタンをクリックして適切な選択を行い、既存の系列の最終点以降に予測を含めた新しい系列を作成する場合、元の系列と生成された残差系列には、新しい観測の欠損値が含まれる。
- 一部の変換 (対数変換など) では、元の系列に含まれる特定の値の欠損データが作成される。

系列の先頭または最終点の欠損データにより特定の問題が発生することはありません。これらの欠損データにより、使用できる系列の長さが短くなるだけです。系列の途中にあるギャップ (**埋め込まれた**欠損データ) は、より深刻な問題となる可能性があります。問題の程度は、使用している分析手続きによって異なります。

[欠損値の置き換え] ダイアログ ボックスでは、既存の変数から新しい**時系列**変数を作成し、欠損値を、複数ある方法のいずれかによって計算された推定値に置き換えることができます。デフォルトの新規変数名は、その作成に使用されている既存の変数の最初の 6 文字で、下線および順番が後に付いています。たとえば、変数「**価格**」に対する新しい変数名は「**価格_1**」になります。新しい変数は、元の変数から定義された値ラベルを保持します。

図 8-32
[欠損値の置換] ダイアログ ボックス



時系列変数の欠損値を置き換えるには

▶ メニューから次の項目を選択します。

変換
欠損値の置き換え

▶ [変数名と方法] の [方法] ボックスの一覧から欠損値の置換に使用する推定方法を選択します。

▶ 変数の一覧から欠損値の置換を行う変数を選択します。

オプションとして、次の選択が可能です。

- デフォルトの新規変数名を上書きする変数名を入力する。
- 選択した変数に対する推定方法を変更する。

欠損値置き換えのための推定方法

系列平均。欠損値を系列全体の平均値と置換します。

周囲平均値。欠損値を、有効な周囲の値の平均値と置換します。周囲値のスパンは、その平均値の計算に使用された欠損値の上および下の有効値の数です。

周囲中央値。欠損値を有効な周囲の値の中央値と置換します。周囲値のスパンは、その中央値の計算に使用された欠損値の上および下の有効値の数です。

線型補間。線型補間を使って欠損値を置換します。欠損値の前の最後の有効値および欠損値の後の最初の有効値が補間に使用されます。系列の最初のケースまたは最後のケースに欠損値がある場合、欠損値は置換されません。

その点における線型トレンド。欠損値を、その点の線型トレンドと置換します。既存の系列は、1 から *n* に尺度化されたインデックス変数で回帰されます。欠損値はその予測値で置換されます。

予測モデルによるデータのスコアリング

予測モデルを一連のデータに適用するプロセスは、データの **スコアリング**と呼ばれます。SPSS、Clementine、および AnswerTree には、回帰モデル、クラスタ モデル、ツリー モデル、ニューラル ネットワーク モデルなどの予測モデルを構築するための手続きがあります。モデルが構築されると、モデル指定は、モデルの再構築に必要なすべての情報を含む XML ファイルとして保存できます。その後、SPSS サーバー製品では、XML モデル ファイルを読み取り、そのモデルをデータセットに適用することができます。

例。ローン申請書では、当該の申請者とローンのさまざまな側面を基にリスクが見積もられます。リスク モデルから取得した信用スコアは、ローン申請書を承認または拒否するのに使用されます。

スコアリングは、データの変換として扱われます。モデルは、予測した結果を得るために、特定の変数のグループ（モデルで指定された予測変数）に適用する一連の数値変換として内部的には表されます。この意味では、特定のモデルを使用したデータのスコアリングのプロセスは、本質的に、平方根関数のような任意の関数を一連のデータに適用するのと同じです。

スコアリングを利用できるのは SPSS サーバーだけです。ディストリビュート アナリシス モードで作業するユーザーが、インタラクティブにスコアリングを実行できます。大きいデータ ファイルのスコアリングを行う場合は、SPSS Batch Facility (SPSS サーバーに付属する実行可能な別バージョン) を利用することもできます。SPSS Batch Facility の使用方法の詳細は、『SPSS Batch Facility User's Guide』を参照してください。これは、SPSS サーバー製品の CD に PDF ファイルとして収録されています。

スコアリングでは、次の処理が行われます。

- XML (PMML) 形式のファイルからモデルを 読み込む。
- ApplyModel や StrApplyModel 関数 ([変数の計算] ダイアログ ボックス) を使用して、スコアを新しい変数として計算する。

`ApplyModel` または `StrApplyModel` 関数の詳細は、『Command Syntax Reference』の「Transformation Expressions」にある「Scoring Expressions」を参照してください。

次の表は、モデル指定を XML へエクスポートする場合に利用できる手続きの一覧です。エクスポートされたモデルは、前述のように、新しいデータをスコアリングするために SPSS サーバーで使用できます。SPSS サーバーでスコアリングができるすべてのモデルの一覧は、`ApplyModel` 関数の説明を参照してください。

手続き名	コマンド名	オプション
判別分析判別分析	`DISCRIMINANT`	Base
線型回帰	`REGRESSION`	Base
TwoStep クラスタ	`TWOSTEP CLUSTER`	Base
一般化線型モデル	`GENLIN`	Advanced Models
コンプレックス サンプルの一般線型モデル	`CSGLM`	Complex Samples
コンプレックス サンプルのロジスティック回帰	`CSLOGISTIC`	Complex Samples
コンプレックス サンプルの順序回帰	`CSORDINAL`	Complex Samples
ロジスティック回帰	`LOGISTIC REGRESSION`	Regression
多項ロジスティック回帰	`NOMREG`	Regression
分類ツリー	`TREE`	Tree

保存したモデルの読み込み

[モデルのロード] ダイアログ ボックスでは、XML (PMML) 形式で保存された予測モデルを読み込むことができます。この機能を利用できるのは、ディストリビュート アナリシス モードで作業中の場合だけです。モデルの読み込みは、モデルを使ったデータのスコアリングに必要な最初のステップです。

図 8-33
[モデルのロード] の出力

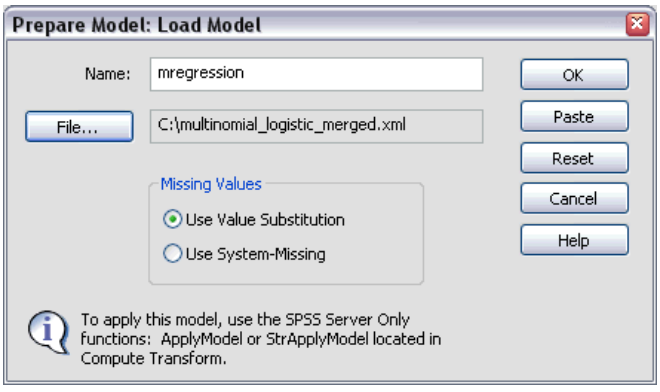
SPSS Variables	Model Variables					
	Name	Label	Type	Width	Role	Measurement
Age_Group	Age_Group	Age group	Numeric	8	Factor	Nominal
Frequency	Frequency	Number of purchases	Numeric	7	Covariate	Continuous
Gender	Gender	Gender	String	6	Factor	Nominal
Has_Broadband	Has_Broadband	Has home broadband connection	Numeric	10	Factor	Nominal
Has_Child	Has_Child	Has children	Numeric	10	Factor	Nominal
Income_Group	Income_Group	Income group	Numeric	10	Factor	Nominal
Log_Amount	Log_Amount	Log-transformed amount	Numeric	8	Covariate	Continuous
Recency	Recency	Days since last purchase	Numeric	10	Covariate	Continuous
Response	Response	Response	Numeric	10	Target	Nominal

モデルを読み込むには

▶ メニューから次の項目を選択します。

変換
モデルの準備
モデルのロード...

図 8-34
[モデルの準備: モデルのロード] ダイアログ ボックス



▶ このモデルに付ける名前を入力します。読み込むモデルには一意の名前を付ける必要があります。

▶ [ファイル] をクリックしてモデル ファイルを選択します。[ファイルを開く] ダイアログ ボックスが表示され、ディストリビュート アナリシス モードで利用できるファイルが

表示されます。このダイアログ ボックスには、SPSS サーバーがインストールされているマシン上のファイルと、ローカル コンピュータの共有フォルダや共有ドライブにあるファイルが表示されます。

注: データをスコアリングすると、モデル ファイルの変数と同じ名前を持つアクティブなデータセット内の変数に対してモデルが適用されます。コマンド シンタックスを使って、元のモデルの変数と、アクティブなデータセット内の別の変数をマップすることができます (MODEL HANDLE コマンドを参照してください)。

名前。このモデルを識別するための名前。有効なモデル名の規則は変数名 (85 ページの 5 章変数名を参照) の規則と同様ですが、最初の文字として $ を使用することができます。ApplyModel または StrApplyModel 関数を使用してデータをスコアリングするときに、この名前を使ってモデルを指定します。

ファイル。モデル指定が記述されている XML (PMML) ファイル。

欠損値 (変数の定義)

このオプションでは、スコアリングの実行中に、モデルで定義された予測変数の欠損値が見つかった場合の処理方法を制御します。スコアリングのコンテキストに欠損値があれば、次のいずれかが行なわれます。

- 予測変数には、値は含まれません。数値型変数であれば、システム欠損値になります。また、文字型変数であれば、null 文字列になります。
- 与えられた予測変数に対し、この値はモデル内ですでにユーザー欠損値として定義されています。モデル内ではなく、アクティブなデータセット内でユーザー欠損値として定義された値は、スコアリングのプロセスにおいては、欠損値としては処理されません。
- 予測変数はカテゴリ型であるのに対し、この値はモデル内で定義されているいずれのカテゴリでもありません。

値の置き換えを使用。スコアリングするケースに欠損値がある場合、置き換え値を使用します。欠損値が置き換えられる値の決定方法は、予測モデルの種類によって異なります。

- **SPSS モデル。**線型回帰モデルおよび判別モデルの独立変数の場合、モデルを構築して保存するときに、欠損値の置き換え値として平均値が指定されると、スコアリング計算時に見つかった欠損値の代わりにこの平均値が使用され、スコアリングが続行されます。平均値を使用できない場合は、システム欠損値が返されます。
- **AnswerTree モデル & TREE コマンド モデル。**CHAID モデルおよび Exhaustive CHAID モデルの場合、欠損分割変数に対して最大子ノードが選択されます。最大子ノードは、学習サンプル ケースを使用する子ノードのうち、最大の母集団を持つ子ノードです。C &RT モデルと QUEST モデルでは、(使用可能であれば) 代理分割変数が最初に使用されます。(代理分割変数とは、別の予測変数を使用することで、元の分割変数に最も近くなるようにする分割変

数です)。代理分割変数が指定されていない、またはすべての代理分割変数が欠損している場合は、最大子ノードが使用されます。

- **Clementine モデル。**線型回帰モデルは、SPSS モデルの説明と同様に処理されます。ロジスティック回帰モデルは、ロジスティック回帰モデルの説明と同様に処理されます。C &RT Tree モデルは、AnswerTree モデルの場合の C &RT モデルの説明と同様に処理されます。
- **ロジスティック回帰モデル。**ロジスティック回帰モデルの共変量に対し、保存されたモデルの一部として予測変数の平均値が含まれている場合、スコアリング計算時に見つかった欠損値の代わりにこの平均値が使用され、スコアリングが続行されます。予測変数がカテゴリ変数の場合 (たとえば、ロジスティック回帰モデルの因子)、または平均値を使用できない場合は、システム欠損値が返されます。

システム欠損値を使用。スコアリングするケースに欠損値がある場合、システム欠損値が返されます。

読み込まれたモデルのリストの表示

現在読み込み済みのモデルのリストを表示することができます。メニューから次の項目を選択します (ディストリビュート アナリシス モードでのみ選択可能)。

変換
モデルの準備
モデルの一覧表示...

モデル ハンドルの表が生成されます。この表には、現在読み込まれているすべてのモデルと、モデルに割り当てられた名前 (モデル ハンドルといいます)、モデルの種類、モデル ファイルのパス、および欠損値の処理方法が表示されます。

図 8-35
読み込まれたモデルのリスト

Model Handles

Name	Type	Input File Name	Validation handling	Missing handling
mregression	NOMREG	C:\models\mregression.xml	ON	SUBSTITUTE
tree	TREE	C:\models\tree.xml	ON	SUBSTITUTE

コマンド シンタックスで利用できる追加機能

[モデルのロード] ダイアログ ボックスから、選択内容をシンタックス ウィンドウに貼り付けて、MODEL HANDLE コマンド シンタックスを編集することができます。これにより、次の操作を行うことができます。

- 元のモデル内の変数と、アクティブなデータセット内の別の変数をマップする (MAP サブコマンドを使用)。デフォルトでは、モデル ファイルの変数と同じ名前を持つアクティブなデータセット内の変数に対してモデルが適用されます。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章 9

ファイルの操作とファイルの変換

データ ファイルは、特定の要求に合った理想的な形式で構成されているとは限りません。データ ファイルの結合、異なる順序でのデータの並べ替え、ケースのサブグループの選択、またはケースをグループ化することによる分析単位の変更などを行うことができます。次に示すものを含め、SPSS にはさまざまなファイル変換機能があります。

データの並べ替え。1 つ以上の変数値に基づいてケースの並べ替えができます。

変数とケースの入れ換え。SPSS は行をケースとして、列を変数として読み込みます。この順序が逆のデータ ファイルでは、行と列を入れ換えてデータを正しい書式で読み込むことができます。

ファイルの結合。2 つ以上のデータ ファイルを結合することができます。変数は同じでケースが違うファイル、またはケースは同じで変数が違うファイルを結合することができます。

ケースのサブグループの選択。分析をケースのサブグループに制限することもできますが、異なるサブグループに対して同時に分析を実行することもできます。

データのグループ集計。1 つ以上のグループ化変数に基づいてケースを集計することによって分析の単位を変更することができます。

データの重み付け。重み付け変数の値に基づいて分析用ケースに重みをつけることができます。

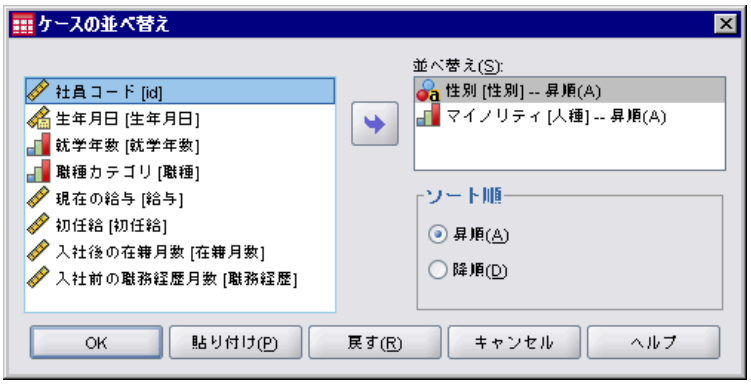
データの再構成。データを再編成して、複数のケースから 1 つのケース (レコード)、または 1 つのケースから複数のケースを作成できます。

ケースの並べ替え

このダイアログ ボックスで、1 つ以上の並べ替え変数の値に基づいてデータ ファイルのケース (行) を並べ替えます。ケースは、昇順または降順で並べ替えができます。

- 複数の並べ替え変数を選択すると、ケースは [並べ替え] ボックスのリスト直前の変数のカテゴリ内にある各変数別に並べ替えられます。たとえば、最初の並べ替え変数として**性別**を選択し、2 番目の並べ替え変数として **人種**を選択すると、ケースは、各性別カテゴリ内で人種の分類別に並べ替えられます。
- 並べ替えの順序は、ロケールで定義された順序に従います (文字コードの数値としての順序とは必ずしも一致しません)。デフォルトのロケールは、オペレーティング システムのロケールです。ロケールの設定は、[オプション] ダイアログ ボックス ([編集] メニュー) の [全般] タブにある [言語])で管理できます。

図 9-1
[ケースの並べ替え] ダイアログ ボックス



ケースを並べ替えるには

▶ メニューから次の項目を選択します。

データ
ケースの並べ替え...

▶ 1 つ以上の並べ替え変数を選択します。

変数の並べ替え

ユーザー指定の変数属性を含む変数属性 (たとえば、変数名、データ「型、尺度) の変数に基づいてアクティブ データセットの変数を並び替えできます。

- 値は、昇順または降順で並べ替えができます。
- ユーザー指定の変数属性の元の (事前に並べ替えておいた) 変数順序を保存できます。
- ユーザー指定の変数属性の値の並べ替えは、[変数ビュー] に現在表示されているユーザー指定の変数属性に制限されます。

ユーザー指定の属性の詳細は、ユーザー指定の変数属性を参照してください。

変数を並べ替えるには

データ エディタの [変数ビュー] の中で、

- 属性列見出しを右クリックし、コンテキスト メニューから **[昇順で並べ替え]** または **[降順で並べ替え]** を選択します。

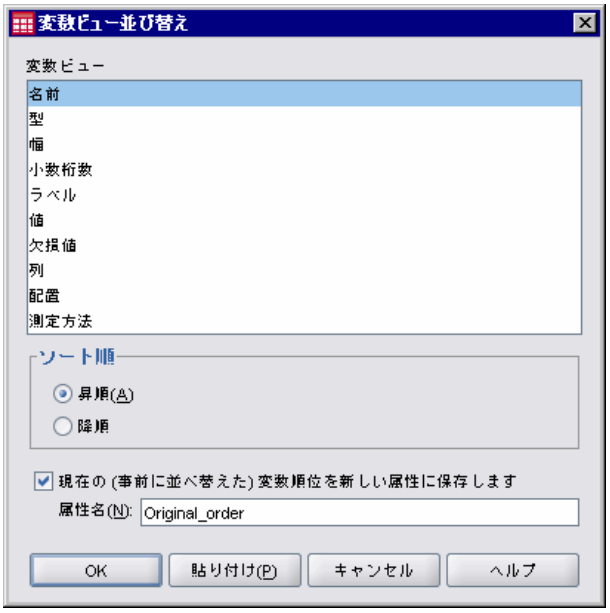
または

- 開いているウィンドウのメニューから、次の項目を選択します。
  データ
  変数の並べ替え...

- 変数を並べ替えするのに使用したい変数を選択しています。

- 並べ替え順序 (昇順または降順) を選択します。

図 9-2
[変数ビュー並び替え] ダイアログ



- 変数属性のリストはデータ エディタの [変数ビュー] に表示された属性列名と一致します。
- ユーザー指定の変数属性の元の (事前に並べ替えておいた) 変数順序を保存できます。各変数では、属性値は並べ替えより位置を示す整数です、したがってユーザー指定の属性の値に基づいて変数を並べ替えると元の変数順序を回復できます。

行と列の入れ替え

[行と列の入れ換え] では、ケース (行) が変数になり、変数 (列) がケースになるように、最初のデータ ファイルにある行と列が入れ換えられる新しいデータ ファイルを作成します。SPSS は、新しい変数名を自動的に作成してそのリストを表示します。

- 元の変数名を基にした新しい文字型変数 case_lbl が自動的に作成されます。
- 一意の値を持つ ID または名前変数がアクティブなデータセットに含まれる場合は、それを名前変数として使うことができ、その値は行と列を入れ換えたデータ ファイルで変数名として使用されます。値が数値変数の場合、変数名は *V* で始まりその後に数値が続きます。
- ユーザー欠損値は、行と列を入れ換えたデータ ファイルではシステム欠損値に変換されます。これらの値を保持するには、データ エディタの [変数ビュー] で欠損値の定義を変更します。

変数とケースを入れ換えるには

▶ メニューから次の項目を選択します。
データ
行と列の入れ換え...

▶ ケースに入れ換える 1 つ以上の変数を選択します。

データ ファイルの結合

SPSS を使うと、2 種類の方法で 2 つのファイルからデータを結合することができます。次のようにできます。

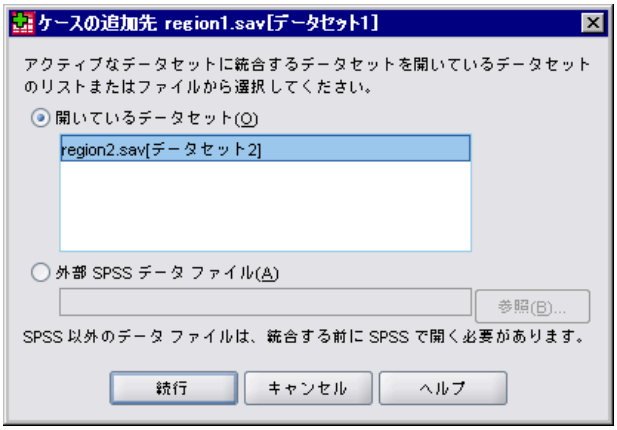
- 作業データセットを別のオープン データセットあるいは変数が同じでケースの異なる SPSS 形式のデータ ファイルと結合します。
- 作業データセットを別のオープン データセットあるいはケースが同じで変数の異なる SPSS 形式のデータ ファイルと結合します。

ファイルを結合するには

▶ メニューから次の項目を選択します。
データ
ファイルの結合

▶ [**ケースの追加**] または [**変数の追加**] を選択します。

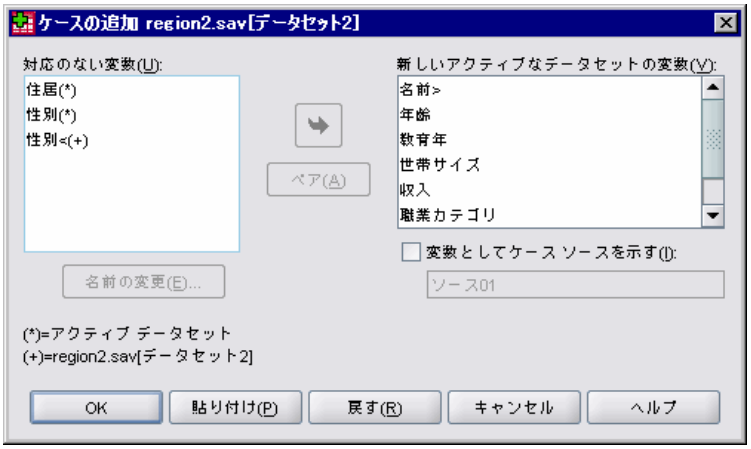
図 9-3
結合するファイルの選択



ケースの追加

[ケースの追加] により、作業データセットを 2 番目のデータセットあるいは変数 (列) が同じでケース (行) の異なる SPSS 形式のデータ ファイルと結合します。たとえば、2 か所の販売区域にいる顧客に対して同じ情報を記録して、そのデータを各地域で個別のファイルとして管理していたとします。2 番目のデータセットには、外部 SPSS 形式のデータ ファイルまたは現在のセッションで使用できるデータセットを選択することができます。

図 9-4
[ケースの追加] ダイアログ ボックス



対応のない変数。新しく結合されたデータ ファイルから除外する変数。作業データセットからの変数は、アスタリスク (*) で示されます。別のデータセットからの変数は、プラス記号 (+) で示されます。デフォルトでは、このリストには次のものが含まれます。

- 他のファイルにある変数名と一致しないデータ ファイルからの変数。対応のない変数からペアを作成してそれを新しく結合したファイルに含めることができます。
- 一方のファイルで数値データとして定義し、他方のファイルで文字型データとして定義した変数。数値型変数は、文字型変数と結合することはできません。
- 幅が等しくない文字型変数。文字型変数の幅は、両方のデータ ファイルで同じでなければなりません。

新しい作業データセットの変数。新しく結合したデータ ファイルに含める変数。デフォルトでは、名前とデータの型 (数値または文字型) の両方が一致する変数のすべてがリスト上に含められます。

- 結合したファイルに変数を含めたくない場合、それをリストから除外できます。
- 結合したファイル中の対応のない変数は、その変数を含まないファイルからケースに対して欠損データを持ちます。

変数としてケース ソースを示す。各ケースのソース データ ファイルを示します。この変数の値は、作業データセットからのケースの場合は 0 で、外部データ ファイルからのケースの場合は 1 です。

変数が同じでケースの異なるデータ ファイルを結合するには

▶ 結合するデータファイルを少なくとも 1 つ開きます。複数のデータセットが開いている場合、作業データセットに結合するデータセットを 1 つ作成します。このファイルのケースが新しく結合したデータ ファイルの最初に表示されます。

▶ メニューから次の項目を選択します。

データ
ファイルの結合
ケースの追加...

▶ データセットまたは SPSS 形式のデータ ファイルを選択して、作業データセットと結合します。

▶ [新しい作業データセットの変数] ボックスのリスト内に必要ない変数があれば除去します。

▶ 2 つのファイルに別の変数名で記録されているが同じ情報を表す変数のペアを [対応のない変数] ボックスのリストから追加します。たとえば、誕生日の変数名が、あるファイルでは **誕生日**に、別のファイルでは **生年月日**になっていることがあります。

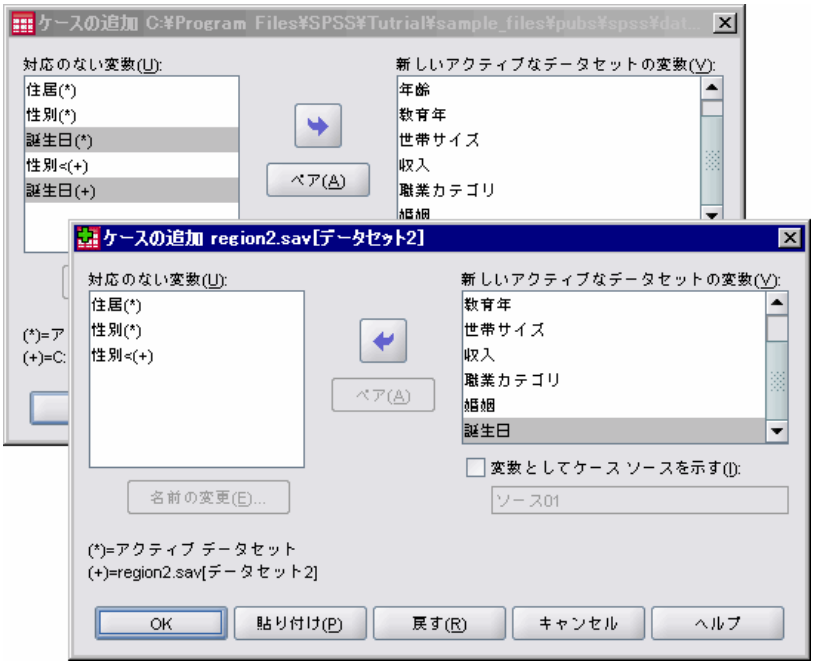
対応のない変数から変数のペアを選択するには

▶ [対応のない変数] ボックスのリスト内の変数の 1 つをクリックします。

▶ リストに載っているその他の変数を Ctrl+ クリックします。(Ctrl キーを押しながら、マウスの左ボタンをクリックします)。

▶ [ペア] をクリックして、変数のペアを [新しいアクティブなデータセットの変数] ボックスのリストに移動します。(作業データセットからの変数名が、結合したファイルの変数名として使用されます)。

図 9-5
Ctrl キーを押しながらクリックして変数のペアを選択



ケースの追加: 名前の変更

結合したデータ ファイルに含めるために [対応のない変数] ボックスから [新しい作業データ ファイルの変数] ボックスに変数を移動する前に、作業データセットまたは別のデータセットから変数の名前を変更することができます。変数の名前を変更すると、次を行うことができます。

- 変数のペアとして作業データセットからの名前ではなく別のデータセットからの変数名を使うことができます。
- 名前が同じでも型が一致しないかまたは文字幅が違う 2 つの変数を含めることができます。たとえば、アクティブなデータセットにある数値型変数の **性別** と、別のデータセットにある文字型変数の **性別** の両方を取り込む場合は、最初にどちらか一方の名前を変更しておく必要があります。

ケースの追加: 辞書情報

作業データセット内にある既存の辞書情報 (変数ラベルと値ラベル、ユーザー欠損値、表示書式) は、結合したデータ ファイルに適用されます。

- 変数の辞書情報が作業データセットで未定義の場合、別のデータセットから取り込んだ辞書情報が使用されます。
- 作業データセット内にある変数に対して値ラベルまたはユーザー欠損値が定義されている場合、別のデータセット内のその変数に対する追加の値ラベルまたはユーザー欠損値は無視されます。

3 つ以上のデータ ソースの結合

コマンド シンタックスを使うことで、50 個までのデータセット、データ ファイルのいずれかまたはその両方を結合することができます。詳細は、『Command Syntax Reference』([ヘルプ] メニューから利用可能) にある ADD FILES コマンドの項を参照してください。

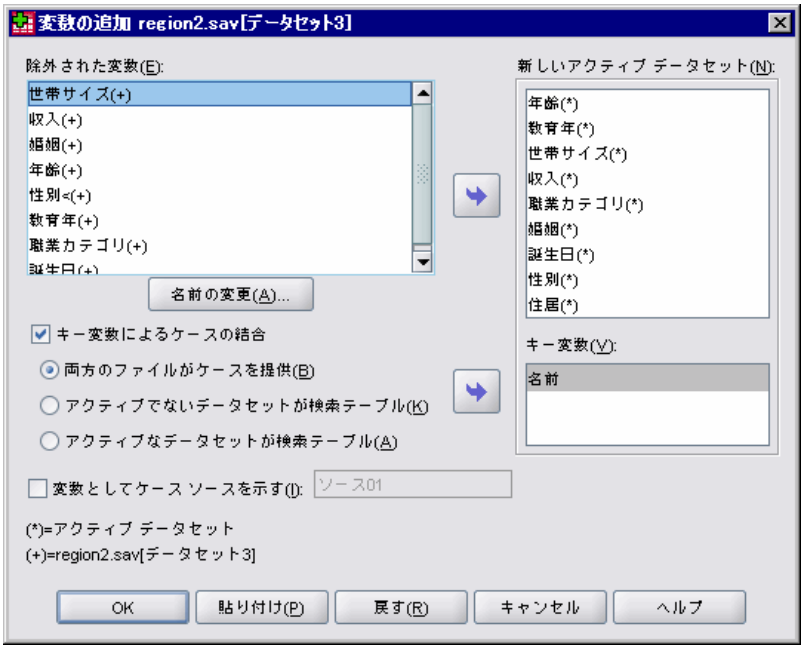
変数の追加

[変数の追加] により、作業データセットを別のオープン データセットあるいはケース (行) が同じで変数 (列) の異なる SPSS 形式のデータ ファイルと結合します。たとえば、検定前結果を含むデータ ファイルを検定後結果を含むデータ ファイルと結合することもできます。

- ケースは、両方のデータセットで同じ順序で並べ替えなければなりません。
- ケースを結合するために 1 つ以上のキー変数を使用する場合、キー変数の昇順で 2 つのデータセットを並べ替えなければなりません。
- 2 つ目のデータ ファイルにある変数名で、作業データセット内の変数名に重複するものは、デフォルトでは除外されます。SPSS ではこれらの変数には重複する情報が含まれていると仮定しているためです。

変数としてケース ソースを示す。各ケースのソース データ ファイルを示します。この変数の値は、作業データセットからのケースの場合は 0 で、外部データ ファイルからのケースの場合は 1 です。

図 9-6
[変数を追加] ダイアログ ボックス



除外された変数。新しく結合されたデータ ファイルから除外する変数。デフォルトでは、このリストには別のデータセットから取り込んだ変数名で作業データセット内の変数名と重複するものが含まれます。作業データセットからの変数は、アスタリスク (*) で示されます。別のデータセットからの変数は、プラス記号 (+) で示されます。名前が重複していて除外された変数を結合したファイルに取り込みたい場合は、名前を変更して取り込む変数のリストに追加します。

新しい作業データセット。新しく結合したデータセットに含める変数。デフォルトでは、両方のデータセットにある一意的なすべての変数名がリストに取り込まれます。

キー変数。一方のデータセット中に他方のデータセットに結合されるケースを持たないケースがいくつかある場合 (すなわち、一方のデータセットで欠損しているケースが複数ある場合)、2 つのデータセットから取り込んだケースを識別し正しく結合させるためにキー変数を使用します。キー変数をテーブル検索ファイルで使うこともできます。

- キー変数は両方のデータセットで同じ名前でなければなりません。
- 両方のデータセットはキー変数の昇順で並べ替える必要があり、さらに [キー変数] ボックスのリスト内の変数の順序はその並べ替え順と同じでなければなりません。
- キー変数で結合しないケースは結合したファイルに取り込まれますが、他のファイルからのケースとは結合しません。結合しないケースは取り出されるファイル内にある変数のみに対して値を持ち、他のファイルからの変数にはシステム欠損値を持ちます。

[非作業データセットが検索テーブル] または [作業データセットが検索テーブル]。検索テーブル (**テーブル ルックアップ ファイル**) では、各「ケース」のデータを他のデータ ファイルにある複数のケースに適用できます。たとえば、一方のファイルに家族の個人的な情報 (たとえば、性別、年齢、教育) が含まれていて、他方のファイルには家族の全体的な情報 (たとえば、総収入、家族の規模、住所) が含まれている場合、テーブル検索ファイルとして家族データのファイルを使い、結合したデータ ファイル内家族の各個人に共通の家族データを適用することもできます。

ケースが同じで変数の異なるファイルを結合するには

- 結合するデータファイルを少なくとも 1 つ開きます。複数のデータセットが開いている場合、作業データセットに結合するデータセットを 1 つ作成します。
- メニューから次の項目を選択します。
  データ
  ファイルの結合
  変数の追加...
- データセットまたは SPSS 形式のデータ ファイルを選択して、作業データセットと結合します。

キー変数を選択するには

- [除外された変数] ボックスのリスト内にある外部ファイルの変数 (+) から変数を選択します。
- **[キー変数によるケースの結合]** チェックボックスをオンにします。
- 変数を [キー変数] ボックスに追加します。

  キー変数は、作業データセットおよび別のデータセットの両方に含まれていなければなりません。両方のデータセットはキー変数の昇順で並べ替える必要があり、さらに [キー変数] ボックスのリスト内の変数の順序はその並べ替え順と同じでなければなりません。

変数の追加: 名前の変更

結合したデータ ファイルに含めるために [新しい作業データ ファイルの変数] ボックスに変数を移動する前に、作業データセットまたは別のデータセットから変数の名前を変更することができます。これは、2 つのファイル内の名前は同じだが異なる情報を持つ 2 つの変数を同時に取り込みたい場合に特に有効です。

3 つ以上のデータ ソースの結合

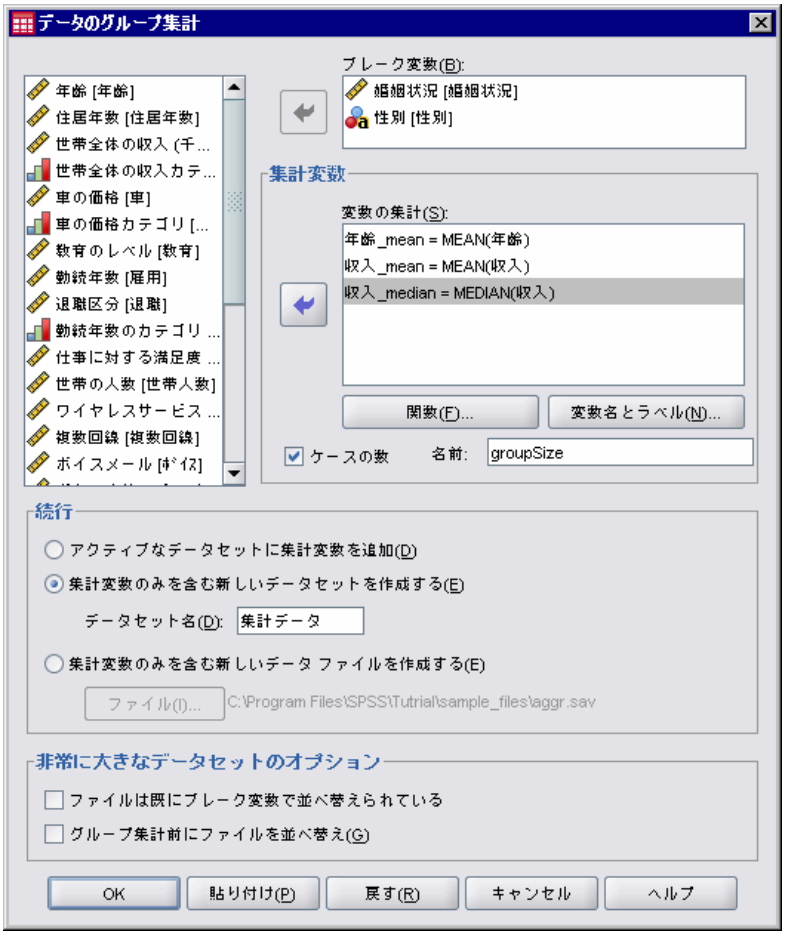
コマンド シンタックスを使うことで、50 個までのデータセット、データ ファイルのいずれかまたはその両方を結合することができます。詳細は、『Command Syntax Reference』([ヘルプ] メニューから利用可能) にある MATCH FILES コマンドの項を参照してください。

データのグループ集計

[データのグループ集計] では、アクティブなデータセット内のケースのグループが 1 つのケースに集計されます。また、新しく集計されたファイルが作成されたり、集計データを含むアクティブなデータセット内に新しい変数が作成されます。ケースは、1 つ以上のブレーク (グループ化) 変数の値に基づいて集計されます。

- 新しい集計データ ファイルを作成した場合、新しいデータ ファイルには、ブレーク変数で定義されたグループごとに 1 つのケースが含まれます。たとえば、2 つの値を持った 1 つのブレーク変数がある場合、新しいデータ ファイルには 2 つのケースしか含まれません。
- 集計変数をアクティブなデータセットに追加した場合、データ ファイル自体は集計されません。ブレーク変数と同じ値を持つ各ケースは新しい集計変数に対して同じ値を受け取ります。たとえば、**性別**が唯一のブレーク変数である場合、すべての男性は、平均年齢を表す新しい集計変数に対して同じ値を受け取ります。

図 9-7
[データのグループ集計] ダイアログ ボックス



ブレーク変数。ケースは、ブレーク変数の値に基づいてまとめてグループ分けされます。ブレーク変数値の一意的な各結合で 1 つのグループが定義されます。新しい集計データ ファイルの作成時、すべてのブレーク変数が、既存の名前と辞書情報を含む新しいファイルに保存されます。ブレーク変数には、数値型または文字型のいずれかを指定できます。

集計変数。ソース変数は、新しい集計変数を作成するために集計関数とともに使用されます。集計変数名の後には、オプションとして変数ラベル、集計関数の名前、およびかっこで囲んだソース変数名が続きます。

デフォルトの集計変数名を新しい変数名で上書きしたり、記述的な変数ラベルを付けたり、さらに集計データ値を計算するために使用する関数を変更することもできます。各ブレーク グループ内のケース数を持つ変数を作成することもできます。

データ ファイルを集計するには

▶ メニューから次の項目を選択します。
データ
グループ集計...

▶ 集計データを作成するため、ケースのグループ化方法を定義する 1 つ以上のブレーク変数を選択します。

▶ 1 つ以上の集計変数を選択します。

▶ [関数] をクリックし、各集計変数に集計関数を選択します。

集計結果の保存

集計変数をアクティブなデータセットに追加したり、新しく集計データ ファイルを作成することができます。

- **作業するデータファイルに集計変数を加える。**集計関数に基づく新しい変数は差作業するデータファイルに加えられます。データ ファイル自体は集計されません。ブレーク変数と同じ値を持つ各ケースは新しい集計変数に対して同じ値を受け取ります。
- **集計変数のみを含む新しいデータセットを作成する。**現在のセッションの新しいデータセットに集計データを保存します。そのデータセットには、集計されたケースを定義するブレーク変数や、集計関数によって定義されたすべての集計変数が含まれています。アクティブなデータセットは影響されません。
- **集計変数のみを含む新しいデータファイルを作成する。**外部データ ファイルに集計データを保存します。そのファイルには、集計されたケースを定義するブレーク変数や、集計関数によって定義されたすべての集計変数が含まれています。アクティブなデータセットは影響されません。

大きいデータ ファイルの並べ替えオプション

非常に大きいデータ ファイルの場合、事前に並べ替えておいたデータを集計するのがより効果的です。

ファイルは既にブレーク変数で並べ替えられている。データがブレーク変数値ですでにソートされている場合、このオプションを使うと、手続きはより少ないメモリで迅速に行われます。このオプションは慎重に使用してください。

- データは、ブレーク変数の値によって、[データのグループ集計] 手続きで指定したブレーク変数と同じ順序で並べ替える必要があります。
- 変数をアクティブなデータセットに追加している場合は、データをブレーク変数の値による昇順で並べ替えたときに限り、このオプションを選択します。

集計する前にファイルを並べ替えます。大きなデータファイルを持つ非常にまれなインスタンスでは、場合によっては、集計前にブレーク変数の値によってデータファイルを並び替える必要があります。メモリーまたはパフォーマンスに問題がない場合は、このオプションは推奨されません。

データのグループ集計: 集計関数の定義

このダイアログ ボックスは、[データのグループ集計] ダイアログ ボックス内にある [集計変数] ボックスのリスト内で選択した変数に、集計したデータ値を計算するときに使用する関数を指定します。集計関数には、次のものがあります。

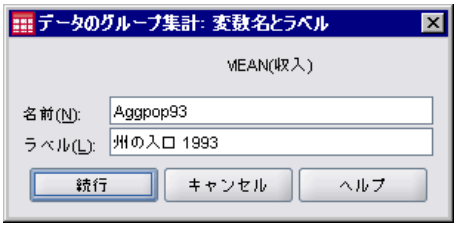
- 数値型変数の集計関数 (平均値、中央値、標準偏差、合計値を含む)
- ケースの数 (重み付けなし、重み付けあり、非欠損値、欠損値を含む)
- 超過 (未満) のパーセントまたは割合
- 範囲内 (外) のパーセントまたは割合

図 9-8
[集計関数の定義] ダイアログ ボックス

データのグループ集計: 変数名とラベル

[データのグループ集計] では、新規データ ファイル内の集計変数にデフォルトの変数名が割り当てられます。このダイアログ ボックスを使うと、[集計変数] ボックスのリスト内で選択した変数の変数名を変更し、記述的な変数ラベルを付けることができます。 詳細は、85 ページの 5 章の変数名を参照してください。

図 9-9
[変数名とラベル] ダイアログ ボックス

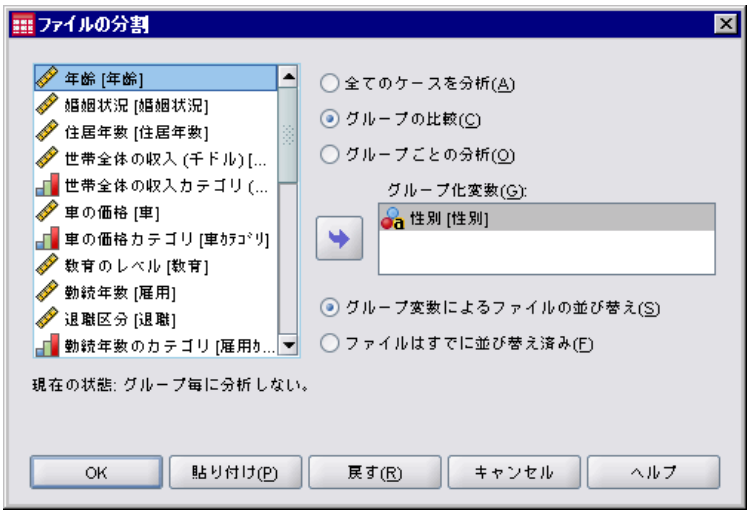


ファイルの分割

[ファイルの分割] では、分析を行うため、1 つ以上のグループ化変数の値に基づいてデータ ファイルを個別のグループに分割します。複数のグループ化変数を選択すると、[グループ化変数] ボックスのリスト内の直前の変数のカテゴリ内で各変数別にグループ化されます。たとえば、最初のグループ化変数として **性別**を選択し、2 番目のグループ化変数として **人種**を選択すると、ケースは、各性別カテゴリ内で人種の分類別にグループ化されます。

- 最大 8 つまでのグループ化変数を指定することができます。
- 長い文字型変数 (8 バイトを超える文字型変数) の場合は 8 バイトが 1 つの変数としてカウントされ、8 つまでがグループ化変数として使用できます。
- ケースは、[グループ化変数] ボックスに表示されている変数と同じ順序で、グループ化変数の値別に並べ替えられている必要があります。データ ファイルがまだ並べ替えられていない場合は、**[グループ化変数によるファイルの並び替え]** をクリックします。

図 9-10
[ファイルの分割] ダイアログ ボックス



グループの比較。ファイルの分割グループは、比較目的として一緒に提示されます。ピボット テーブルでは、単一のピボット テーブルが作成され、各ファイル分割変数はテーブル次元間を移動することができます。図表は、各ファイル分割グループごとに個別の図表が作成されて、ビューアにいっしょに表示されます。

グループごとの分析。各手続きから作成される結果はすべて、それぞれのファイルの分割グループに対して、個別に表示されます。

分析用にデータ ファイルを分割するには

▶ メニューから次の項目を選択します。

データ
ファイルの分割...

▶ [**グループの比較**] または [**グループごとの分析**] を選択します。

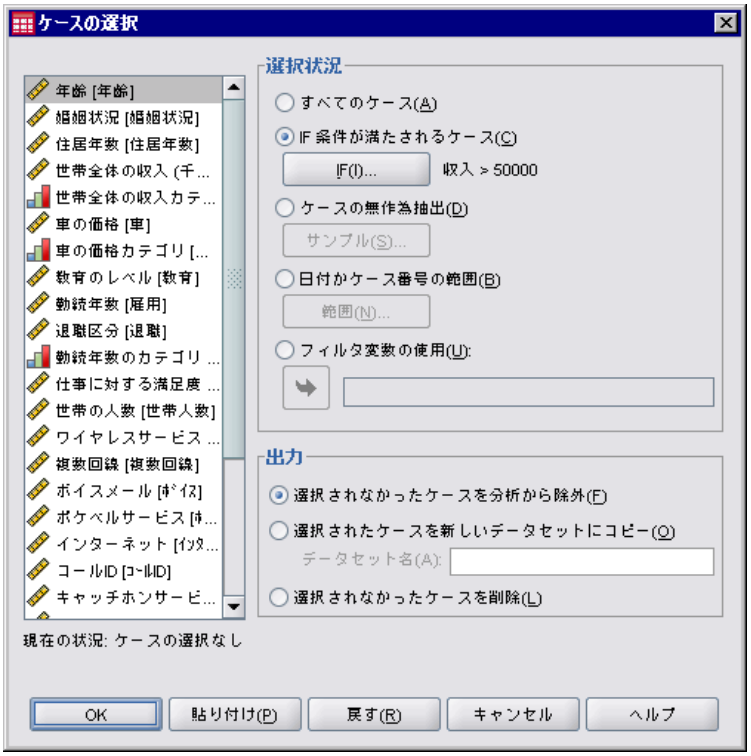
▶ 1 つ以上のグループ化変数を選択します。

ケースの選択

[ケースの選択] は、変数と複雑な式を含む基準に基づく、ケースのサブグループを選択するためのいくつかの方法を提供します。ケースの無作為抽出を選択することもできます。サブグループを定義するときに使用する基準には、次のものを含めることができます。

- 変数の値と範囲
- 日付と時刻の範囲
- ケース (行) 番号
- 算術式
- 論理式
- 関数

図 9-11
[ケースの選択] ダイアログ ボックス



全てのケース。ケースのフィルタリングをオフにして、すべてのケースを使用します。

IF 条件が満たされるケース。ケースを選択するには条件式を使用します。条件式の結果が真の場合、ケースが選択されます。結果が偽または欠損の場合、ケースは選択されません。

ケースの無作為抽出。近似パーセントまたは正確な数のケースに基づいて無作為サンプルを選択します。

日付かケース番号の範囲。一定範囲のケース数または一定範囲の日付/時刻に基づいてケースを選択します。

フィルタ変数の使用。フィルタ変数としてデータ ファイルから選択した数値型変数を使用します。0 以外の値を持つケース、あるいはフィルタ変数のないケースを選択します。

出力 (X11 ARIMA)

このセクションでは、選択されていないケースの処理方法を制御します。選択されなかったケースの処理には、次のオプションのどちらかを選択できます。

- **選択されなかったケースを分析から除外。**選択されなかったケースは分析から除外されますが、データセットには残っています。[分析から除外] をオフにすると、選択されなかったケースをこのセッションの後に使用できます。無作為抽出を選択するか、条件式に基づいてケースを選択した場合、「filter_$」という名前の変数が生成されます。選択されたケースではこの変数に 1 が、選択されなかったケースでは 0 が含まれます。
- **選択されたケースを新しいデータセットにコピー。**新しいデータセットに、選択したケースがコピーされます。元のデータセットには、影響はありません。選択していないケースは、新しいデータセットにはコピーされず、そのまま元のデータセットの中に残っています。
- **選択されなかったケースの削除。**選択されなかったケースは、データセットから削除されます。削除されたケースを回復するには、変更を保存しないでファイルを終了し、再びファイルを開くしかありません。データ ファイルに変更を保存すると、ケースは永久に削除されます。

注: 選択されなかったケースを削除してファイルを保存した場合は、そのケースを回復することはできません。

ケースのサブグループを選択するには

▶ メニューから次の項目を選択します。
データ
ケースの選択...

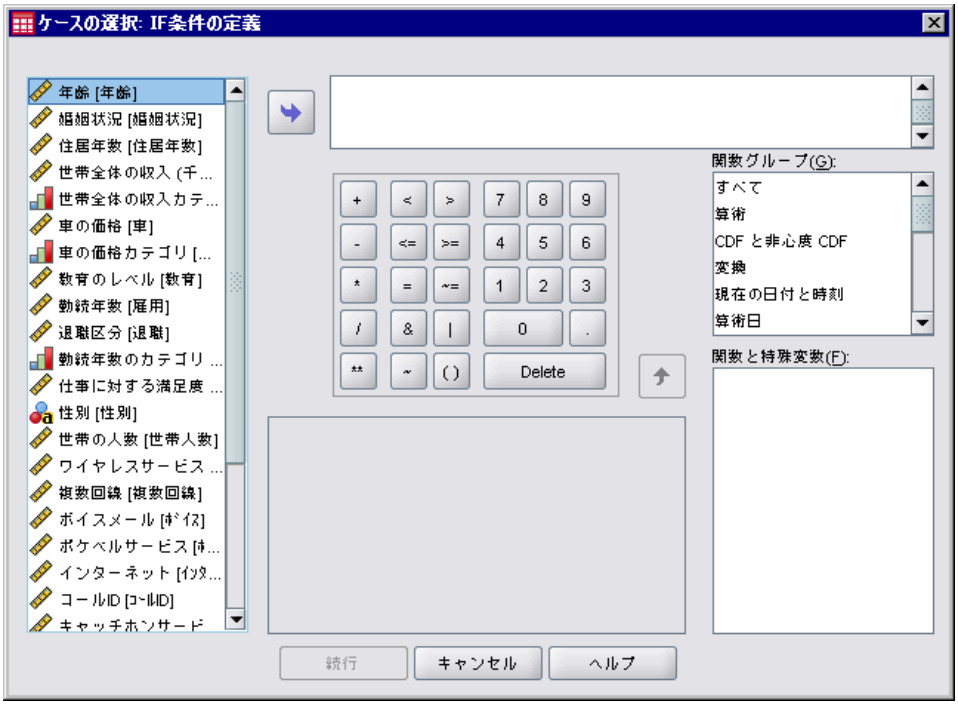
▶ ケースを選択する方法を 1 つ選択します。

▶ ケースを選択するための基準を指定します。

ケース選択: IF 条件の定義

このダイアログ ボックスでは、条件式を使ってケースのサブグループを選択することができます。条件式は、ケースごとに **真**の値、**偽**の値、または **欠損値**のいずれかを返します。

図 9-12
[ケースの選択: IF 条件の定義] ダイアログ ボックス

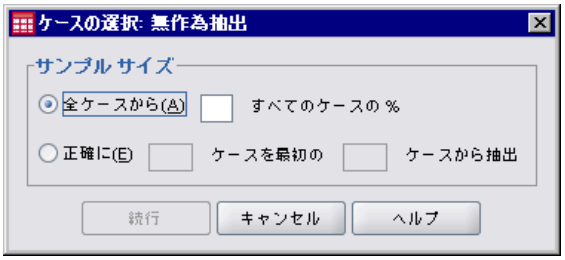


- 条件式の結果が **真**の場合、ケースは、選択したサブグループに含まれます。
- 条件式の結果が **偽**または **欠損**の場合、ケースは、選択したサブグループには含まれません。
- 条件式はそのほとんどが、計算パッド上の 6 つの関係演算子 (<、>、<=、>=、=、~=) を使用して表されます。
- 条件式には、変数名、定数、算術演算子、数値、および 関数、論理変数、関係演算子を使用できます。

ケースの選択: 無作為抽出

このダイアログ ボックスでは、ケースのおおよそのパーセントまたは正確な数に基づいて無作為抽出を選択することができます。抽出は非復元的に行われるため、同じケースを複数回選択することはできません。

図 9-13
[ケースの選択: 無作為抽出] ダイアログ ボックス



全ケースの。近似的に指定されたケースの割合で任意のサンプルを選択します。SPSS は各ケースに対して独立した疑似無作為決定を行うので、選択されたケースのパーセントを指定したパーセントに近づけることができます。データ ファイル内にケースが多くあるほど、選択されたケースのパーセントは指定したパーセントに近くなります。

正確に。ケースの数をユーザーが指定します。いくつのケースからサンプルを生成するかも指定しなければなりません。この 2 番目の数はデータ ファイル内にあるケースの総数以下でなければなりません。この数がデータ ファイル内にあるケースの総数よりも多いと、その数に比例してサンプルのケースは要求された数よりも少なくなってしまいます。

ケースの選択: 範囲の定義

このダイアログ ボックスでは、一定範囲のケース数、または一定範囲の日付または時刻に基づいてケースを選択します。

- ケースの範囲は、データ エディタに表示される行番号に基づいて定義されます。
- 日付と時刻の範囲は、定義された日付変数 ([データ] メニューの [日付の定義]) を持つ **時系列データ**にだけ利用することができます。

図 9-14
ケースの範囲を選択するための [ケースの選択: 範囲の定義] ダイアログ ボックス (定義された日付変数はなし)

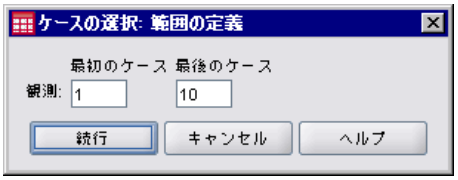
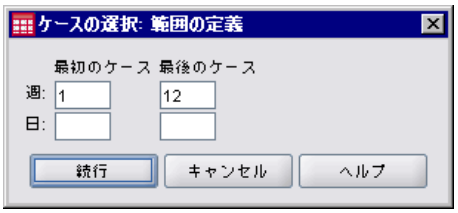


図 9-15
定義された日付変数を持つ時系列データのための [ケースの選択: 範囲の定義] ダイアログ ボックス

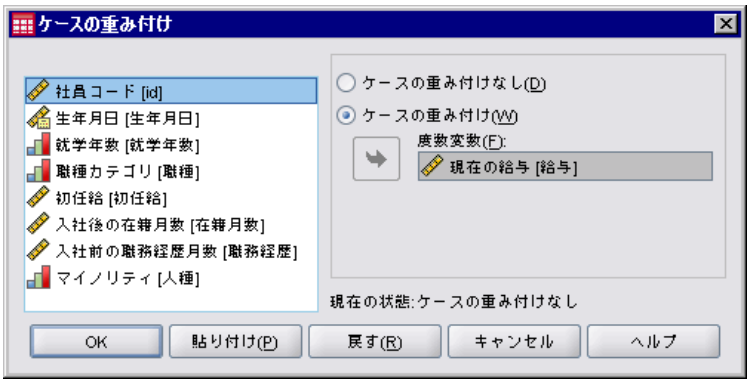


ケースの重み付け

[ケースの重み付け] では、統計分析を行うため (複製をシミュレートすることにより) ケースに異なる重みを付けることができます。

- 重み付け変数の値は、データ ファイル内の 1 つのケースが表す観測数を示していなければなりません。
- 重み付け変数に対してゼロ、マイナスの値、または欠損値のあるケースは、分析から除外されます。
- 小数点付きの値は有効です。それが意味がある場合は正確に使用され、(クロス集計表のような) ケースが集計される場合にも使用されます。

図 9-16
[ケースの重み付け] ダイアログ ボックス



重み付け変数をいったん適用すると、別の重み付け変数を選択するか、重み付けをオフにするまで有効になっています。重み付けデータ ファイルを保存すると、重み付け情報はデータ ファイルとともに保存されます。重み付けは、ファイルを重み付け形式で保存した後でも、いつでもオフにすることができます。

クロス集計表での重み付け。[クロス集計表] 手続きには、ケースの重みを処理するためのいくつかのオプションがあります。詳細は、302 ページの 16 章のクロス集計表のセル表示の設定を参照してください。

散布図とヒストグラムでの重み付け。散布図とヒストグラムでは、ケースの重み付けのオンまたはオフの切り替えができますが、重み付け変数に 0、マイナス値、または欠損値があるケースには影響を与えません。これらのケースは、図表内から重み付けをオフにした場合でも図表から除外されたままになっています。

ケースに重みを付けるには

▶ メニューから次の項目を選択します。
データ
ケースの重み付け...

▶ [**ケースの重み付け**] をクリックします。

▶ [度数変数] ボックスに度数変数を選択します。

度数変数の値は、ケースの重み付けとして使用されます。たとえば、度数変数として 3 の値を持つケースは重み付けしたデータ ファイル内では 3 つのケースを表します。

データの再構成

再構成データ ウィザードを使用して、使用する手続きに合わせてデータを再構成できます。このウィザードにより、現在のファイルが新しく再構成されたファイルに置き換えられます。このウィザードでは、次の処理ができます。

- 選択された変数をケースに再構成
- 選択されたケースを変数に再構成
- すべてのデータを置き換える

データを再構成するには

▶ メニューから次の項目を選択します。
データ
再構成...

▶ 実行する再構成の種類を選択します。

▶ 再構成するデータを選択します。

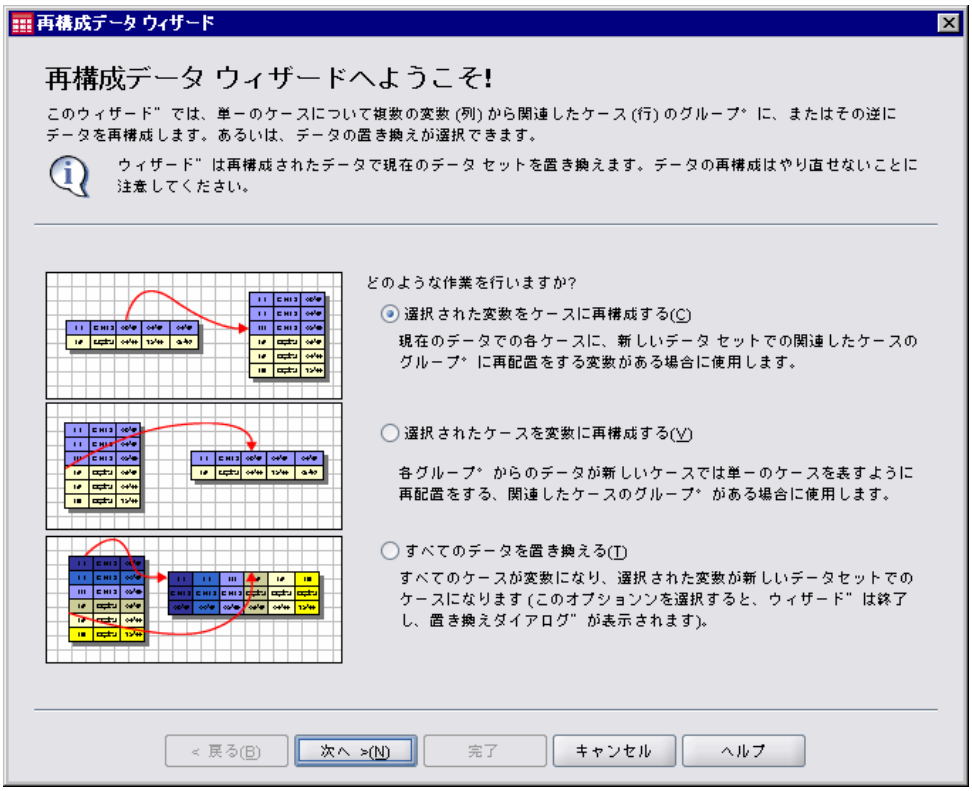
オプションとして、次の選択が可能です。

- 識別変数を作成する。識別変数を作成すると、新しいファイルの値が元のファイルのどの値から作成されているかを識別できます。
- 再構成を実行する前に、データを並べ替え
- 新規ファイルのオプションを定義
- コマンド シンタックスをシンタックス ウィンドウに貼り付け

再構成データ ウィザード: 種類の選択

再構成データ ウィザードを使用して、データを再構成します。最初に表示されるダイアログ ボックスで、実行する再構成の種類を選択します。

図 9-17
再構成データ ウィザード



- **選択された変数をケースに再構成する。**データに含まれている関連する列のグループを、行のグループとして新規データ ファイルに表示する場合に、このオプションを選択します。この再構成を選択すると、[**変数からケースへ**] というステップが表示されます。
- **選択されたケースを変数に再構成する。**データに含まれている関連する行のグループを、列のグループとして新規データ ファイルに表示する場合に、このオプションを選択します。この再構成を選択すると、[**ケースから変数へ**] というステップが表示されます。
- **すべてのデータを置き換える。**データを置き換える場合に、このオプションを選択します。新しいデータでは、行は列に、列は行にすべて置き替えられます。この再構成を選択すると、再構成データ ウィザードが閉じ、[行と列の置き換え] ダイアログ ボックスが開きます。

データの再構成方法の決定

変数には、分析の対象となる情報（例、測定や得点など）が含まれます。一方、**ケース**は観測（例、個人など）を表します。**単純な**データ構造では、各変数が 1 つの列、各ケースが 1 つの行となります。たとえば、あるクラスに在籍する生徒全員の検定の得点を測定する場合、得点の値はすべて 1 つの列に表示され、行は生徒ごとに 1 つずつ作成されます。

データを分析するときには、多くの場合ある条件下で変数がどのように変化するかを分析します。条件には、特定の試験的治療、人口統計、または時点などが含まれます。データ分析では、分析対象の条件を **因子**と呼びます。因子を分析する場合は、データ構造が**複雑に**なります。1 つの変数に関する情報が複数の列に含まれる（たとえば、因子のレベルごとに列が存在する）場合や、1 つのケースに関する情報が複数の行に含まれる（たとえば、因子のレベルごとに行が存在する）場合があります。再構成データ ウィザードを使用して、複雑なデータ構造を持つファイルの再構成が簡単に実行できます。

ウィザードで選択する再構成の種類は、現在のファイルの構成と再構成後の新規ファイルの構成によって決まります。

現在のファイルのデータ構造。現在のデータ構造は、因子が**別個の**変数に記録されている（ケース グループ）場合と、変数と**いっしょに**記録されている（変数グループ）場合があります。

- **ケース グループ。**現在のファイルで、変数と条件が別個の列に記録されている場合。次に例を示します。

変数	因子
8	1
9	1
3	2
1	2

この例では、関連している最初の 2 行が **ケース グループ**になります。これらの行には、同じ因子レベルのデータが含まれています。SPSS データ分析では、データがこのような構造である場合、因子を **グループ化変数**と呼びます。

- **列グループ。**現在のファイルで、変数と条件が同じ列に記録されている場合。次に例を示します。

var_1	var_2
8	3
9	1

この例では、関連している 2 列が **変数グループ**になります。これらは、因子レベル 1 については var_1、因子レベル 2 については var_2 という、同じ変数に対するデータを含んでいます。SPSS のデータ分析では、多くの場合、データをこのように構成する場合に因子を**反復測度**と呼びます。

新規ファイルのデータ構造。新規ファイルのデータ構造は、通常、データ分析に使用する手続きによって決まります。

- **ケース グループを必要とする手続き。**グループ化変数を必要とする分析を実行するには、データの構造をケース グループにする必要があります。このような分析の例には、**1 変量**、**多変量**、**分散成分**の一般線型モデル、混合モデル、OLAP キューブ、**独立サンプル**の t 検定、ノンパラメトリック検定などがあります。現在のデータ構造が変数グループであるときにこれらの分析を実行するには、[**選択された変数をケースに再構成する**] を選択します。
- **変数グループを必要とする手続き。**反復測定を分析するには、データ構造を変数グループにする必要があります。このような分析の例には、**反復測定**の一般線型モデル、**時間依存共変量**の Cox 回帰分析、**対応のあるサンプル**の t 検定、**対応サンプル**のノンパラメトリック検定などがあります。現在のデータ構造がケース グループであるときにこれらの分析を実行するには、[**選択されたケースを変数に再構成する**] を選択します。

変数からケースへ再構成する例

この例では、因子 *A* および *B* の検定の得点がそれぞれ別個の列に表示されています。

図 9-18
変数からケースへ再構成する現在のデータ

	スコア_a	スコア_b
1	11014.00	864.00
2	684.00	636.00
3	810.00	638.00

このデータを独立サンプルの *t* 検定で分析するとします。現在のデータには、「**スコア_a**」と「**スコア_b**」で構成されている列グループがありますが、この手続きで必要となる **グループ化変数**がありません。再構成データ ウィザードで [**選択された変数をケースに再構成する**] を選択します。1 つの変数グループを「**得点**」という新しい変数に再構成し、「**グループ**」というインデックスを作成します。次の図には、新しいデータ ファイルが示されています。

図 9-19
変数からケースに再構成された新しいデータ

	グループ	スコア
1	スコア_a	11014.00
2	スコア_b	864.00
3	スコア_a	684.00
4	スコア_b	636.00
5	スコア_a	810.00
6	スコア_b	638.00

これで、独立サンプルの *t* 検定を実行するときに、**グループ**をグループ化変数として使用できます。

ケースから変数へ再構成する例

この例では、各被験者の検定の得点が治療の前と後の 2 度記録されています。

図 9-20
ケースから変数へ再構成する現在のデータ

	ID	スコア	時間
1	1	1014.00	bef
2	1	864.00	aft
3	2	684.00	bef
4	2	636.00	aft

このデータを、対応のあるサンプルの *t* 検定で分析するとします。現在のデータ構造はケース グループですが、この手続きで必要となる対応のある変数の **反復測度**がありません。再構成データ ウィザードで [**選択されたケースを変数に再構成する**] を選択します。**ID** を使用して現在のデータの行グループを識別し、**時間**を使用して新規ファイルの変数グループを作成します。

図 9-21
ケースから変数に再構成された新しいデータ

	ID	aft	bef
1	1	864.00	1014.00
2	2	636.00	684.00
3			

これで、対応のあるサンプルの *t* 検定を実行するときに、**bef** と **aft** を変数のペアとして使用できます。

再構成データ ウィザード (変数からケースへ): 変数グループの数

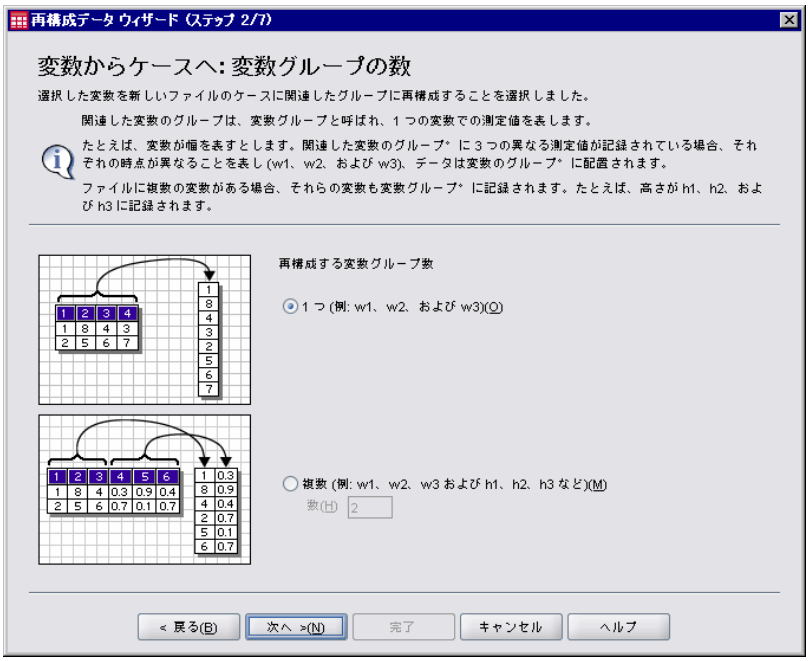
注: このステップが表示されるのは、変数グループから行への再構成を選択した場合です。

このステップでは、新規ファイルで再構成する現在のファイルの変数グループ数を決定します。

現在のファイルに含まれる変数グループの数。現在のデータに変数グループがいくつ存在しているかを調べます。**変数グループ**と呼ばれる関連する列のグループでは、同じ変数の反復測度が別々の列に記録されます。たとえば、現在のデータに **幅**を記録する **w1**、**w2**、および **w3** という 3 つの列がある場合、変数グループの数は 1 つになります。さらに、**高さ**を記録する **h1**、**h2**、および **h3** という 3 つの列がある場合、変数グループは 2 つあることになります。

新規ファイルに作成する変数グループの数。新規データ ファイルに再構成する変数グループの数を決定します。すべての変数グループを新規ファイルで再構成する必要はありません。

図 9-22
再構成データ ウィザード: 変数グループの数 (ステップ 2)



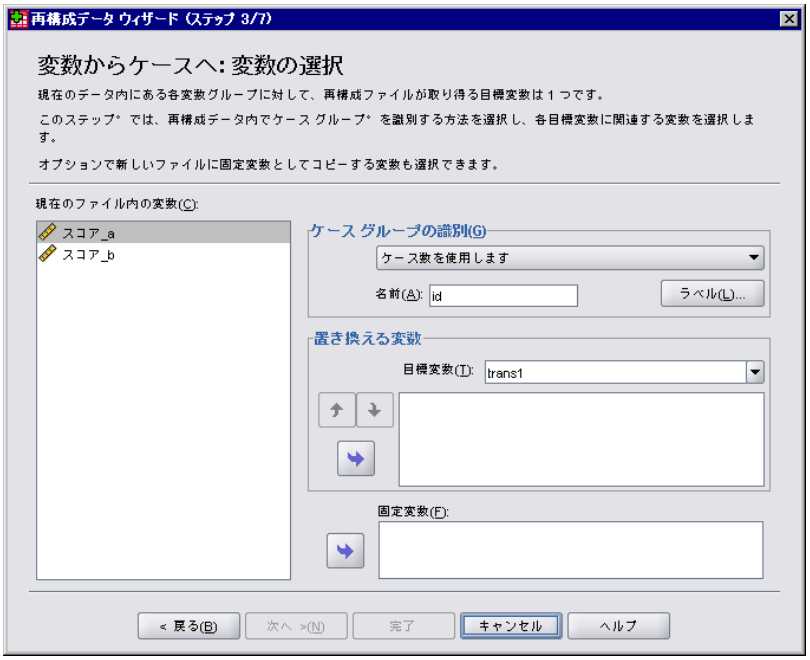
- **1 つ。**新規ファイルに、現在のファイルに含まれる 1 つの変数グループから再構成された変数が 1 つ作成されます。
- **複数。**新規ファイルに、再構成された複数の変数が作成されます。このステップで指定した数は、次のステップに影響します。次のステップでは、指定した数の新しい変数が自動的に作成されます。

再構成データ ウィザード (変数からケースへ): 選択 変数

注: このステップが表示されるのは、変数グループから行への再構成を選択した場合です。

このステップでは、現在のファイルに含まれる変数を新規ファイルでどのように使用するかを指定します。新規ファイルで行を識別する変数を作成することもできます。

図 9-23
再構成データ ウィザード: 変数の選択 (ステップ 3)



新しい行の識別方法。新規データ ファイルに、現在のデータ ファイルの行を識別する変数を作成し、新しい行グループの作成元を示すことができます。この識別変数には、ケースの連続番号や変数の値を使用できます。[ケース グループの識別] のコントロールを使用して、新規ファイルの識別変数を定義します。セルをクリックして、デフォルトの変数名を変更し、識別変数にわかりやすい変数ラベルを付けることができます。

新規ファイルで再構成の対象となる変数。前のステップでは、再構成する変数グループの数を指定しました。ウィザードは、グループごとに 1 つの新しい変数を作成します。変数グループの値は、新規ファイルの変数に表示されます。[置き換える変数] のコントロールを使用して、新規ファイルで再構成する変数を定義します。

再構成する変数を 1 つ指定するには

▶ 変換する変数グループを含む変数を [置き換える変数] リストに配置します。グループ内のすべての変数は、同じ型 (数値型か文字型のどちらか) でなければなりません。

変数グループには、同じ変数を何度も含めることができます (変数はソース変数リストから移動するのではなく、コピーされます)。変数の値は新規ファイルで反復されます。

再構成する変数を複数指定するには

▶ [目標変数] ドロップダウン リストから、定義する最初の目標変数を選択します。

▶ 変換する変数グループを含む変数を [置き換える変数] リストに配置します。グループ内のすべての変数は、同じ型 (数値型か文字型のどちらか) でなければなりません。変数グループに、同じ変数を 2 度以上含めることができます。(変数は、ソース変数リストから移されるのではなく、コピーされます。値は新しいファイル内で繰り返されます)。

▶ 定義する次の目標変数を選択し、利用するすべての目標変数に対して変数の選択手順を繰り返します。

- 同一の変数を同じ目標変数グループに複数回含めることはできますが、複数の目標変数グループに含めることはできません。
- それぞれの目標変数グループ リストには、同数の変数が含まれている必要があります。(2 度以上表示されている変数は数に含まれます)。
- 目標変数グループの数は、前のステップで指定した変数グループの数によって決まります。デフォルトの変数名はこのステップで変更できますが、再構成する変数グループの数を変更するには、前のステップに戻る必要があります。
- 次のステップに進む前に、利用するすべての目標変数に対して変数グループを定義する必要があります (ソース リストの変数を選択)。

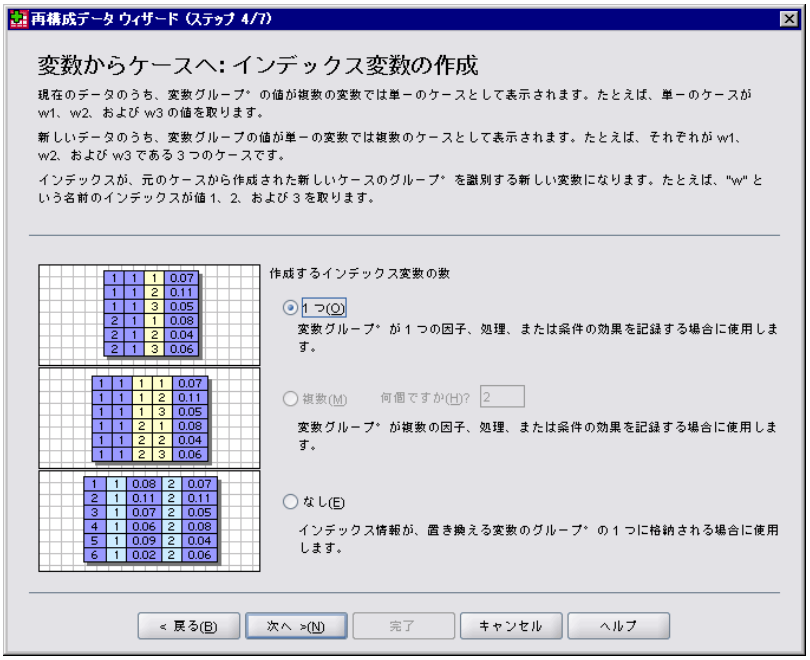
新規ファイルにコピーする変数。再構成しない変数は、新規ファイルにコピーできます。これらの変数の値は、新しい行にコピーされます。新規ファイルにコピーする変数を [固定変数] リストに移動します。

再構成データ ウィザード (変数からケースへ): インデックスの作成 変数

注: このステップが表示されるのは、変数グループから行への再構成を選択した場合です。

このステップでは、インデックス変数を作成するかどうかを決定します。インデックスとは、新規行の作成元変数に基づいて、行グループを順に識別する新しい変数です。

図 9-24
再構成データ ウィザード: インデックス変数の作成 (ステップ 4)



新規ファイルに作成するインデックス変数の数。インデックス変数は、手続きのグループ化変数として使用できます。ほとんどの場合、インデックス変数は 1 つで十分ですが、現在のファイルの変数グループが複数の因子レベルを反映している場合は、複数のインデックス変数が必要となることがあります。

- **1 つ。**インデックス変数が 1 つ作成されます。
- **複数。**複数のインデックスが作成されます。作成するインデックスの数を入力してください。このステップで指定する数は、次のステップに影響します。次のステップでは、指定した数のインデックスが自動的に作成されます。
- **なし。**新規ファイルにインデックス変数を作成しない場合に選択します。

変数からケースへの再構成で 1 つのインデックスを作成する例

現在のデータには、**幅**という変数グループ 1 つと、**時間**という因子 1 つが含まれています。幅は 3 度測定され、**w1**、**w2**、および **w3** に記録されています。

図 9-25
インデックス 1 つを構成する現在のデータ キーワード

	被験者	w1	w2	w3
1	1	6.70	4.30	5.70
2	2	7.10	5.90	5.60

この変数グループを **幅**という 1 つの変数に再構成し、番号から成るインデックスを 1 つ作成します。次の表に、新しいデータを示します。

図 9-26
インデックス 1 つを持つ再構成された新しいデータ

	被験者	インデックス	幅
1	1	1	6.70
2	1	2	4.30
3	1	3	5.70
4	2	1	7.10
5	2	2	5.90
6	2	3	5.60

インデックスの値は 1 から始まり、グループの変数ごとに増分されます。元のファイルで新しい行が見つかるたびに、再び 1 から番号が付けられます。これで、グループ化変数を必要とする手続きで、**インデックス**を使用できるようになりました。

変数からケースへの再構成で 2 つのインデックスを作成する例

変数グループに複数の因子が記録されている場合は、複数のインデックスを作成できます。その場合は、最初の因子のレベルが 1 次インデックスとなり、2 番目以降の因子のレベルがその内側で循環するよう現在のデータを配置する必要があります。現在のデータには、**幅**という 1 つの変数グループと、A と B という 2 つの因子が含まれています。データは、因子 B のレベルが因子 A のレベル内で循環するように配置されています。

図 9-27
インデックス 2 つを構成する現在のデータ インデックス

	被験者	w_a1b1	w_a1b2	w_a2b1	w_a2b2
1	1	5.50	6.40	5.80	5.90
2	2	7.40	7.10	6.60	6.70

この変数グループを **幅**という 1 つの変数に再構成し、インデックスを 2 つ作成します。次の表に、新しいデータを示します。

図 9-28
インデックス 2 つを持つ新しい再構成データ

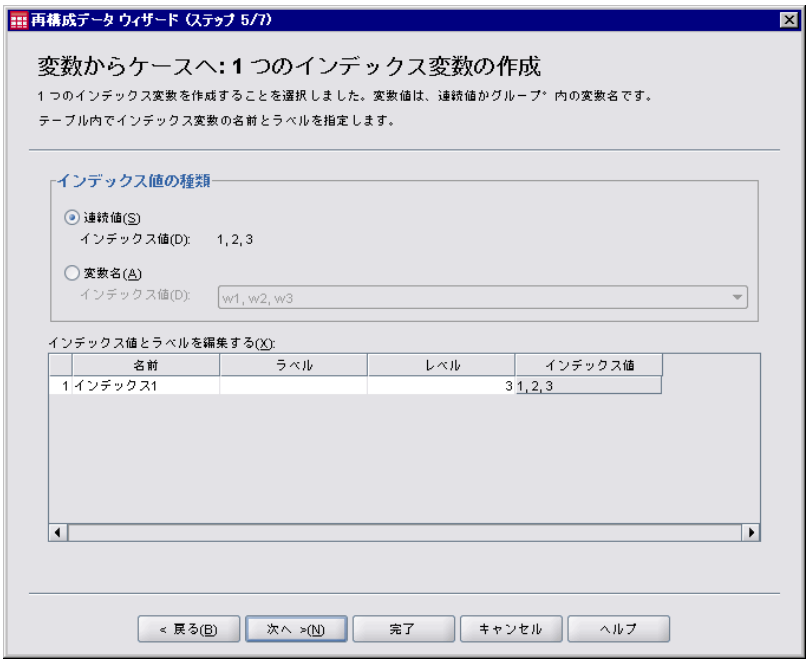
	被験者	インデックス_a	インデックス_b	幅
1	1	1	1	5.50
2	1	1	2	6.40
3	1	2	1	5.80
4	1	2	2	5.90
5	2	1	1	7.40
6	2	1	2	7.10
7	2	2	1	6.60
8	2	2	2	6.70

再構成データ ウィザード (変数からケースへ): 1 つのインデックス変数の作成

注: このステップは、変数グループから行への再構成、および 1 つのインデックス変数の作成を選択した場合に表示されます。

このステップでは、インデックス変数の値を決定します。インデックス変数の値には、連続番号や元の変数グループの変数名を使用できます。また、新しいインデックス変数の名前やラベルを指定することもできます。

図 9-29
再構成データ ウィザード: 1 つのインデックス変数の作成 (ステップ 5)



詳細は、218 ページの変数からケースへの再構成で 1 つのインデックスを作成する例を参照してください。

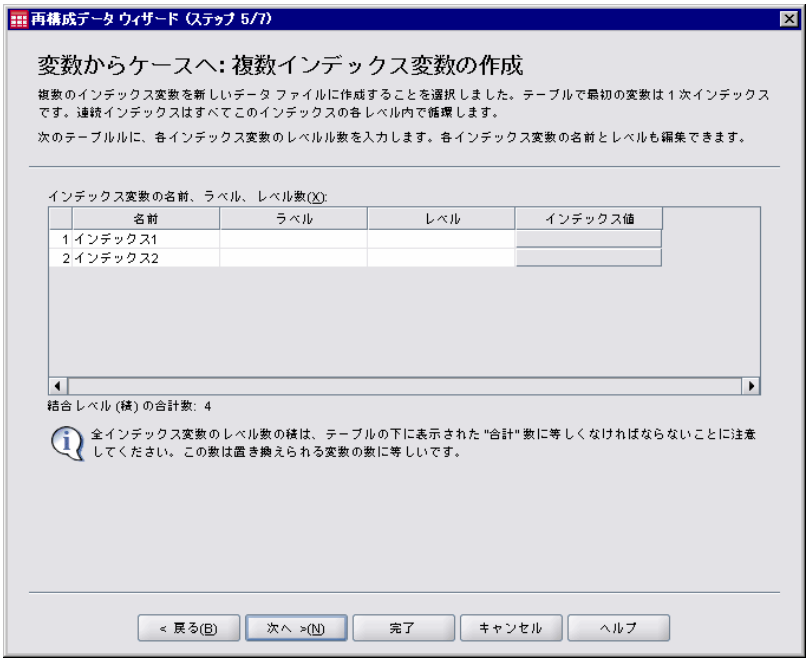
- **連続値**。連続番号が自動的にインデックスの値として割り当てられます。
- **変数名**。選択した変数グループの名前がインデックスの値として使用されます。リストから変数グループを選択します。
- **[名前] と [ラベル]**。セルをクリックして、デフォルトの変数名を変更し、インデックス変数にわかりやすい変数ラベルを付けることができます。

再構成データ ウィザード (変数からケースへ): 複数のインデックス変数の作成

注: このステップは、変数グループから行への再構成、および複数のインデックス変数の作成を選択した場合に表示されます。

このステップでは、各インデックス変数のレベルの数を指定します。また、新しいインデックス変数の名前やラベルを指定することもできます。

図 9-30
再構成データ ウィザード: 複数インデックス変数の作成 (ステップ 5)



詳細は、219 ページの変数からケースへの再構成で 2 つのインデックスを作成する例を参照してください。

現在のファイルに記録されているレベルの数。現在のデータに記録されている因子レベルの数を調べます。**レベル**は、同一の条件下にあるケース グループを定義します。因子が複数ある場合、最初の因子のレベルを 1 次インデックスとし、2 番目以降の因子のレベルがこのレベル内で循環するように現在のデータを配置する必要があります。

新規ファイルに作成するレベルの数。各インデックスのレベルの数を入力します。複数のインデックス変数の値は、常に連続番号になります。この値は 1 から始まり、その後レベルごとに増分されます。最初のインデックスは増分が最も遅く、最後のインデックスは増分が最速です。

組み合わされるレベルの合計。現在のデータに含まれているレベルの数より多くのレベルを作成することはできません。再構成されたデータには処理の組み合せごとに行が 1 つ含まれるので、作成するレベルの数はウィザードによって確認されます。ウィザードは、作成するレベルの積と変数グループの変数の数を比較します。これらの数は一致している必要があります。

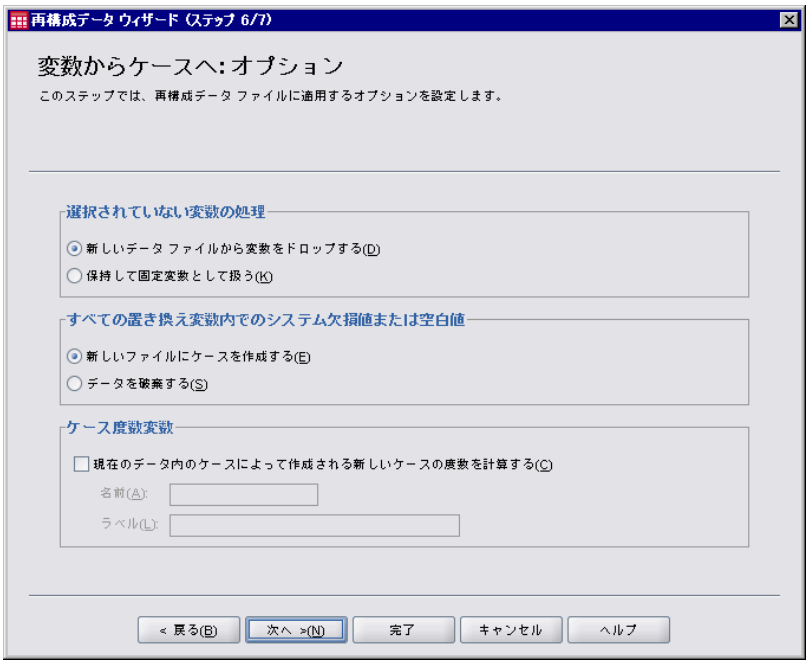
[名前] と [ラベル]。セルをクリックして、デフォルトの変数名を変更し、インデックス変数にわかりやすい変数ラベルを付けることができます。

再構成データ ウィザード (変数からケースへ):オプション

注: このステップが表示されるのは、変数グループから行への再構成を選択した場合です。

このステップでは、再構成される新規ファイルのオプションを指定します。

図 9-31
再構成データ ウィザード: オプション (ステップ 6)



選択されていない変数の破棄。変数選択のステップ (ステップ 3) で、現在のデータから再構成する変数グループ、コピーする変数、および識別変数を選択しました。選択されている変数のデータは、新規ファイルに表示されます。現在のデータにまだ変数が含まれている場合は、それらの変数を破棄するかまたは保持するかを選択できます。

欠損データの保持。ウィザードは、NULL 値が含まれていないかどうか、新しい行となる可能性のあるデータを 1 つずつ調べます。**NULL 値**とは、システム欠損値または空白値です。NULL 値だけを含む行は、保持するかまたは破棄するかを選択できます。

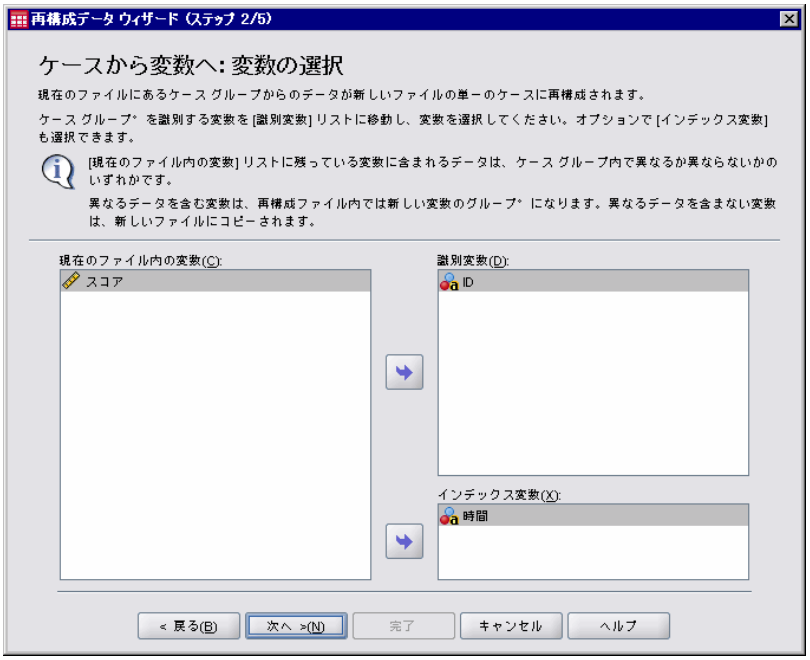
度数変数の作成。新規ファイルに、**度数変数**を作成できます。この変数には、現在のデータの行から生成された新しい行の数が含まれます。新規ファイルから NULL 値を破棄して、現在のデータに含まれる特定の行から異なる数の新しい行を生成する場合は、度数変数を使用すると便利です。セルをクリックして、デフォルトの変数名を変更し、度数変数にわかりやすい変数ラベルを付けることができます。

再構成データ ウィザード (ケースから変数へ): 選択 変数

注: このステップは、ケース グループから列への再構成を選択した場合に表示されます。

このステップでは、現在のファイルに含まれる変数を新規ファイルでどのように使用するかを指定します。

図 9-32
再構成データ ウィザード: 変数の選択 (ステップ 2)



現在のデータのケース グループを識別する変数。ケース グループとは、測定する観測単位 (個人や機関など) が同じであるという点で、互いに関連している行のグループです。新規ファイルで各ケース グループを 1 つの行に統合するには、現在のファイルでグループの識別に使用されている変数を指定する必要があ

ります。現在のファイルのケース グループを識別する変数を [識別変数] リストに移動します。現在のデータ ファイルの分割に使用される変数は、ケース グループの識別にも自動的に使用されます。ウィザードは、識別値の新たな組み合せが見つかるたびに新しい行を作成します。このため、[識別変数] リストに表示されている変数と同じ順序になるように、現在のファイルのケースを識別変数の値で並べ替える必要があります。現在のデータ ファイルがまだ並べ替えられていない場合は、次のステップで実行できます。

新規ファイルでの新規変数グループの作成方法。元データでは、変数は 1 つの列に表示されます。新しいデータ ファイルでは、変数は複数の新しい列に表示されます。**インデックス変数**とは、ウィザードが新しい列を作成するのに使用する現在のデータの変数のことです。再構成されたデータには、新しい列の一意な値ごとに新しい変数が 1 つ含まれます。新規変数グループの形成に使用する変数を [インデックス変数] リストに移動します。表示されるオプションを選択すると、新しい列をインデックスで並べ替えることもできます。

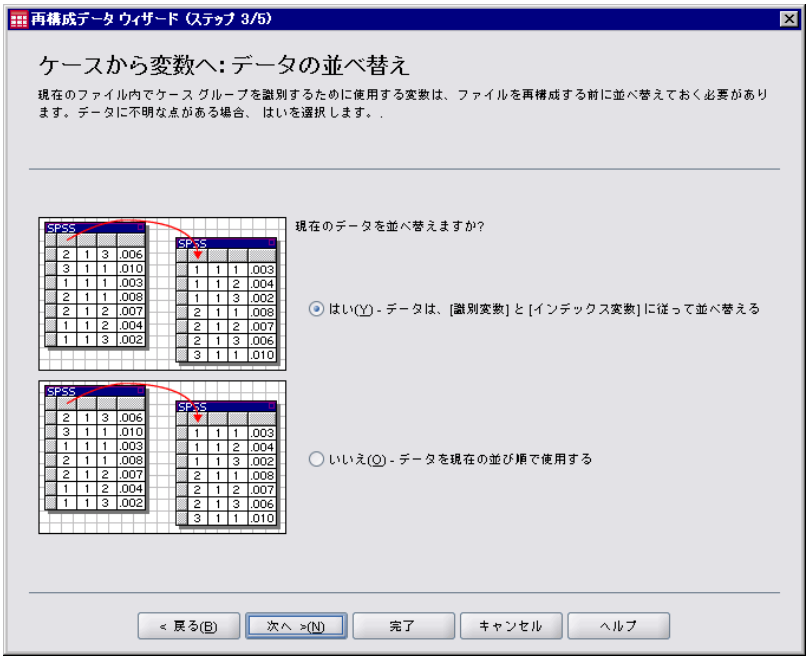
他の列の処理。[現在のファイル内の変数] リストに残っている変数の処理方法は、ウィザードによって自動的に決められます。ウィザードは、各変数のデータ値がケース グループ内で変化するかどうかを調べます。変化する場合、その値は新規ファイルで変数グループとして再構成されます。変化しない場合、その値は新規ファイルにコピーされます。

再構成データ ウィザード (ケースから変数へ): ソート データ

注: このステップは、ケース グループから列への再構成を選択した場合に表示されます。

このステップでは、再構成する前に現在のファイルを並べ替えるかどうかを決定します。ウィザードは、識別値の新たな組み合せが見つかるたびに新しい行を作成するため、ケース グループを識別する変数の順に、データを並べ替える必要があります。

図 9-33
再構成データ ウィザード: データの並べ替え (ステップ 3)



現在のファイルの行の順序。現在のデータの順序、およびケース グループの識別に使用する変数 (前のステップで指定している) を調べます。

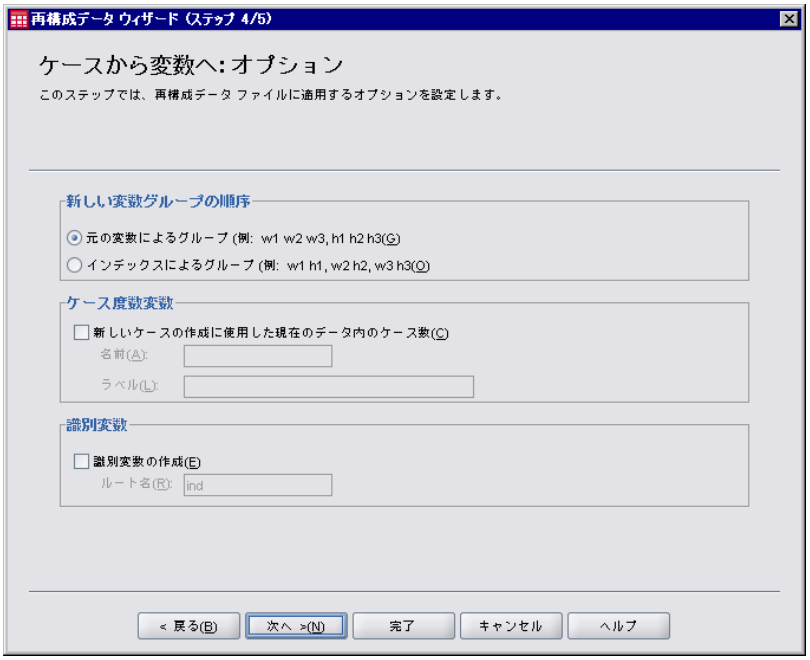
- **はい。**現在のデータは自動的に識別変数の順に並べ替えられ、前のステップで [識別変数] に表示されている変数と同じ順序になります。データが識別変数の順に並べ替えられていない場合、または並べ替えが実行されたかどうか不明である場合に選択します。このオプションでは個別のデータ パスが必要ですが、再構成のときに行を正確に並べ替えることができます。
- **いいえ。**現在のデータは並べ替えられません。現在のデータが、ケース グループを識別する変数の順にすでに並べ替えられている場合に選択します。

再構成データ ウィザード (ケースから変数へ): オプション

注: このステップは、ケース グループから列への再構成を選択した場合に表示されます。

このステップでは、再構成される新規ファイルのオプションを指定します。

図 9-34
再構成データ ウィザード: オプション (ステップ 4)



新規ファイルでの新しい変数グループの順序。

- **元の変数によるグループ。**同一の変数から作成された新しい変数をいっしょにグループ化します。
- **インデックスによるグループ。**インデックス変数の値に基づいて変数をグループ化します。

例。再構成する変数は w と h であり、インデックスは **月**です。

w	h	月

変数ごとにグループ化すると、次の表のようになります。

w.jan	w.feb	h.jan

インデックスごとにグループ化すると、次の表のようになります。

w.jan	h.jan	w.feb

度数変数の作成。新規ファイルに、度数変数を作成できます。この変数には、新規データ ファイルの行の作成元となった現在のデータの行数が含まれます。

識別変数の作成。インデックス変数を使用して、新規データ ファイルに **識別変数** を作成します。インデックス変数の各固有値に対し、新しい変数が 1 つ作成されます。識別変数は、ケースの値の有無を通知します。ケースが値を持っている場合、指示変数の値は 1 になります。ケースが値を持っていない場合は 0 になります。

例。インデックス変数は**製品**です。顧客が購入した製品を記録します。元のデータとしては、次のものがあります。

顧客	製品
1	chick
1	eggs
2	eggs
3	chick

識別変数を作成すると、**製品**の各固有値に対し、新しい変数が 1 つ作成されます。再構成データは、次のようになります。

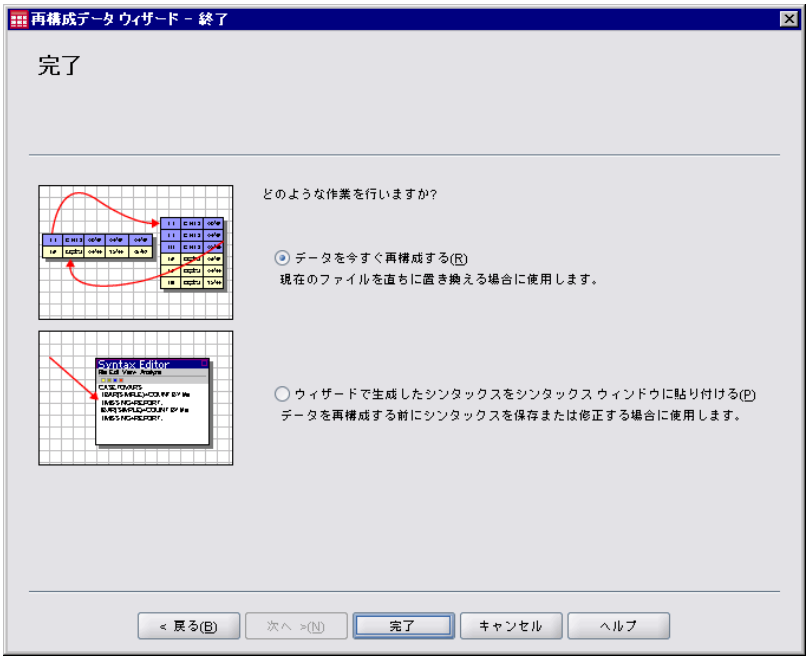
顧客	indchick	indeggs
1	1	1
2	0	1
3	1	0

この例では、再構成データを使用して、顧客が購入する製品の度数を求めることができます。

再構成データ ウィザード: 終了

これは再構成データ ウィザードの最終ステップです。どのように処理するかを決定します。

図 9-35
再構成データ ウィザード: 終了



- **データを今すぐ再構成する。**再構成された新規ファイルが作成されます。現在のファイルを直ちに置換する場合に、このオプションを選択します。

注: 元データに重みが付けられている場合、重みとして使用されている変数が再構成されるとき、およびこのような変数が新規ファイルから破棄されるときを除いて、新しいデータにも重みが付けられます。

- **ウィザードで生成したシンタックスをシンタックス ウィンドウに貼り付ける。**シンタックス ウィンドウに、ウィザードによって生成されたシンタックスが貼り付けられます。現在のファイルを置換する準備が完了していない場合、シンタックスを修正する場合、または後で使用できるようにシンタックスを保存する場合に、このオプションを選択します。

章
10

出力の作業

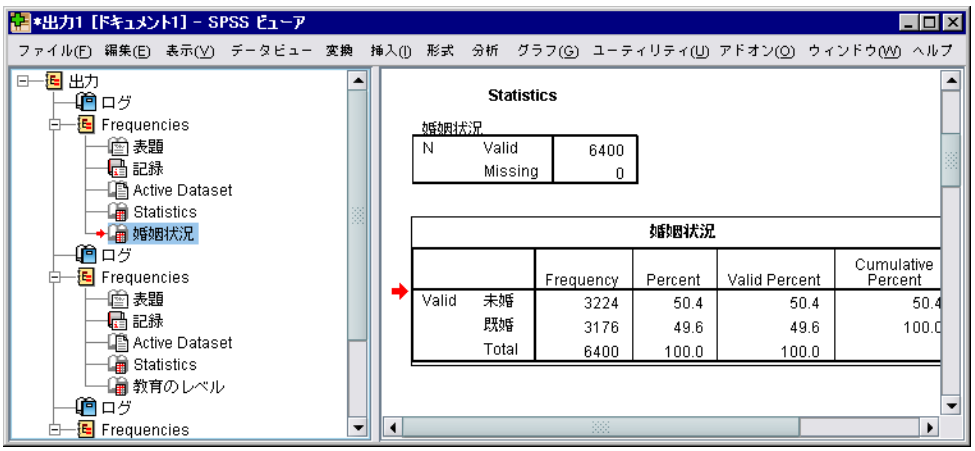
SPSS で手続きを実行すると、その結果はビューアというウィンドウに表示されます。このウィンドウでは、見たい出力に容易に移動することができます。出力を操作して、希望する出力を正確に取り入れた文書を作成することもできます。

ビューア (一般オプション)

出力結果はビューアに表示されます。ビューアを使用すると、次のことができます。

- 出力結果を走査する
- 選択したテーブルと図表を表示したり隠したりする
- 選択した項目を移動することによって出力の表示順を変更する
- 項目を [ビューア] とその他のアプリケーション間で移動する

図 10-1
ビューア (一般オプション)



ビューアは、2 つのウィンドウ枠に分割されています。

- 左のウィンドウ枠には、出力内容のアウトラインが表示されます。
- 右のウィンドウ枠には統計表、図表、およびテキスト出力が表示されます。

対応テーブルまたは図表に直接移動するには、アウトラインの項目をクリックします。アウトライン ウィンドウ枠の右境界線をクリックし、ドラッグするとアウトライン ウィンドウ枠の幅を変更することができます。

結果の表示と非表示

ビューアでは、手続き全体から個別のテーブルまたは結果を選択的に表示することも、隠すこともできます。この手順は内容ウィンドウ枠に表示される出力の量を減らしたいときに有効です。

テーブルと図表を隠すには

- ビューアのアウトライン ウィンドウ枠にある項目のブック アイコンをダブルクリックします。

  または

- 項目をクリックして選択します。
- メニューから次の項目を選択します。

  表示
  隠す

  または

- アウトラインのツール バーで、閉じた本の (隠す) アイコンをクリックします。

  開いたブック (表示) アイコンがアクティブ アイコンになり、現在項目が隠れていることを示します。

手続きの結果を隠すには

- アウトライン ウィンドウ枠で手続き名の左にあるボックスをクリックします。

  これで手続きの結果がすべて隠れて、アウトライン表示が閉じます。

出力の移動、削除、コピー

項目を 1 つずつ、またはグループ単位でコピー、移動、または削除して、結果の配置を変えることができます。

ビューア内で出力を移動するには

- アウトライン枠または内容ウィンドウ枠で項目を選択します。
- 選択した項目を別の場所にドラッグ アンド ドロップします。

ビューア内で出力を削除するには

▶ アウトライン枠または内容ウィンドウ枠で項目を選択します。

▶ Delete キーを押します。

または

▶ メニューから次の項目を選択します。

編集
削除

初期配置の変更

デフォルトでは、すべての出力は左に配置されます。新規項目の初期位置合わせを変更するには

▶ メニューから次の項目を選択します。

編集
オプション

▶ [**ビューア**] タブをクリックします。

▶ [初期出力状態] グループで、項目の種類 (たとえば、ピボット テーブル、グラフ、テキスト出力) を選択します。

▶ 希望する配置のオプションを選択します。

出力項目の配置の変更

▶ アウトラインまたは出力内容のウィンドウ枠で、位置合わせしたい項目を選択します。

▶ メニューから次の項目を選択します。

書式
左位置合わせ

または

書式
中揃え

または

書式
右詰め

ビューアのアウトライン

アウトライン ウィンドウ枠には、ビューア文書の目次が表示されます。アウトライン ウィンドウ枠を使用すると、出力の走査および出力表示のコントロールを行うことができます。アウトライン ウィンドウ枠内のほとんどの処理は、内容ウィンドウ枠に対応する影響を与えます。

- アウトライン ウィンドウ枠で項目を選択すると、内容ウィンドウ枠内に対応する項目が表示されます。
- アウトライン ウィンドウ枠で項目を移動すると、内容ウィンドウ枠内の対応する項目が移動します。
- アウトライン表示を閉じると、閉じたレベル中のすべての項目からの出力が隠れます。

アウトライン表示の制御。アウトライン表示を制御するには、次の操作を行います。

- アウトライン表示を開き、閉じます。
- 選択した項目のアウトライン レベルを変更します。
- アウトライン表示内の項目の大きさを変更します。
- アウトライン表示で使用するフォントを変更します。

アウトライン表示を閉じたり開くには

▶ 閉じる、または開くアウトライン項目の左にあるボックスをクリックします。

または

▶ アウトラインの項目をクリックします。

▶ メニューから次の項目を選択します。

表示
閉じる

または

表示
開く

アウトライン レベルを変更するには

▶ アウトライン枠内にある項目をクリックします。

▶ アウトラインのツールバーで左矢印をクリックして項目のレベルを上げます (項目を左に移動します)。

または

アウトラインのツールバーで右矢印をクリックして項目のレベルを下げます (項目を右に移動します)。

または

▶ メニューから次の項目を選択します。

編集
アウトライン
レベルを上げる

または

編集
アウトライン
レベルを下げる

アウトライン レベルの変更は、アウトライン レベル内で項目を移動した後で特に有効です。項目を移動すると、項目のアウトライン レベルが変化するので、アウトラインのツールバーで左矢印ボタンと右矢印ボタンを使用すると最初のアウトライン レベルに戻すことができます。

アウトライン項目のサイズを変更するには

▶ メニューから次の項目を選択します。

表示
アウトライン サイズ

▶ アウトラインのサイズ（**小**、**中**、または **大**）を選択します。

アウトラインのフォントを変更するには

▶ メニューから次の項目を選択します。

表示
アウトラインのフォント...

▶ フォントを選択します。

項目をビューアに追加する

ビューアでは、表題、新規テキスト、図表、または外部テキスト ファイルのような項目を追加することができます。

表題やテキストを追加するには

テーブルまたは図表に連結されていないテキスト項目は、ビューアに追加することができます。

▶ テーブル、図表、または表題またはテキストの前にくる他のオブジェクトをクリックします。

▶ メニューから次の項目を選択します。

挿入
新しい表題

または

挿入
新しいテキスト

▶ 新しいオブジェクトをダブルクリックします。

▶ テキストを入力します。

テキスト ファイルを追加するには

▶ ビューアのアウトライン枠または内容ウィンドウ枠で、テキストを追加する位置の直前にくるテーブル、図表、その他のオブジェクトをクリックします。

▶ メニューから次の項目を選択します。

挿入
テキスト ファイル...

▶ テキスト ファイルを選択します。

テキストを編集するには、そのテキストをダブル クリックします。

オブジェクトを SPSS ビューアに貼り付ける

他のアプリケーションのオブジェクトも、ビューアに貼り付けることができます。[後に貼り付け] か [形式を選択して貼り付け] のどちらかを使用します。いずれの方法でも、選択したオブジェクトの後に新しいオブジェクトが貼り付けられます。[形式を選択して貼り付け] は、貼り付けるオブジェクトの形式を選択したい場合に使用します。

[ビューア] の情報の検索および置換

▶ [ビューア] の情報の検索と置換には、メニューから次の項目を選びます。

編集
検索

または

編集
置換

図 10-2
[検索と置換] ダイアログ ボックス

検索と置換を使用すると、次のことができます。

- すべての文書または選択した項目だけを検索。
- 現在の場所から上下に検索。
- アウトライン枠または内容ウィンドウ枠への検索を制限するか、両方の枠を検索。
- 非表示項目を検索。内容ウィンドウ枠 (たとえば、デフォルトでは非表示の記録テーブル) およびピボット テーブルの非表示の行および列などの項目を含みます。
- 検索基準を大文字と小文字が区別される一致に制限。
- ピボット テーブルの検索基準をすべてのセルの内容の一致に制限。

非表示の項目およびピボット テーブル層

- 複合多次元のピボット テーブルの現在表示されている層の下の層は非表示と考えられなく、非表示の項目が検索に含まれていなくても検索領域に含まれます。
- 非表示項目はデフォルトでは非表示の、またはテーブルを編集し、指定した非表示の列と行を選択して手動で非表示にしたピボット テーブルの列と行 (たとえば、空白の行および列はデフォルトでは非表示) および内容ウィンドウ

枠 (アウトライン枠の閉じた本のアイコン付きの項目、またはアウトライン枠の閉じられたブロック内に含んだ) を含みます。非表示項目は、**[隠されている項目を含む]** を明示的に選択すると検索に含まれます。

- 両方の場合に、検索テキストまたは値を含む隠れたまたは見えない要素が見つかれば表示されますが、項目は元の状態に戻ります。

出力を他のアプリケーションにコピー

出力オブジェクトは、コピーして、ワード プロセッサ、またはスプレッドシートなど他のアプリケーションに貼り付けることができます。さまざまな形式で出力を貼り付けることができます。使用するアプリケーションにより、次の形式のいくつかが利用できます。

ピクチャ (メタファイル)。ピボット テーブル、テキスト出力、図表はメタファイルのピクチャとして貼り付けることができます。絵の形式は他のアプリケーションで大きさを変更することができ、他のアプリケーションの機能を使用して一定量の編集を行うこともできます。ピクチャとして貼り付けたピボット テーブルは、すべての罫線とフォント特性を保持します。この形式は、Windows オペレーティング システムでのみ使用できます。

RTF (リッチ テキスト形式)。ピボット テーブルは、RTF 形式で他のアプリケーションに貼り付けることもできます。ほとんどのアプリケーションでは、後で他のアプリケーションで編集可能なテーブルとしてピボット テーブルを貼り付けます。

注: Microsoft Word では、横幅が極端に長い表は表示されないことがあります。

ビットマップ。ピボット テーブルと図表はビットマップ形式で他のアプリケーションに貼り付けることができます。

BIFF。テーブルの内容は、スプレッドシートに貼り付けて、数値精度を保持できます。

テキスト。テーブルの内容は、テキストとしてコピーして貼り付けることができます。この手順は、テキストしか送受信できない電子メールなどのアプリケーションには有効です。

使用するアプリケーションが複合利用可能形式をサポートしているなら、形式を選択できる 形式を選択して貼り付け...] メニュー項目があるか、利用可能な形式の一覧が自動的に表示されます。

出力項目をコピーして他のアプリケーションに貼り付けるには

▶ アウトライン枠または内容ウィンドウ枠でオブジェクトを選択します。

▶ ビューアのメニューから、次の項目を選択します。

編集
コピー

▶ 使用するアプリケーションのメニューから、次の項目を選択します。

編集
貼り付け

または

編集
形式を選択して貼り付け...

貼り付け。出力は多数の形式でクリップボードにコピーされます。それぞれのアリケーションで、貼り付けに最も適した形式を決定します。

形式を選択して貼り付け。SPSS は、出力を複数の形式でクリップボードにコピーします。[形式を選択して貼り付け] を使用すると、使用するアプリケーションで利用できる形式のリストから希望する形式を選択することができます。

出力のエクスポート

[出力のエクスポート] では、ビューア出力を、HTML、テキスト、Word/RTF、Excel、PowerPoint (PowerPoint 97 以降のバージョンが必要です)、および PDF の各形式で保存します。各種書式のグラフィック ファイルとして図表をエクスポートできます。

注: PowerPoint へのエクスポートは、Windows オペレーティング システム上のみで利用でき、Student 版では利用できません。

出力をエクスポートするには

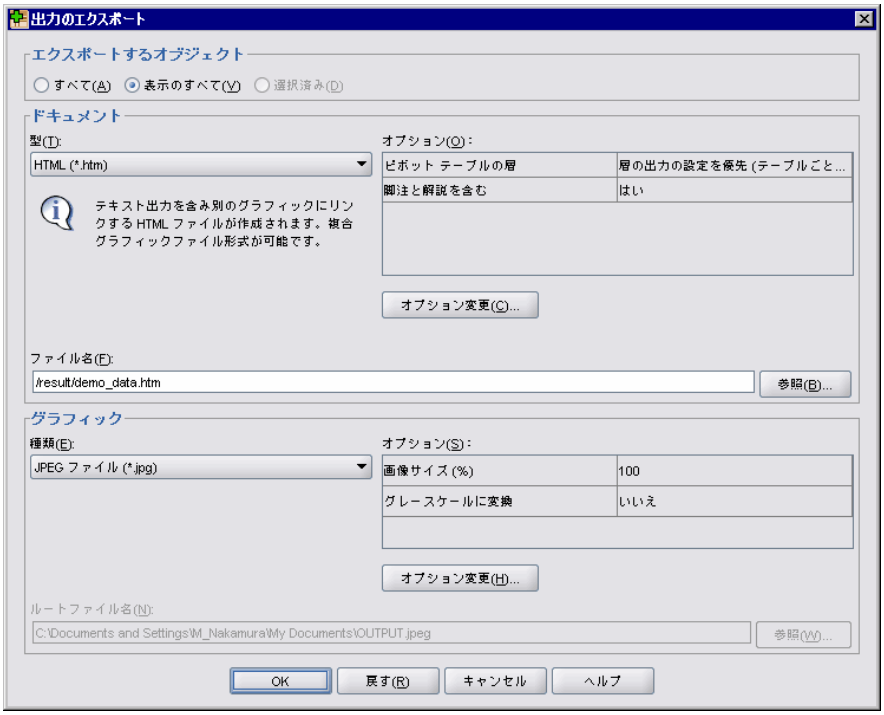
▶ ビューアをアクティブ ウィンドウにします (ウィンドウ内の任意の場所をクリックします)。

▶ メニューから次の項目を選択します。

ファイル
エクスポート...

▶ ファイル名 (または図表には接頭辞) を入力して、エクスポート書式を選択します。

図 10-3
[出力のエクスポート] ダイアログ ボックス



エクスポートするオブジェクト。ビューア内のすべてのオブジェクト、表示されているすべてのオブジェクト、または選択したオブジェクトのみをエクスポートできます。

文書の種類。次のオプションを使用できます。

- Word/RTF (*.doc)。ピボット テーブルは、セルの罫線、フォント スタイル、背景色など、あらゆる書式属性とともに、Word テーブルとしてエクスポートされます。テキスト出力は書式付きの RTF としてエクスポートされます。Windows オペレーティング システム上では、図表は EMF (拡張メタファイル) 形式でドキュメントに挿入されます。他のオペレーティング システムでは、図表は PNG 形式で挿入されます。テキスト出力は、常に固定ピッチ (等幅) フォントで表示され、同じフォント属性のままでエクスポートされます。スペース区切りテキスト出力を正しく位置合わせするには、固定ピッチ フォントが必要です。

  注: Microsoft Word では、横幅が極端に長い表は表示されないことがあります。

- Excel (*.xls)。ピボット テーブルの行、列、セルは、Excel の行、列、セルとしてエクスポートされます。この場合、セルの罫線、フォント スタイル、背景色など、すべての書式属性も反映されます。テキスト出力は、すべてのフォント属性とともにエクスポートされます。テキスト出力の各行は、Excel ファイルの行になります。この場合、1 行の内容はすべて 1 つのセルに入れられます。図表は含まれません。

- **HTML (*.htm)。**ピボット テーブルは HTML テーブルとしてエクスポートされます。テキスト出力は、事前にフォーマットされた HTML としてエクスポートされます。図表は参照により埋め込まれます。HTML 文書を含む適当な形式で図表をエクスポートする必要があります (たとえば、PNG および JPEG)。
- **Portable Document Format (*.pdf)。**出力はすべて、[印刷プレビュー] に表示されるとおりにエクスポートされます。またその際、すべての書式属性は維持されます。
- **PowerPoint ファイル (*.ppt)。**ピボット テーブルは Word テーブルとしてエクスポートされ、PowerPoint ファイルへはそれぞれのピボット テーブルが個別のスライドへ 1 つずつ埋め込まれます。ピボット テーブルは、セルの罫線、フォント スタイル、背景色など、あらゆる書式属性が保持されます。図表は TIFF 形式でエクスポートされます。テキスト出力は含まれません。

  注: PowerPoint へのエクスポートは、Windows オペレーティング システム上のみで利用でき、Student 版では利用できません。
- **テキスト (*.txt)。**出力形式は、テキスト形式、UTF-8、および UTF-16 を含みます。ピボット テーブルは、タブ区切り書式またはスペース区切り書式でエクスポートできます。すべてのテキスト出力はスペース区切り書式でエクスポートされます。図表では、各図表に対してエクスポートする図表のファイル名を示す行がテキスト ファイルに挿入されます。
- **なし (グラフィックスのみ)。**利用できるエクスポートの書式には、EPS、JPEG、TIFF、PNG、および BMP があります。Windows オペレーティング システムでは、EMF (拡張メタファイル) 形式も利用できます。

出力管理システム。すべての出力またはユーザーが指定した種類の出力を、テキスト、HTML、XML、または SPSS 形式データ ファイルとして自動的にエクスポートできます。詳細は、531 ページの 48 章の出力管理システムを参照してください。

HTML、Word/RTF、および Excel オプション

次のオプションが、ピボット テーブルを Word/RTF、Excel、および HTML にエクスポートするときに利用できます。

- **ピボット テーブルの層。**デフォルトでは、ピボット テーブルの除外および包括は各ピボット テーブルのテーブル プロパティによりコントロールされます。この設定を上書きして、すべての層を含むことも、または現在見えている層以外のすべてを除くこともできます。詳細は、265 ページの 11 章のテーブル プロパティ: 印刷を参照してください。
- **脚注と解説を含む。**すべてのピボット テーブルの脚注および解説の除外および包括をコントロール。

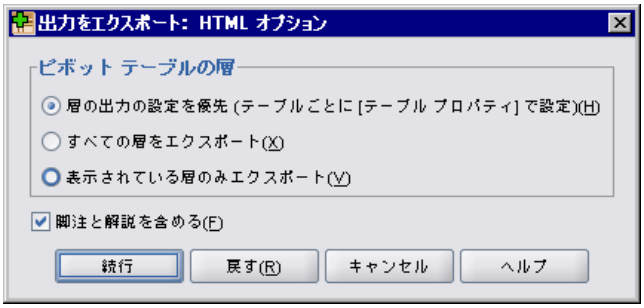
注: HTML では、図表をエクスポートする画像ファイル形式をコントロールすることもできます。詳細は、243 ページの図表エクスポート オプションを参照してください。

HTML、Word/RTF、および Excel のエクスポート オプションを設定するには

▶ エクスポート書式として、[HTML ファイル]、[Word/RTF ファイル]、または [Excel ファイル] を選択します。

▶ [オプション変更] をクリックします。

図 10-4
HTML、Word/RTF、および Excel エクスポート出力オプション



PowerPoint オプション

PowerPoint では次のオプションを使用できます。

- **ピボット テーブルの層。**デフォルトでは、ピボット テーブルの除外および包括は各ピボット テーブルのテーブル プロパティによりコントロールされます。この設定を上書きして、すべての層を含むことも、または現在見えている層以外のすべてを除くこともできます。詳細は、265 ページの 11 章のテーブル プロパティ: 印刷を参照してください。
- **脚注と解説を含む。**すべてのピボット テーブルの脚注および解説の除外および包括をコントロール。
- **スライド表題としてビューアのアウトラインのエントリを使用します。**エクスポートの際に作成されるそれぞれのスライドに表題が付けられます。各スライドには、ビューアからエクスポートされた項目が 1 つずつ含まれます。表題はビューアのアウトラインのウィンドウ枠のアウトライン項目から作成されます。

PowerPoint エクスポート オプションを設定するには

▶ エクスポートの形式として [PowerPoint] を選択します。

▶ [オプション変更] をクリックします。

注: PowerPoint へのエクスポートは、Windows オペレーティング システムでのみ使用できます。

PDF のオプション

PDF では次のオプションを使用できます。

ブックマークを埋め込む。このオプションを選択すると、ビューアのアウトラインのエントリに対応するブックマークが PDF 文書に挿入されます。ブックマークを埋め込むことにより、ビューアのアウトライン ウィンドウ枠と同様に、多数の出力オブジェクトを含む文書内の移動が簡単になります。

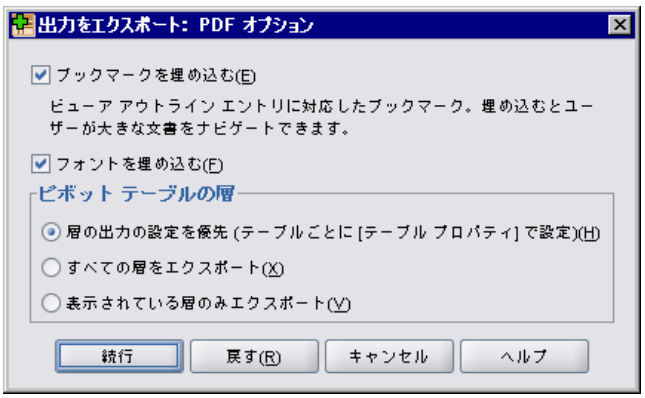
フォントを埋め込む。フォントを埋め込むと、PDF 文書をどのコンピュータでも同じように表示させることができます。フォントが埋め込まれていない場合、PDF 文書を表示（または印刷）するとき、コンピュータで利用できない一部のフォントに対して代用フォントが使用されます。ただし、この場合は、最適な結果が得られないことがあります。

ピボット テーブルの層。デフォルトでは、ピボット テーブルの除外および包括は各ピボット テーブルのテーブル プロパティによりコントロールされます。この設定を上書きして、すべての層を含むことも、または現在見えている層以外のすべてを除くこともできます。詳細は、265 ページの 11 章のテーブル プロパティ: 印刷を参照してください。

PDF エクスポート オプションを設定するには

- エクスポートの形式として [Portable Document Format] を選択します。
- [オプション変更] をクリックします。

図 10-5
[PDF:オプション] ダイアログ ボックス



PDF 出力に影響するその他の設定

ページ設定/ページ属性。PDF 文書のページ サイズ、方向、余白、ページのヘッダーやフッターの内容と表示、および図表の印刷サイズは、各種ページ設定オプションおよびページ属性オプションで制御します。

テーブル プロパティ/テーブルルック。幅の広いテーブルや縦に長いテーブルの縮小と拡大、およびテーブル層の印刷は、各テーブルのテーブル プロパティで制御します。これらのプロパティは、テーブルルックに保存することもできます。詳細は、265 ページの 11 章のテーブル プロパティ: 印刷を参照してください。

デフォルト/現在のプリンタ。PDF 文書の解像度 (DPI) は、デフォルトの解像度設定または現在選択されているプリンタに適用される現在の解像度設定となります ([ページ設定] で変更できます)。最大解像度は 1200 DPI です。プリンタ設定の方が解像度が高い場合、PDF 文書の解像度は 1200 DPI になります。

注: 高解像度の文書を低解像度プリンタで印刷すると、最適な印刷結果が得られないことがあります。

テキスト オプション

テキストのエクスポートでは次のオプションを使用できます。

ピボット テーブル形式。ピボット テーブルは、タブ区切り書式またはスペース区切り書式でエクスポートできます。スペース区切り書式では、次のことがコントロールできます。

- **列幅。**[Autofit] は、表示内容を折り返さなく、各列は列の最大幅のラベルまたは値と同じ幅です。**[カスタム]** は、列の次の行に折り返す幅を超える値およびテーブルのすべての列に適用される最大列幅を設定します。
- **行と列の罫線。**行と列の罫線を作成するのに使用された文字を制御します。行と列の罫線を非表示にするには、値にスペースを入力します。

ピボット テーブルの層。デフォルトでは、ピボット テーブルの除外および包括は各ピボット テーブルのテーブル プロパティによりコントロールされます。この設定を上書きして、すべての層を含むことも、または現在見えている層以外のすべてを除くこともできます。 詳細は、265 ページの 11 章のテーブル プロパティ: 印刷を参照してください。

脚注と解説を含む。すべてのピボット テーブルの脚注および解説の除外および包括をコントロール。

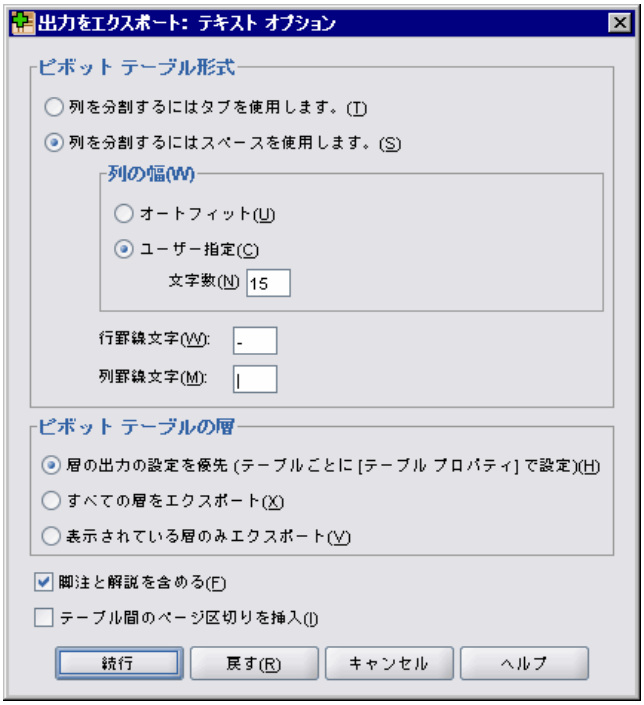
テーブル間のページ区切りを挿入。各テーブル間に用紙送り/ページ区切りが挿入されます。複数の層を持つピボット テーブルの場合、このオプションにより層と層の間にページ区切りが挿入されます。

テキスト エクスポートのオプションを設定するには

▶ エクスポートの形式として [Text] を選択します。

▶ **[オプション変更]** をクリックします。

図 10-6
[テキスト オプション] ダイアログ ボックス



図表エクスポート オプション

HTML およびテキスト文書では、図表のエクスポートに関してはグラフィック形式を選択でき、各グラフィック形式では各種のオプション設定をコントロールできます。

エクスポートした図表のオプションとグラフィック形式を選択します。

- 文書の種類として、HTML、Text、または **なし (グラフィックのみ)** を選択します。
- ドロップダウン リストからグラフィックファイル形式を選択します。
- 選択したグラフィック ファイル形式のオプションを変更するには **[オプションの変更]** をクリックします。

JPEG 図表のエクスポート オプション

- **イメージ サイズ。**元の図表のサイズのパーセントは最大 200 パーセントです。
- **グレースケールに変換。**色を網掛けのグレーに変更します。

BMP 図表のエクスポート オプション

- **イメージ サイズ。**元の図表のサイズのパーセントは最大 200 パーセントです。
- **ファイル サイズを縮小するためにイメージを圧縮します。**可逆的圧縮技術は、イメージ品質を損なわずに小さなファイルを作成します。

PNG 図表エクスポート オプション

イメージ サイズ。元の図表のサイズのパーセントは最大 200 パーセントです。

カラー深度。エクスポートされる図表での色の数を決定します。ある色数で保存されたチャートには、そのチャートで実際に使用されている最低限の色数と保存した色数で表現できる最大限の色数が適用されます。たとえば、3 つの色 (赤、白、および黒) を含む図表を 16 色で保存する場合、図表は 3 色に保持されます。

- また、図表での色の数が保存した色数で表現できる色数を上回る場合、その図表の色を再現するために色がディザリングされます。
- [**現在のスクリーン深度**] は、コンピュータのモニターに現在表示されている色の数です。

EMF および TIFF 図表のエクスポート オプション

イメージ サイズ。元の図表のサイズのパーセントは最大 200 パーセントです。

注: EMF (拡張メタファイル)形式は、Windows オペレーティング システムでのみ使用できます。

EPS 図表のエクスポート オプション

イメージ サイズ。元のイメージ サイズのパーセントを (200 パーセントまで) 選択するか、またはピクセルでイメージ幅を指定できます (高さは幅値および縦横比により決定されます)。エクスポートされたイメージは元のイメージに比例します。

TIFF プレビュー イメージを含む。EPS イメージを伴うプレビューを TIFF 形式で保存して、画面上に EPS イメージを表示できないアプリケーションでも表示できるようにします。

フォント。EPS イメージでのフォントの実装を制御します。

- **フォント参照の使用。**出力デバイスが図表内で使用されているフォントに対応している場合、そのフォントを使用します。対応していないフォントについては、代替フォントが使用されます。
- **フォントを曲線に置換。**フォントを PostScript 曲線データに置換します。テキスト自体は、EPS グラフィックスを編集するアプリケーションのテキストのように編集できなくなります。出力デバイスが図表内で使用されているフォントに対応していない場合に便利なオプションです。

ビューアの印刷

ビューア ウィンドウの内容を印刷するたのオプションが 2 つあります。

全ての表示されている出力。内容枠に現在表示されている項目のみを印刷します。隠れている項目（アウトライン枠に閉じたブック アイコンのある項目、または閉じたアウトライン層に隠れた項目）は印刷されません。

選択部分。アウトライン枠か内容ウィンドウ枠、またはその両方で現在選択されている項目のみを印刷します。

出力と図表を印刷するには

- ビューアをアクティブ ウィンドウにします（ウィンドウ内の任意の場所をクリックします）。
- メニューから次の項目を選択します。

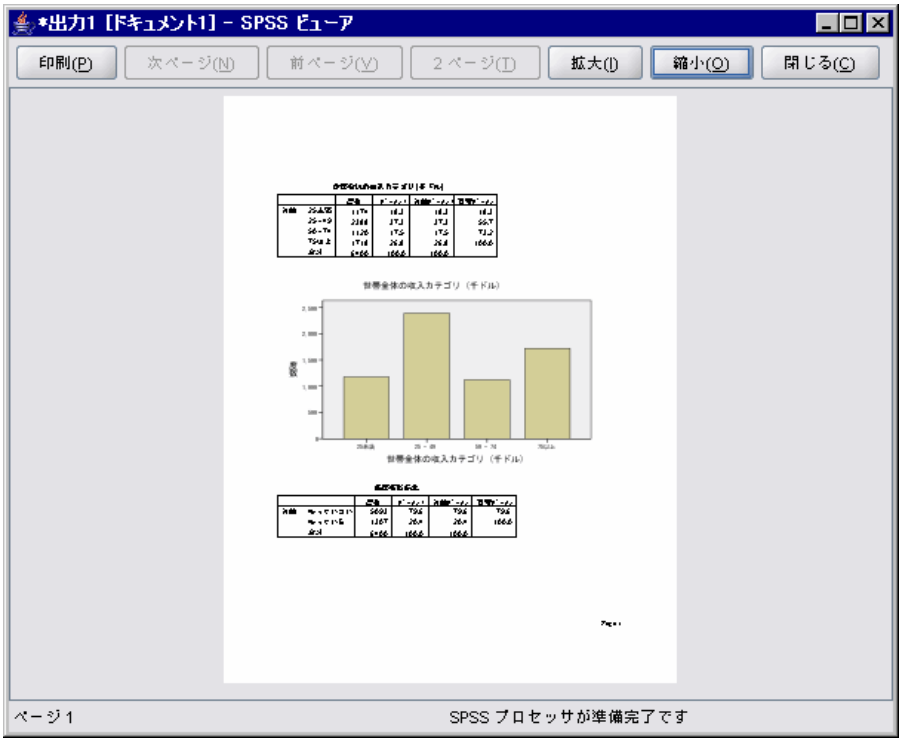
  ファイル
  印刷...
- 希望する印刷設定を選択します。
- [OK] をクリックして印刷します。

印刷プレビュー

[印刷プレビュー] は、ビューア文書の各ページに印刷されるものを表示します。ビューア文書を実際に印刷する前に [印刷プレビュー] でチェックするとよいでしょう。次のような項目はビューアの内容枠には表示されていない場合がありますが、[印刷プレビュー] ではそれが確認できるからです。

- 改ページ
- ピボット テーブルの隠れた層
- 幅の広いテーブルでのブレーク
- 各ページに印刷されるヘッダーとフッター

図 10-7
印刷プレビュー



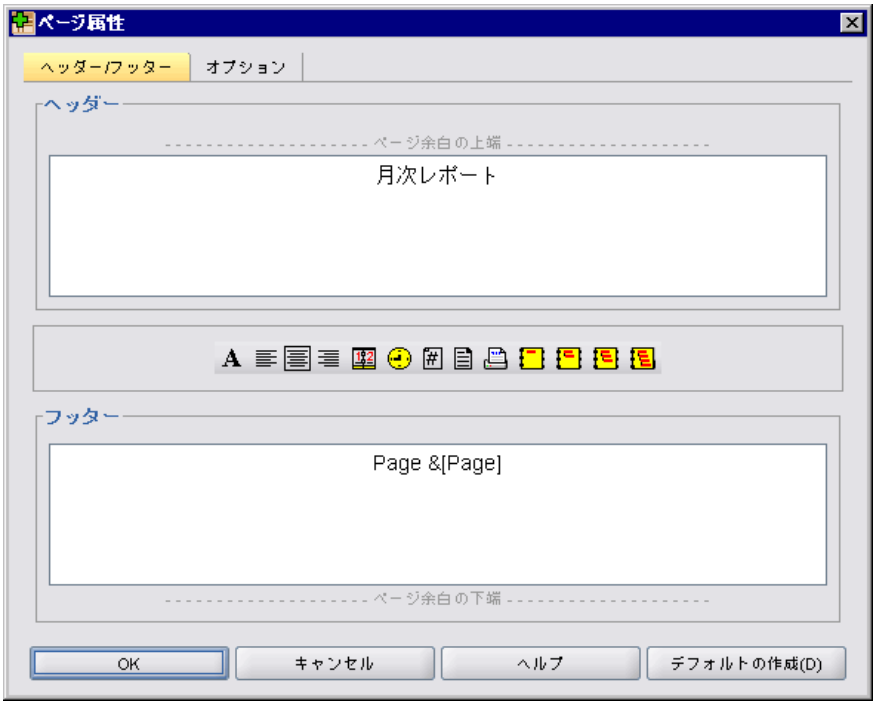
ビューアで選択している出力がある場合は、プレビューでは選択した出力しか表示できません。すべての出力のプレビューを表示する場合は、ビューアで何も選択されていないことを確認してください。

ページ属性: ヘッダーとフッター

ヘッダーとフッターは、各ページの上端と下端に印刷する情報です。どんなテキストでもヘッダーとフッターとして入力することができます。ダイアログ ボックスの中央にあるツールバーを使用すると、次の項目を挿入することもできます。

- 日付と時刻
- ページ番号
- ビューアのファイル名
- アウトラインの見出しラベル
- ページの表題と副題

図 10-8
[ページ属性] ダイアログ ボックス、[ヘッダー/フッター] タブ



- **[デフォルトの作成]** は、新規 [ビューア] 文書のデフォルト設定として指定した設定を使用します。
- アウトラインの見出しラベルは、各ページの最初の項目が第 1、第 2、第 3、第 4 のいずれのレベルのアウトライン見出しであるかを示します。
- ページの表題と副題には、現在のページの表題と副題が表示されます。これらを作成するには、[ビューア挿入] メニューの [新しいページ表題] を使用するか、または TITLE および SUBTITLE コマンドを使用します。ページの表題または副題を指定していない場合、この設定は無視されます。

  注: 新しいページの表題および副題のフォント特性は、[オプション] ダイアログ ボックス ([編集] メニュー) の [ビューア] タブ (アクセスするには、[編集] メニューで [オプション] を選択) 上で設定します。既存のページの表題および副題のフォント特性を変更する場合には、ビューアで表題や副題を編集します。

ヘッダーとフッターがページ上にどのように印刷されるかを確認するには、[ファイル] メニューの [印刷プレビュー] を選択します。

ページのヘッダーとフッターを挿入するには

▶ ビューアをアクティブ ウィンドウにします (ウィンドウ内の任意の場所をクリックします)。

▶ メニューから次の項目を選択します。

ファイル
ページ属性...

▶ [**ヘッダー/フッター**] タブをクリックします。

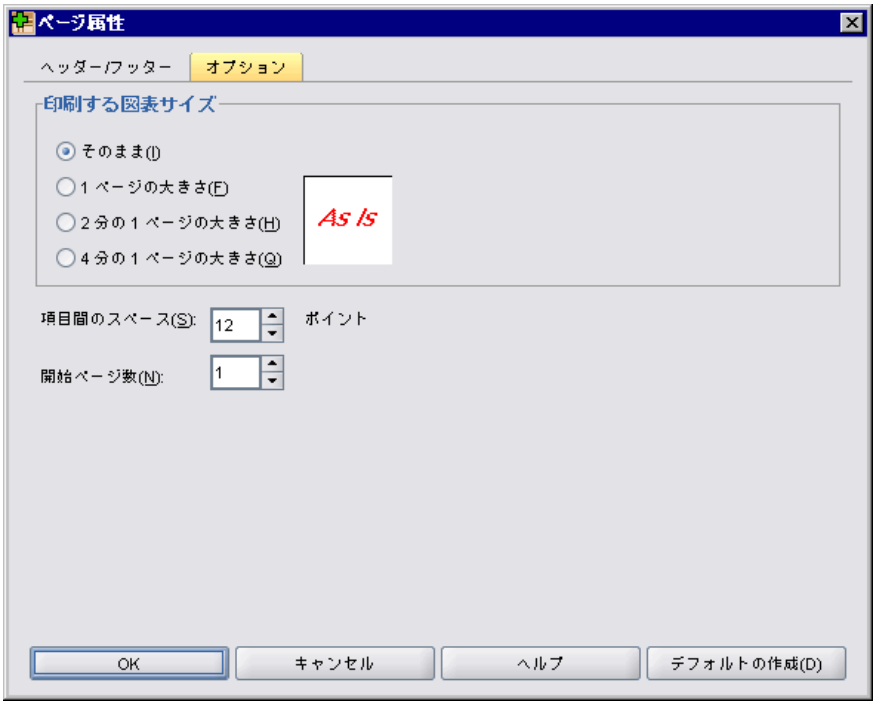
▶ 各ページに表示するヘッダーやフッターを入力します。

ページ属性: オプション

このダイアログ ボックスでは、図表の印刷サイズ、印刷する出力項目間のスペース、およびページの番号付けを制御します。

- **印刷された図表サイズ。**定義されたページの大きさに対する図表の印刷サイズを制御します。図表の縦横比 (幅対高さの比) は、図表の印刷サイズにかかわらず一定です。図表の全体的な印刷サイズは、その高さと幅両方で制限されます。図表の外側の枠がページの左右の枠限界にいったん達すると、ページ高さの方向もそれ以上は拡張されなくなります。
- **項目間のスペース。**印刷される項目間のスペースを制御します。各ピボットテーブル、図表、およびテキスト オブジェクトは個別の項目です。この設定は、ビューアでの項目の表示には影響しません。
- **開始ページ数。**指定した番号から始まって順番にページ番号を付けます。
- **デフォルトの作成。**このオプションは、新規 [ビューア] 文書のデフォルト設定として指定した設定を使用します。

図 10-9
[ページ属性] ダイアログ ボックス、[オプション] タブ



図表の印刷サイズ、ページの番号付け、印刷項目間のスペースを変更するには

- ビューアをアクティブ ウィンドウにします (ウィンドウ内の任意の場所をクリックします)。
- メニューから次の項目を選択します。

  ファイル
  ページ属性...
- [**オプション**] タブをクリックします。
- 設定を変更して、[OK] をクリックします。

出力の保存

ビューアの内容はビューア文書として保存することができます。保存される文書にはビューアの両方の枠、アウトライン枠と内容枠が含まれます。

ビューア文書を保存するには

▶ [ビューア] ウィンドウのメニューから、次の項目を選択します。

ファイル
データの上書き保存

▶ 文書名を入力して、**[保存]** をクリックします。

外部書式 (HTML やテキスト書式など) で結果を保存するには、[ファイル] メニューの [エクスポート] を使用します。

章
11

ピボット テーブル

多くの結果は、インタラクティブにピボットできるテーブルで提示されます。つまり、行、列、および層の配置を変更することができます。

ピボット テーブルの操作

ピボット テーブルを操作するためのオプションには、次のものがあります。

- 行と列の入れ替え
- 行と列の移動
- 多次元層の作成
- 行と列のグループ化と非グループ化
- 行、列、および他の情報を表示および非表示
- 行ラベルおよび列ラベルの回転
- 項目の定義の取得

ピボット テーブルをアクティブにする

ピボット テーブルをアクティブにしたり、または修正する前に、テーブルを **アクティブ** にする必要があります。テーブルをアクティブにするには次のようにします。

▶ テーブルをダブルクリックします。

または

▶ テーブルを右クリックし、コンテキスト メニューから [**内容の編集**] を選択します。

▶ サブメニューから、[**ビューア**] または [**別のウィンドウ**] を選択します。

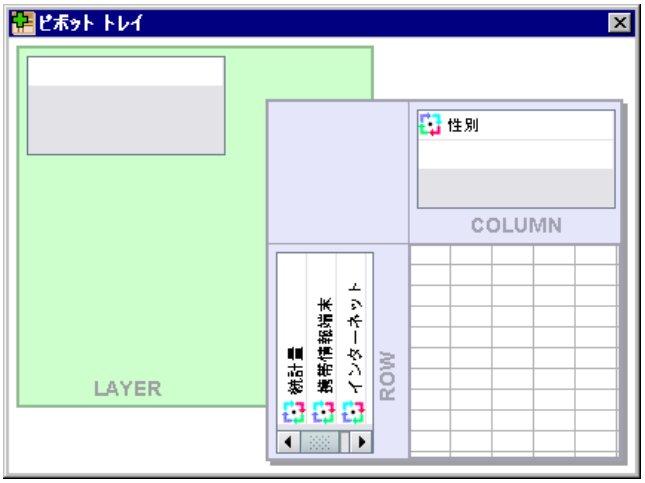
- デフォルトでは、ダブルクリックしてテーブルをアクティブ化すると、ビューア ウィンドウの非常に大きなテーブル以外のすべてがアクティブ化します。詳細は、509 ページの 45 章の[ピボット テーブル] オプションを参照してください。
- 同時に 1 個以上のピボット テーブルをアクティブ化したいなら、別のウィンドウでテーブルをアクティブ化する必要があります。

テーブルをピボットする

- ピボット テーブルをアクティブにします。
- メニューから次の項目を選択します。

  ピボット
  ピボット トレイ

図 11-1
ピボット トレイ



テーブルは、行、列および層の 3 つの次元があります。次元は複数の要素を持つことができます (またはまったくなし)。次元内、または次元間の要素を移動させてテーブル組織を変更できます。要素を移動するには、望む位置にドラッグ アンド ドロップしてください。

次元内の要素の表示順を変更

テーブルの次元内の要素 (行、列、または層) の表示順序を変更するには、次を実行します。

- ピボット トレイがまだオンになっていない場合、[ピボット テーブル] メニューから次の項目を選択します。

  ピボット
  ピボット トレイ

- [ピボット トレイ] の次元内の要素をドラッグ アンド ドロップします。

次元要素内の行と列を移動

- テーブル一覧 (ピボット トレイではなく) 内では、移動したい行と列のラベルをクリックします。

▶ ラベルを新しい位置へドラッグします。

▶ コンテキスト メニューから、**[前に挿入]** または **[入れ替え]** を選択します。

注: [編集] メニューの [ドラッグしてコピー] がオンになって**いないか** (チェックマークが付いていないか) 確認します。[ドラッグしてコピー] がオンになっている場合は、オフにします。

行と列の入れ替え

行と列を入れ替えたいときは、ピボット トレイを使用する簡単な代替方法があります。

▶ メニューから次の項目を選択します。

ピボット
行と列の入れ替え

この操作は、行要素のすべてを列次元にドラッグし、列要素のすべてを行次元にドラッグするのと同じ効果があります。

行または列のグループ化

▶ グループ化したい行または列のラベルを選択します (クリック アンド ドラッグするか、または複数のラベルを選択するには Shift キーを押しながらクリックします)。

▶ メニューから次の項目を選択します。

編集
グループ化

グループ ラベルは自動的に挿入されます。グループ ラベルをダブルクリックしてラベル テキストを編集します。

図 11-2
行と列のグループとラベル

		グループ ラベル		
		女性	男性	合計
グループ ラベル	一般職	206	157	363
	管理職	0	27	27
	役員	10	74	84

注: 既存のグループに行または列を追加するには、現在グループ内にある項目をいったんグループ解除してください。それから追加の項目を含む新しいグループを作成しなければなりません。

行または列のグループ解除

- ▶ グループ解除する列または行のグループ ラベルの好きな場所をクリックします。
- ▶ メニューから次の項目を選択します。

編集
グループ解除

グループ解除では、グループ ラベルが自動的に削除されます。

行ラベルおよび列ラベルの回転

テーブルの最も内側の列ラベルと最も外側の行ラベルの垂直および水平表示間のラベルを回転できます。

- ▶ メニューから次の項目を選択します。

フォーマット
列内側のラベルを回転

または

フォーマット
列外側のラベルを回転

図 11-3
回転した列ラベル

		性別		
		女性	男性	合計
職種	一般職	206	157	363
	管理職	0	27	27
	役員	10	74	84
合計		216	258	474

		性別		
		女性	男性	合計
職種	一般職	206	157	363
	管理職	0	27	27
	役員	10	74	84
合計		216	258	474

回転できるのは、最も内側の列ラベルと最も外側の行ラベルだけです。

層を使用した作業

各カテゴリまたはカテゴリの結合に個別の 2 次元テーブルを表示することもできます。テーブルは、一番上の層だけが見えるようにして、層で重なっていると考えることができます。

層の作成と表示

層を作成するには次のようにします。

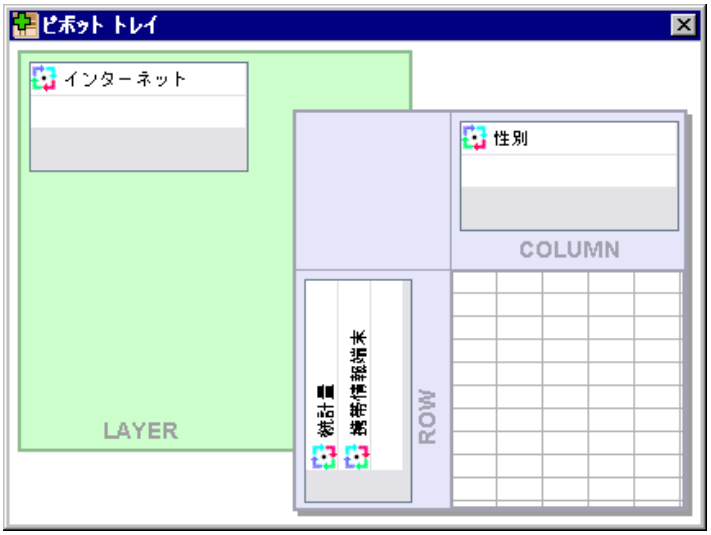
▶ ピボット テーブルをアクティブにします。

▶ ピボット トレイがまだオンになっていない場合、[ピボット テーブル] メニューから次の項目を選択します。

ピボット
ピボット トレイ

▶ 行列次元を層次元に要素をドラッグします。

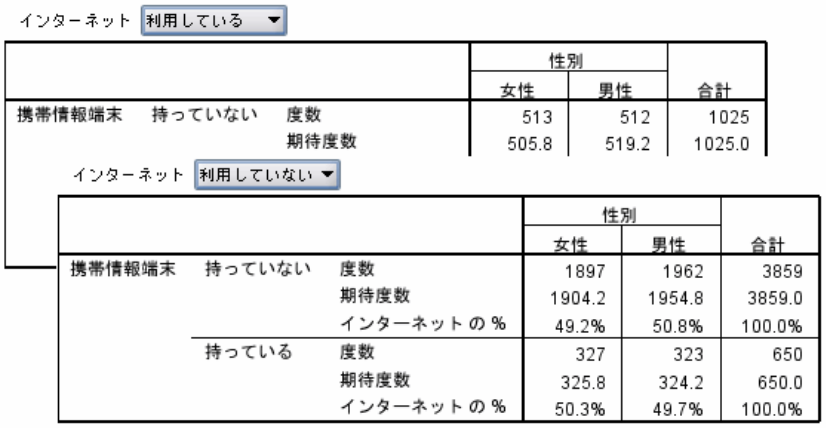
図 11-4
カテゴリを層へ移動



層次元に要素を移動すると多次元テーブルが作成され、単一の 2 次元分割だけが表示されます。見えるテーブルは、一番上の層のテーブルです。たとえば、は

い/いいえカテゴリ変数が層次元にあるなら、多次元テーブルには 2 層があり、1 個は **はい** カテゴリで、もう 1 個は **いいえ** カテゴリです。

図 11-5
個別層内のカテゴリ



インターネット 利用している

			性別		
			女性	男性	合計
携帯情報端末	持っていない	度数	513	512	1025
		期待度数	505.8	519.2	1025.0

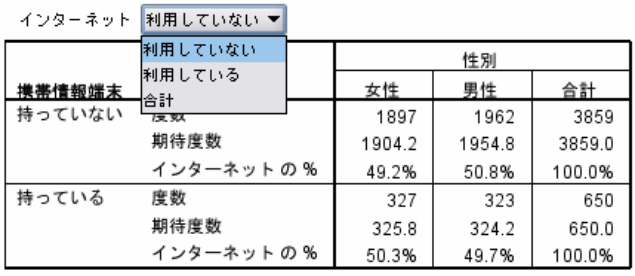
インターネット 利用していない

			性別		
			女性	男性	合計
携帯情報端末	持っていない	度数	1897	1962	3859
		期待度数	1904.2	1954.8	3859.0
		インターネット の %	49.2%	50.8%	100.0%
	持っている	度数	327	323	650
		期待度数	325.8	324.2	650.0
		インターネット の %	50.3%	49.7%	100.0%

表示済み層の変更

▶ 層のドロップダウン リストからカテゴリを選択します (ピボット テーブルではなく、ピボット トレイ)。

図 11-6
ドロップダウン リストから層を選択



インターネット 利用していない

利用していない
利用している
合計

携帯情報端末		性別		
		女性	男性	合計
持っていない	度数	1897	1962	3859
	期待度数	1904.2	1954.8	3859.0
	インターネット の %	49.2%	50.8%	100.0%
持っている	度数	327	323	650
	期待度数	325.8	324.2	650.0
	インターネット の %	50.3%	49.7%	100.0%

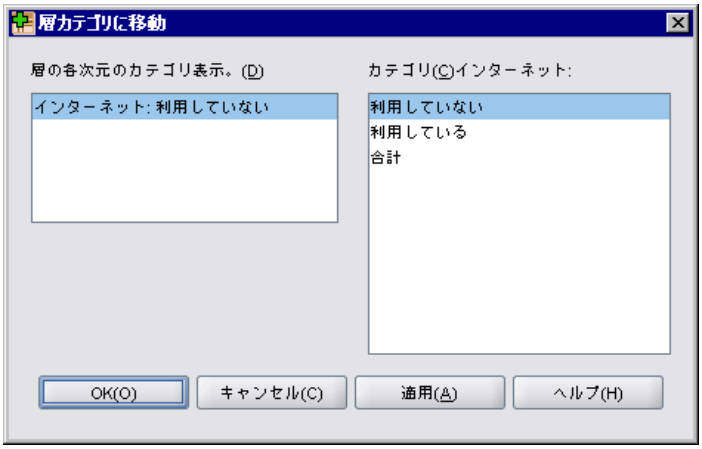
層カテゴリに移動

[層カテゴリに移動] では、ピボット テーブルで層を変更することができます。このダイアログ ボックスは、数多くの層があるとき、または 選択した層に多くのカテゴリがあるときに、特に有効です。

▶ メニューから次の項目を選択します。

ピボット
指定の層へ...

図 11-7
[層カテゴリに移動] ダイアログ ボックス



- ▶ 表示カテゴリ リストで層次元を選択します。カテゴリ リストは、選択した次元のすべてのカテゴリを表示します。
- ▶ カテゴリ一覧では、望むカテゴリを選択し、[OK] または[適用] をクリックします。

項目の表示と非表示

次のセルを隠すことができます。

- 次元ラベル
- 行または列内のラベル セルとデータ セルを含むカテゴリ
- カテゴリ ラベル (データ セルは表示)
- 脚注、表題、および解説

テーブルの行と列を非表示にする

- ▶ 非表示にしたい行または列のカテゴリ ラベルを右クリックします。
- ▶ コンテキスト メニューから次の項目を選択します。

選択
データとラベルのセル

- ▶ カテゴリ ラベルを再度右クリックし、コンテキスト メニューから [**カテゴリ非表示**] を選択します。

または

- ▶ コンテキスト メニューから [**非表示**] を選択します。

テーブルの行と列を表示する

- ▶ 非表示の行と列と同様同じ次元の他の行と列を右クリックします。
- ▶ コンテキスト メニューから次の項目を選択します。

  選択
  データとラベルのセル
- ▶ メニューから次の項目を選択します。

  表示
  [次元変数名] のすべてのカテゴリを表示

  または
- ▶ アクティブしたピボット テーブルのすべての非表示行と列を表示するには、メニューから次の項目を選択します。

  表示
  すべて表示

  テーブル内の非表示の行と列が表示されます。(このテーブルに対してテーブル プロパティで **[空白の行と列を隠す]** が選択されていると、完全に空白の行または列が隠れたままになります)。

次元ラベル表示と非表示

- ▶ 次元ラベルまたは次元内のカテゴリ ラベルを選択します。
- ▶ [ビュー] メニューまたはコンテキスト メニューから **[次元ラベル非表示]** または **[次元ラベル表示]** を選択します。

テーブル表題の表示と非表示

凡例の表題を変更するには

- ▶ 表題の選択。
- ▶ コンテキスト メニューから **[非表示]** を選択します。

非表示の表題を表示するには

- ▶ ビュー メニューから **[すべてを表示]** を選択します。

テーブルルック

テーブルルックは、テーブルの外観を定義した 1 セットのプロパティです。すでに定義されているテーブルルックを選択することもできますが、ユーザー独自のテーブルルックを作成することもできます。

- テーブルルックを適用する前または後で、セル プロパティを使用して個別セルまたはグループのセルの書式を変更することができます。編集したセルの書式は、新しいテーブルルックを適用するときでも残っています。 詳細は、266 ページの セル プロパティを参照してください。
- オプションとして、すべてのセルを、現在のテーブルルックで定義されたセルの書式に戻すことができます。こうすると、編集されているすべてのセルが戻ります。テーブルルック ファイルのリスト内の **[表示のまま]** を選択すると、編集されたセルはどれも現在のテーブル プロパティに戻ります。

注:SPSS の以前のバージョンで作成されたテーブルルックは、バージョン 16.0 以降では使用できません。

テーブルルックを適用または保存するには

▶ ピボット テーブルをアクティブにします。

▶ メニューから次の項目を選択します。

書式
テーブルルック...

図 11-8
[テーブルルック] ダイアログ ボックス

▶ ファイルのリストからテーブルルックを選択します。別のディレクトリのファイルを選択するには、**[参照]** をクリックします。

▶ [OK] をクリックして、選択したピボット テーブルにテーブルルックを適用します。

テーブルルックを編集または作成するには

- [テーブルルック] ダイアログ ボックスで、ファイルのリストからテーブルルックを選択します。
- [**編集**] をクリックします。
- 希望する属性のテーブル プロパティを調整して [OK] をクリックします。
- 編集したテーブルルックを保存するには [**保存**] を、新規のテーブルルックとして保存するには [**別名で保存**] をクリックします。

  テーブルルックを編集した場合、選択したピボット テーブルにだけ影響します。編集したテーブルルックは、テーブルを選択してテーブルルックを再適用しない限り、そのテーブルルックを使用している他のどのテーブルにも適用されません。

テーブル プロパティ

[テーブル プロパティ] を使用すると、テーブルの全般的なプロパティの設定、テーブルの各部に対するセル スタイルの設定、および一連のプロパティをテーブルルックとして保存することができます。次のようにできます。

- 空白の行または列を隠す、印刷プロパティを調整するなどの一般的なプロパティをコントロールできます。
- 脚注マーカーの書式と位置をコントロールできます。
- データ領域内のセル、行と列ラベル、およびテーブルの他の領域に対して特定の書式を決定できます。
- テーブルの各領域を区切る線の幅と色をコントロールできます。

ピボット テーブルのプロパティを変更するには

- ピボット テーブルをアクティブにします。
- メニューから次の項目を選択します。

  書式
  テーブル プロパティ...
- タブ ([**全般**]、[**脚注**]、[**セル書式**]、[**罫線**]、または [**印刷**]) を選択します。
- 希望するオプションを選択します。
- [OK] または [**適用**] をクリックします。

  選択したピボット テーブルに新しいプロパティが適用されます。選択したテーブルの代わりに、テーブルルックに新しいテーブルのプロパティを適用するには、テーブルルックを編集します ([書式] メニュー、[テーブルルック])。

テーブル プロパティ: 全般

いくつかのプロパティは、全体としてテーブルに適用します。次のようにできます。

- 空白の行と列を表示したり、隠したりします。(空白の行または列では、データ セル内に何も入っていません)。
- 列ラベルの位置を制御します。列ラベルは、左上のコーナーに置くこともできますが、入れ子にすることもできます。
- 最大および最小列幅 (ポイント単位で表現) をコントロールします。

図 11-9
[テーブル プロパティ] ダイアログ ボックスの [全般] タブ

全般的なテーブルプロパティを変更するには

▶ [全般] タブをクリックします。

▶ 希望するオプションを選択します。

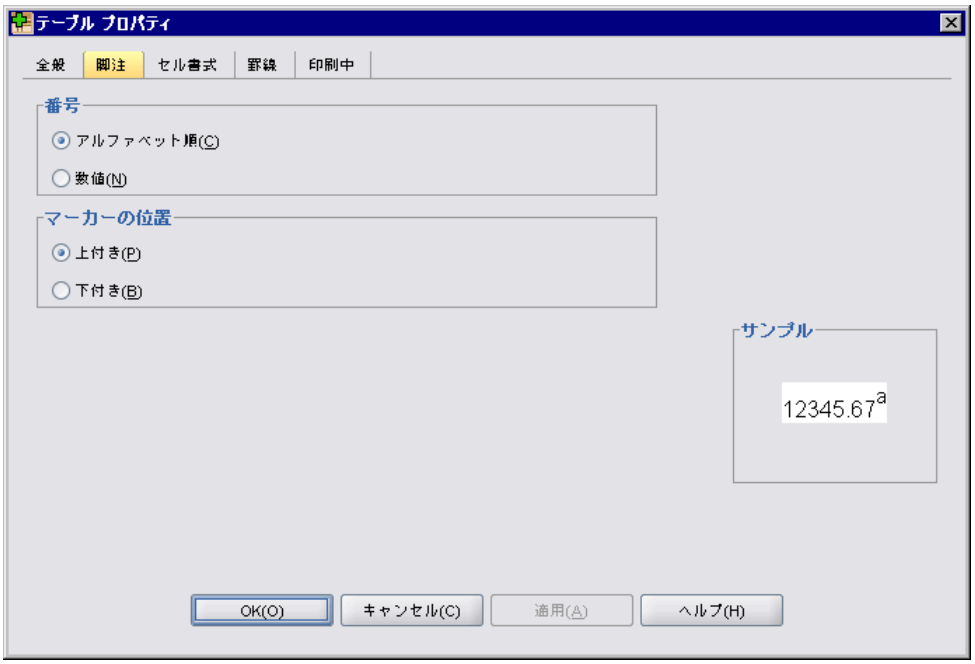
▶ [OK] または [適用] をクリックします。

テーブル プロパティ: 脚注

脚注マーカーのプロパティには、テキストに関連するスタイルおよび位置があります。

- 脚注マーカーのスタイルは、数字 (1、2、3...) または英字 (a、b、c...) のどちらかです。
- 脚注マーカーは、上付き文字または下付き文字としてテキストに添付することができます。

図 11-10
[テーブル プロパティ] ダイアログ ボックスの [脚注] タブ



脚注マーカーのプロパティを変更するには

▶ [脚注] タブをクリックします。

▶ 脚注番号の書式を選択します。

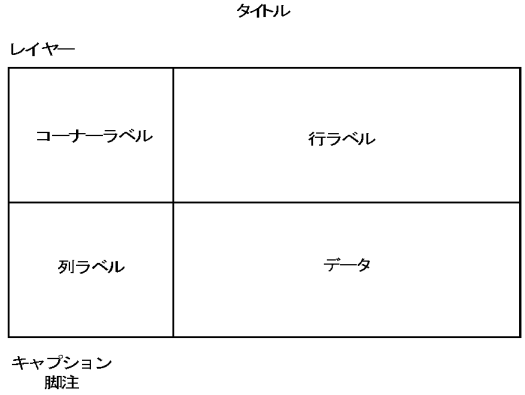
▶ マーカーの位置を選択します。

▶ [OK] または [適用] をクリックします。

テーブル プロパティ: セル書式

書式設定では、1 つのテーブルは、表題、層、コーナー ラベル、行ラベル、列ラベル、データ、解説、および脚注という領域に分割されます。テーブルの各領域に対して、関連するセルの書式を修正することができます。セルの書式には、テキストの特性 (フォント、大きさ、色、スタイルなど)、水平と垂直の位置合わせ、背景の色、および内側セルの余白などがあります。

図 11-11
テーブルの領域



セル書式は、領域に適用されます (情報のカテゴリ)。領域は個々のセルの特性を表すものではありません。この区別は、テーブルを変更するときに、重要になります。

例をあげると、次のようになります。

- 列ラベルのセル形式として太字を指定すると、列次元にどのような情報が現在表示されていても列ラベルは太字で表示されます。項目を列次元から他の次元に移動すると、列ラベルの太字文字は保持できません。
- アクティブなピボット テーブルでセルを強調表示して、ツールバーの [太字] ボタンをクリックして列ラベルを太字にすると、セルの内容をどの次元に移動するかに関係なく、セルの内容は太字のままになっていて、列ラベルは、列次元に移動された他の項目には太字特性を保持しません。

図 11-12
[テーブル プロパティ] ダイアログ ボックスの [セル書式] タブ

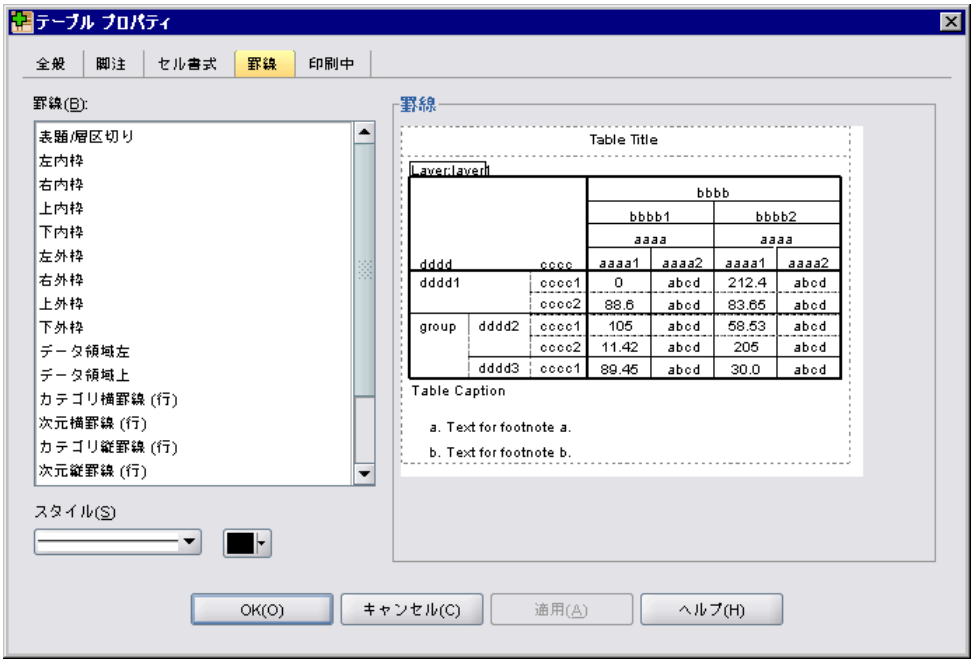
セル書式を変更するには

- [セル書式] タブを選択します。
- ドロップダウン リストから 1 つの領域を選択するか、またはサンプル内の領域をクリックします。
- 領域の特性を選択します。選択した内容はサンプルに反映されます。
- [OK] または [適用] をクリックします。

テーブル プロパティ: 罫線

テーブル内の各罫線位置に対して、線の種類および色を選択することができます。種類として [なし] を選択すると、選択した位置には線が表示されません。

図 11-13
[テーブル プロパティ] ダイアログ ボックスの [罫線] タブ



テーブル罫線を変更するには

- ▶ [罫線] タブをクリックします。
- ▶ リストにある名前をクリックするか、またはサンプル領域にある行をクリックして、罫線の位置を選択します。
- ▶ 線の種類または [なし] を選択します。
- ▶ 色を選択します。
- ▶ [OK] または [適用] をクリックします。

テーブル プロパティ: 印刷

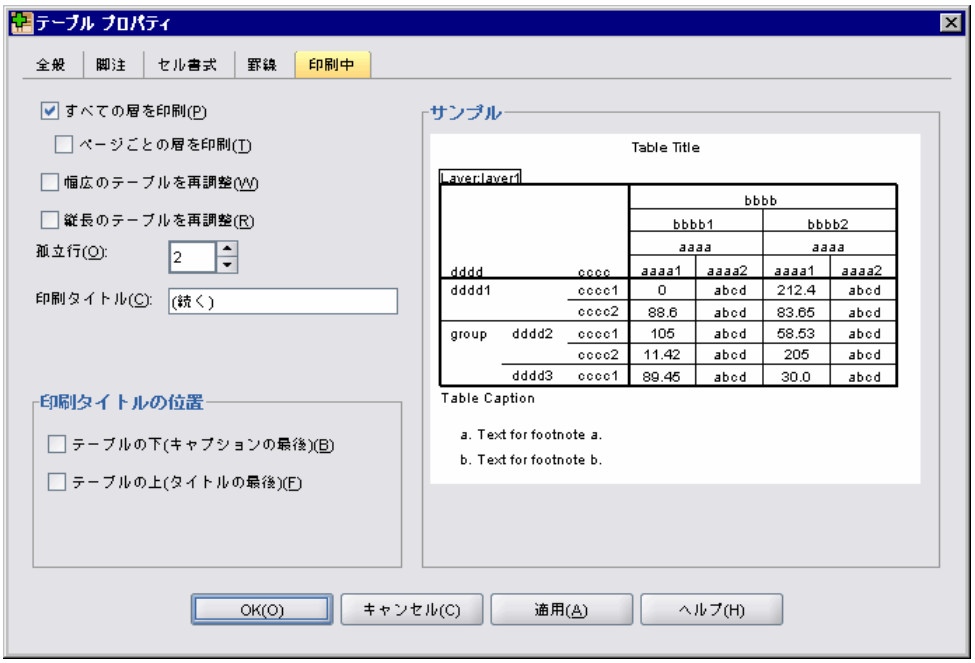
ピボット テーブルの印刷に関する次のプロパティをコントロールできます。

- すべての層またはテーブルの最上層だけを印刷することができます。また、1 ページに 1 層を印刷することもできます。
- 印刷するページに合うようにテーブルを水平または垂直方向に縮小できます。
- テーブルが定義されたページサイズには広すぎるか長すぎる場合、テーブルの印刷部分に取り入れられる行と列の最小数である改ページ後の行や列数をコントロールできます。

(注: ページの上の部分にすでに他の出力があり、現在のページにはめ込むにはテーブルが長すぎるが、定義されたページ長全体には入る大きさである場合、このテーブルは、改ページ後の行/列数の設定に関わりなく自動的に新しいページに印刷されます)。

- 1 ページに入らないテーブルにも印刷タイトルを付けることができます。各ページの下端と各ページの上端に印刷タイトルをつけることができます。どちらも選択されなかった場合には、印刷タイトルは表示されません。

図 11-14
[テーブル プロパティ] ダイアログ ボックスの [印刷] タブ



ピボット テーブルの印刷プロパティをコントロールするには

- ▶ [印刷] タブをクリックします。
- ▶ 希望する印刷オプションを選択します。
- ▶ [OK] または [適用] をクリックします。

セル プロパティ

セル プロパティは、選択した個別のセルに適用されます。値書式、位置合わせ、余白、および色を変更することができます。セルのプロパティは、テーブルのプロパティを無効にするので、テーブルのプロパティを変更しても、個別に適用したセルのプロパティは変更されません。

セル プロパティを変更するには

- テーブルをアクティブにして、テーブル内のセルを選択します。
- 書式メニューまたはコンテキスト メニューから [セル プロパティ] を選択します。

フォントおよび背景

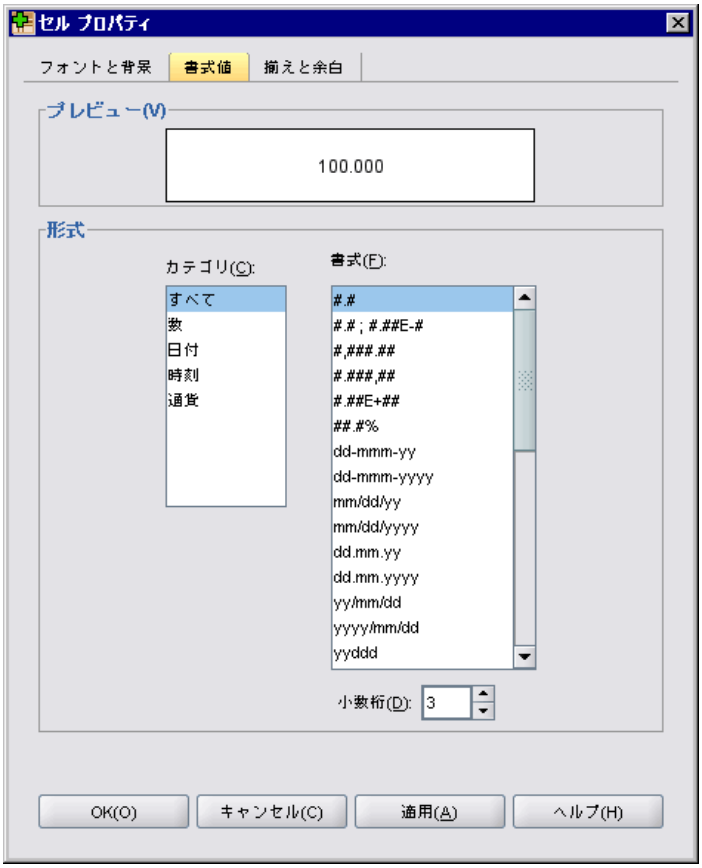
[フォントと背景色] タブはテーブルの選択したセルのフォントのスタイルおよび色および背景色をコントロールします。

図 11-15
[セル プロパティ] ダイアログ ボックス、[フォントと背景色] タブ

形式値

[形式値] タブは選択したセルの値書式をコントロールします。数字、日付、時刻、または通貨の書式を選択でき、さらに表示される小数の桁数を調整できます。

図 11-16
[セル プロパティ] ダイアログ ボックス、[形式値] タブ



位置合せおよび余白

[位置合わせおよび余白] タブは選択したセルの値の水平および垂直位置合わせと上部、下部、左および右余白をコントロールします。**混合** 水平位置合わせは各セルの内容をその型により位置合わせします。たとえば、日付は右に位置合わせされ、テキスト値は左に位置合わせされます。

図 11-17
[セル プロパティ] ダイアログ ボックス、[揃えと余白] タブ

脚注と解説

脚注と解説をテーブルに追加できます。脚注と解説を非表示にでき、脚注マーカーを変更し、脚注の番号をつけ直せます。

ある脚注属性はテーブル プロパティによりコントロールされます。 詳細は、261 ページのテーブル プロパティ: 脚注を参照してください。

脚注と解説を追加

テーブルに解説を追加するには

- 挿入メニューから [解説] を選択します。

脚注は、テーブル内のどの項目にも付加することができます。脚注を追加するには

- アクティブなピボット テーブル内で表題、セル、または解説をクリックします。

▶ 挿入メニューから [**脚注**] を選択します。

解説を表示か非表示にするには

解説を非表示にするには

▶ 解説を選択します。

▶ コンテキスト メニューから [**非表示**] を選択します。

非表示の解説を表示するには

▶ ビュー メニューから [**すべてを表示**] を選択します。

テーブル内の脚注を隠す、または表示するには

脚注を非表示にするには

▶ 脚注を選択します。

▶ [ビュー] メニューから [**非表示**] またはコンテキスト メニューから [**脚注を非表示**] を選択します。

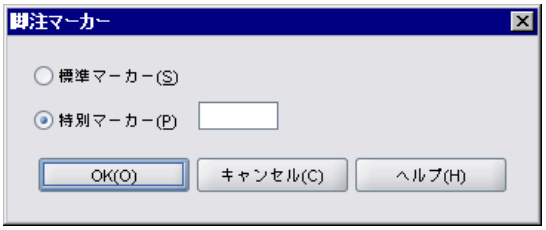
非表示の脚注を表示するには

▶ [ビュー] メニューから [**すべての脚注を表示**] を選択します。

脚注マーカー（特殊）

[脚注マーカー] は、脚注をマークするために使用する文字を変更します。

図 11-18
[脚注マーカー] ダイアログ ボックス



脚注マーカーを変更するには

▶ 脚注を選択します。

▶ 形式メニューから [**脚注マーカー**] を選択します。

▶ 1 つまたは 2 つの文字を入力します。

脚注番号のつけ直し

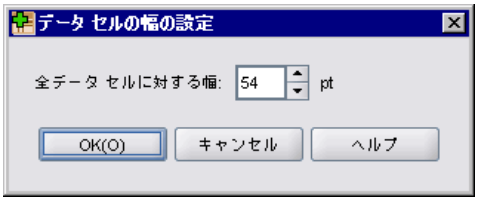
行、列、および層を切り替えてテーブルを回転させると、脚注の順序が変化することもあります。脚注の番号をつけ直すには、次を実行します。

▶ 形式メニューから [**脚注の番号の付け直し**] を選択します。

データセルの幅

すべてのデータセルを同じ幅に設定するときは、[データ セルの幅の設定] ダイアログ ボックスを使用します。

図 11-19
[データ セルの幅の設定] ダイアログ ボックス



すべてのデータ セルの幅を設定するには

▶ メニューから次の項目を選択します。
書式
データ セルの幅の設定...

▶ セル幅として値を入力します。

列の幅の変更

▶ 列の罫線をクリックしてドラッグします。

ピボット テーブル内の隠れた罫線を表示

見えている罫線が少ないテーブルでは、隠れた罫線を表示することができます。こうすると、列幅の変更のような作業を簡単に行うことができます。

▶ [ビュー] メニューから [**グリッド線**] を選択します。

図 11-20
隠れた罫線の表示されたグリッド線

インターネット				性別		
				女性	男性	合計
利用していない	携帯情報端末	持っていない	度数	1897	1962	3859
			期待度数	1903.4	1955.6	3859.0
			携帯情報端末の %	49.2%	50.8%	100.0%
		持っている	度数	327	323	650
			期待度数	320.6	329.4	650.0
	合計					
利用している	携帯情報端末					
	合計					

インターネット				性別		
				女性	男性	合計
利用していない	携帯情報端末	持っていない	度数	1897	1962	3859
			期待度数	1903.4	1955.6	3859.0
			携帯情報端末の %	49.2%	50.8%	100.0%
		持っている	度数	327	323	650
			期待度数	320.6	329.4	650.0
			携帯情報端末の %	50.3%	49.7%	100.0%
		合計	度数	2224	2285	4509
			期待度数	2224.0	2285.0	4509.0
			携帯情報端末の %	49.3%	50.7%	100.0%
利用している	携帯情報端末	持っていない	度数	513	512	1025
			期待度数	512.5	512.5	1025.0
			携帯情報端末の %	50.0%	50.0%	100.0%
		持っている	度数	305	306	611
			期待度数	305.5	305.5	611.0
			携帯情報端末の %	49.9%	50.1%	100.0%
		合計	度数	818	818	1636
			期待度数	818.0	818.0	1636.0
			携帯情報端末の %	50.0%	50.0%	100.0%

ピボット テーブルの行と列を選択

ピボット テーブルでは、行と列全体の選択方法、および選択した行または列を示す強調表示がテーブルの隣接していない領域にまたがるなどのいくつかの制約があります。列と行の全体を選択するには

- 行または列ラベルをクリックします。
- メニューから次の項目を選択します。

編集
選択
データとラベルのセル

または

- 行または列のカテゴリ ラベルを右クリックします。
- コンテキスト メニューから次の項目を選択します。

選択
データとラベルのセル

または

▶ 列または行を、Ctrl キーと Alt キーを押しながらクリックします。

ピボット テーブルの印刷

ピボット テーブルの内容を印刷する場合、テーブルの外観に影響する要素がいくつかあります。ピボット テーブルの属性を変更することでこれらの要素をコントロールできます。

- 多次元ピボット テーブル (複数の層のあるテーブル) では、すべての層を印刷するか、最上層 (見えている層) だけを印刷または選択できます。 詳細は、265 ページのテーブル プロパティ: 印刷を参照してください。
- 縦長のまたは幅の広いピボット テーブルでは、ページに適合するようにテーブルのサイズを自動的に変更することができます。また、テーブルを分割する場合の位置およびページ分割の位置を制御することができます。 詳細は、265 ページのテーブル プロパティ: 印刷を参照してください。
- 1 ページでは広すぎたり長すぎたりするテーブルでは、ページ間のテーブル分割をコントロールできます。

[ファイル] メニューの [印刷プレビュー] を使用して、印刷するピボット テーブルの外観を画面上で見ることができます。

テーブル幅または長さによる分割制御

広すぎるかまたは長すぎるかのどちらかの理由で定義されたページ サイズ内に印刷できないピボット テーブルは、自動的に分割され複数の部分に分けて印刷されます。(幅の広いテーブルの場合、余白に余裕があれば複数のセクションは同じページ上に印刷されます)。次のようにできます。

- 大きなテーブルを分割する場合、行と列の位置を制御できます。
- テーブルを分割する場合にまとめる必要のある行と列を指定できます。
- 定義されたページサイズに合わせて大きなテーブルを再調整することができます。

印刷するピボット テーブルの行分割および列分割を指定するには

▶ ブレークを挿入したい場所の左にある列ラベルをクリックするか、またはブレークを挿入する場所の上にある行ラベルをクリックします。

▶ メニューから次の項目を選択します。

書式
ここで区切る

1 つにまとめたい行または列を指定するには

- まとめておきたい行または列のラベルを選択します。(複数の行または列ラベルを選択するには、クリックしてドラッグするか、または Shift キーを押しながらクリックします)。
- メニューから次の項目を選択します。
  書式
  まとめる

ピボット テーブルからグラフを作成する

- ピボット テーブルをダブルクリックして、アクティブにします。
- 図表内に表示したい行、列、またはセルを選択します。
- 選択された面のどこかを右クリックします。
- コンテキスト メニューの [**グラフを作成**] をクリックし、図表の種類を選択します。

章
12
コマンド シンタックスを使用する作業

一般的なタスクを保存して自動化できるようにするための強力なコマンド言語が用意されています。コマンド言語を使用すると、メニューやダイアログ ボックスにはない機能も実行できます。

SPSS コマンドのほとんどは、SPSS のメニューやダイアログ ボックスから利用できますが、。SPSS コマンド言語を使用しなければ利用できないコマンドやオプションもあります。コマンド言語を使用すると、シンタックス ファイルにジョブを保存することもできるので、日を改めて分析を繰り返したり、自動化したジョブをバッチ機能を使用して実行したりできます。

シンタックス ファイルは、コマンドを含む単なるテキスト ファイルです。シンタックス ウィンドウを開いてコマンド入力することもできますが、次のいずれかの方法を使用して、シンタックス ファイルを容易に作成できます。

- ダイアログ ボックスからコマンド シンタックスを貼り付ける。
- 出力ログからシンタックスをコピーする。
- ジャーナル ファイルからシンタックスをコピーする。

コマンド シンタックスの詳細なリファレンス情報は、2 つの形式で提供されています。1 つは総合的なヘルプ システムに組み込まれているもの、もう 1 つは独立した PDF ファイル (『Command Syntax Reference』) です。この PDF ファイルは、[ヘルプ] メニューから参照できます。

シンタックス ウィンドウにある現在のコマンドに対する状況依存ヘルプは、F1 キーを押すと表示されます。

シンタックスの規則

セッション中にコマンド シンタックス ウィンドウからコマンドを実行すると、**インタラクティブ** モードでコマンドを実行していることになります。

インタラクティブ モードでのコマンド指定には、次の規則が適用されます。

- 各コマンドは、新しい行に指定する必要があります。コマンドは、コマンド行の任意のカラムから開始し、必要なだけの行数まで続けることができます。ただし、END DATA コマンドについては、データの最終行の次の行の先頭カラムに指定する必要があります。

- 各コマンドの末尾には必ず、コマンド終了符としてピリオドを打ちます。ただし、BEGIN DATA の場合は、インライン データが連続して指定されたものとして扱われるよう、終了符を省略するのが適切です。
- コマンド終了符は、各コマンド内の最後の非空白文字でなければなりません。
- コマンド終了符としてのピリオドがない場合、空白行がコマンド終了符として解釈されます。

注: 他のモードでのコマンド実行 (インタラクティブ セッションでの INSERT コマンドや INCLUDE コマンドを使用したコマンド ファイルの実行など) と互換性を維持するため、各行のコマンド シンタックスは、256 バイトを超えないようにしてください。

- サブコマンドはスラッシュ (/) で区切ります。ただし、コマンドの最初のサブコマンドの前のスラッシュは、通常付けても付けなくても構いません。
- 変数名は省略せずに入力します。
- テキストをアポストロフィや引用符で囲む場合は、1 行に収めます。
- ユーザーの地域やロケールの設定とは関係なく、小数点を示すにはピリオド (.) を使用します。
- ピリオドで終わる変数名があると、ダイアログ ボックスを使用して作成したコマンドにエラーが発生する可能性があります。ダイアログ ボックスではこうした変数名は設定できませんが、一般的にもこうした命名は避ける必要があります。

SPSS コマンド シンタックスは大文字と小文字を区別しません。また、コマンドのいろいろな指定に 3 文字または 4 文字の略語を使用できます。1 つのコマンドを指定するのに、行数は何行でも使用できます。次のように、空白と改行は、スラッシュ、かっこ、算術演算子の両側や変数名と変数名の間など、空白を 1 つ挿入してもよいところであれば、ほぼどこにでも追加できます。例をあげると、次のようになります。

```
FREQUENCIES
  VARIABLES=JOBCAT GENDER
  /PERCENTILES=25 50 75
  /BARCHART.
```

かつ

```
freq var=jobcat gender /percent=25 50 75 /bar.
```

上記のシンタックスは、どちらも有効で、同じ結果が生成されます。

INCLUDE ファイル

INCLUDE コマンドを介して実行されるコマンド ファイルに対し、バッチ モード シンタックスの規則が適用されます。

バッチ モードでのコマンド指定には、次の規則が適用されます。

- コマンド ファイル内のすべてのコマンドは、カラム 1 から始まる必要があります。コマンド ファイルを読みやすくするためにコマンド指定をインデントしたい場合は、先頭カラムにプラス記号 (+) またはマイナス記号 (–) を使用できます。
- 1 つのコマンドに複数行を使用する場合、後続の各行ともカラム 1 は空白である必要があります。
- コマンド終止符は省略できます。
- 1 行の長さが 256 バイトを超えてはなりません。超過部分は切り捨てられるます。

INSERT コマンドは、どちらかの規則のセットに準拠するコマンド ファイルを処理できるため、既存のコマンド ファイルで INCLUDE コマンドを使用していない場合は、INSERT コマンドを使用する必要があると考えられます。ダイアログ ボックスの選択肢をシンタックス ウィンドウに貼り付けてコマンド シンタックスを作成すると、コマンドの書式はどの操作にも適したものになります。詳細は、『Command Syntax Reference』([ヘルプ] メニューから PDF 形式で参照可能) を参照してください。

ダイアログ ボックスからのシンタックスの貼り付け

コマンド シンタックス ファイルを作成する一番簡単な方法は、SPSS のダイアログ ボックスで選択し、そのシンタックスをシンタックス ウィンドウに貼り付ける方法です。煩雑な分析では、手順のそれぞれでシンタックスを貼り付けることによって、ジョブ ファイルを作成することができますから、日を改めてその分析を繰り返したり、自動化したジョブを SPSS のバッチ機能を使用して実行したりできます。

[シンタックス エディタ] ウィンドウでは、貼り付けたシンタックスの実行、編集、シンタックス ファイルへの保存ができます。

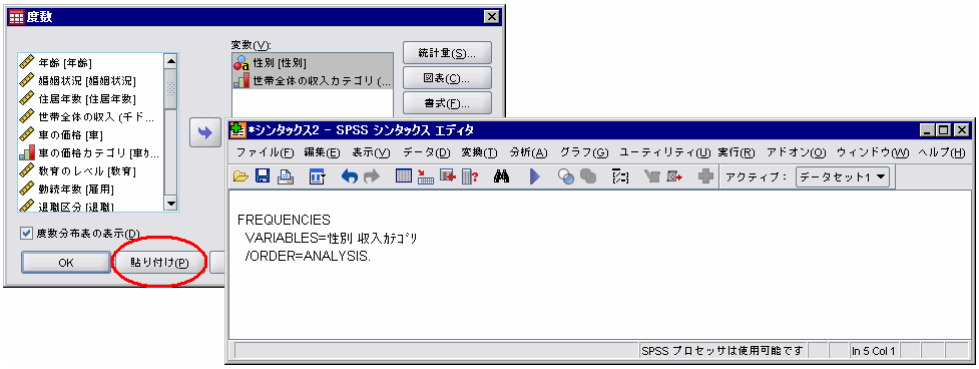
ダイアログ ボックスからシンタックスを貼り付けるには

▶ ダイアログ ボックスを開き、貼り付けたい設定を指定します。

▶ [貼り付け] をクリックします。

コマンド シンタックスが指定のシンタックス ウィンドウに貼り付けられます。[シンタックス エディタ] ウィンドウが開いていないと、SPSS が新たな [シンタックス エディタ] ウィンドウを開き、そこにシンタックスを貼り付けます。

図 12-1
ダイアログ ボックスから貼り付けたコマンド シンタックス

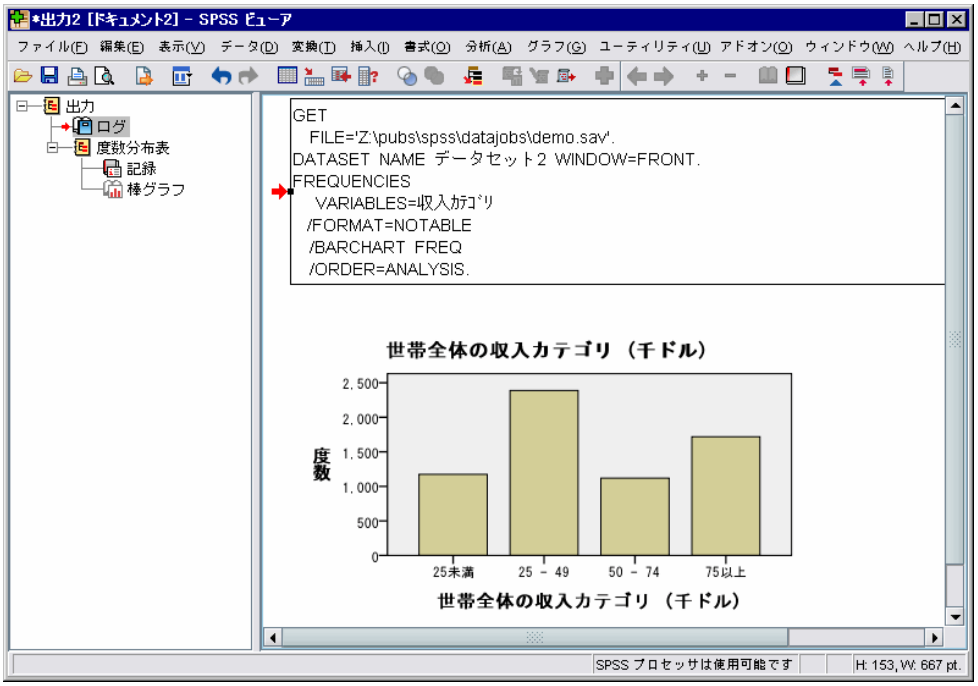


出力ログのシンタックスのコピー

シンタックス ファイルは、ビューアに表示される SPSS のログのコマンド シンタックスをコピーすると作成できます。この方法では、分析を実行する前に、SPSS のビューア設定 ([編集] メニューの [オプション] を選択し、[ビューア] タブを選択) で **[ログのなかにコマンドを表示]** を選択する必要があります。各コマンドが分析の出力と一緒にビューアに表示されます。

[シンタックス エディタ] ウィンドウでは、貼り付けたシンタックスの実行、編集、シンタックス ファイルへの保存ができます。

図 12-2
SPSS ログのコマンド シンタックス



出力ログのシンタックスをコピーするには

- 分析を実行する前に、メニューから次の項目を選択します。

  編集
  オプション...

- [ビューア] タブで、**[ログのなかにコマンドを表示]** を選択します。

  分析を実行すると、ダイアログ ボックスで選択した SPSS コマンドが SPSS ログに記録されます。

- 以前に保存したシンタックス ファイルを開くか、ファイルを新規作成します。シンタックス ファイルを新規作成するには、メニューから次の項目を選択します。

  ファイル
  新規作成
  シンタックス

- ビューアで、SPSS のログの項目をダブルクリックしてアクティブにします。

- コピーしたいテキストを選択します。

- ビューアのメニューから、次の項目を選択します。

  編集
  コピー

▶ シンタックス ウィンドウで、メニューから次の項目を選択します。

編集
貼り付け

コマンド シンタックスを実行するには

▶ 実行するコマンドをシンタックス ウィンドウで強調表示します。

▶ シンタックス エディタののツールバーで [実行] (右向きの三角形) をクリックします。

または

▶ [実行] メニューから、次の項目のうち 1 つを選択します。

- **すべて。**シンタックス ウィンドウのすべてのコマンドを実行します。
- **選択部分。**現在選択されているコマンドを実行します。部分的にハイライト表示になっているコマンドがあれば、それも含まれます。選択をしないと、カーソルの現在位置のコマンドを実行します。
- **現在の位置。**カーソルの現在位置のコマンドを実行します。
- **最後まで。**カーソルの現在位置からそのコマンド シンタックス ファイルの最後までの全コマンドを実行します。

シンタックス エディタのツールバーの [実行] をクリックすると、選択したコマンドが実行されますが、コマンドを選択していない場合は、カーソルが置かれているコマンドが実行されます。

図 12-3
[実行] ボタン

Unicode シンタックス ファイル

Unicode モードでは、セッションの間に作成または修正されたコマンド シンタックス ファイルの保存のデフォルト形式は、また Unicode (UTF-8) です。Unicode 形式コマンド シンタックス ファイルを保存すると 16.0 以前の SPSS のリリースでは読み込むことができません。Unicode モードの詳細は、「[全般] オプション」(498 ページの)を参照してください。

以前のバージョンと変換可能なシンタックス ファイルを保存するには次の項目を実行します。

▶ シンタックス ウィンドウのメニューから次の項目を選択します。

ファイル
名前を付けて保存

- [名前を付けて保存] ダイアログ ボックスのエンコード ドロップ ダウン リストで、**[ロケールのエンコード]** を選択します。ロケールのエンコードは現在のロケールにより決定されます。

複数の Execute コマンド

ダイアログ ボックスから貼り付けたり、ログまたはジャーナルからコピーしたシンタックスには、EXECUTE コマンドが含まれる場合があります。シンタックス ウィンドウから複数のコマンドを実行するとき、通常は EXECUTE コマンドを繰り返す必要はありません。EXECUTE コマンドは 1 回ごとにデータ ファイル全体を読み込むため、データ ファイルが大きい場合は特に、パフォーマンスが低下する場合があります。詳細は、『Command Syntax Reference』(SPSS ウィンドウの [ヘルプ] メニューから利用可能) にある EXECUTE コマンドの項を参照してください。

ラグ関数

注意しておくべき 1 つの例外として、ラグ関数を含む変換コマンドがあります。一連の変換コマンドにおいて、データを読み込む EXECUTE コマンドまたはその他のコマンドが含まれない場合、ラグ関数は、コマンドの順序に関係なく、他のすべての変換が終わった後で計算されます。例をあげると、次のようになります。

```
COMPUTE lagvar=LAG(var1).
COMPUTE var1=var1*2.
```

かつ

```
COMPUTE lagvar=LAG(var1).
EXECUTE.
COMPUTE var1=var1*2.
```

上記のコマンドは、lagvar の値に関して異なる結果を生成します。なぜなら、前者のコマンドでは var1 の変換後の値が使用されるのに対し、後者のコマンドでは元の値が使用されるからです。

章
13

度数（プロビット分析）

度数分布表手続きは、多くのタイプの変数を記述するのに有効な統計と図形を表示します。度数分布表手続きは、データをざっと見るには最適なスタート位置です。

度数レポートと棒グラフでは、昇順または降順で値を配置することもでき、さらにカテゴリを度数別に順序付けて表示することもできます。度数分布表は、変数に多くのカテゴリがあるとき抑制することができます。図表には度数分析（デフォルト）またはパーセントでラベル表示することができます。

例。ある企業の顧客は業種によりどのように分布しているでしょうか?出力から、次のことがわかります。顧客の 37.5% は公の省庁関係者で、24.9% は一般企業、28.1% は学術機関、9.4% は医療機関の関係者です。販売収益のような連続した量的データの場合は、平均製品売上高は 3,576 ドルで、標準偏差が 1,078 ドルということがわかります。

統計量と作図。度数、パーセント、累積パーセント、平均値、中央値、最頻値、合計、標準偏差、分散、範囲、最小値と最大値、平均値の標準誤差、歪度と尖度（両方とも標準誤差付き）、4 分位、ユーザー指定のパーセンタイル、棒グラフ、円グラフ、ヒストグラム。

データ。数値コードまたは文字列を使用してカテゴリ変数をコード化します（名義または順序尺度）。

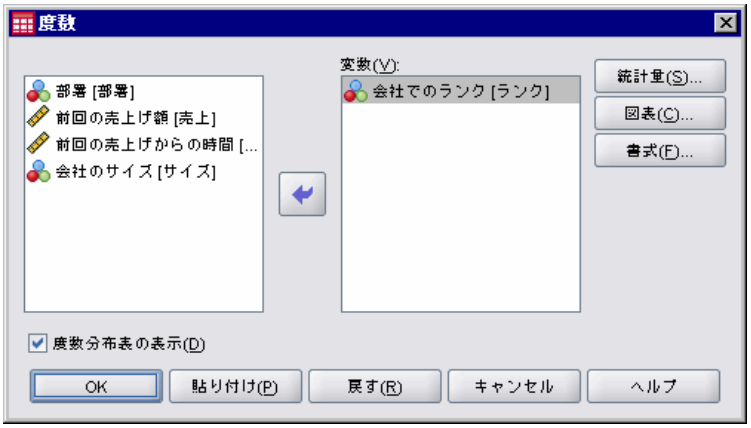
仮定。集計表とパーセントは、どの分布から取り出すデータにも有効ですが、特に順序付けしたカテゴリまたは順序付けしていないカテゴリには、役に立つ統計です。平均値と標準偏差のようなオプションの要約統計のほとんどは、正規理論に基づいていて、分布が対称な量的変数に適しています。中央値、4 分位、およびパーセンタイルのような頑健な統計は正規性の仮定に適合する量的変数にも、適合しない量的変数にも適しています。

度数分布表を取得するには

▶ メニューから次の項目を選択します。

分析
記述統計
度数分布表...

図 13-1
[度数分布表] メーン ダイアログ ボックス



▶ 1 つ以上のカテゴリ変数または量的変数を選択します。

オプションとして、次の選択が可能です。

- 量的変数の記述統計を求めるには [**統計**] をクリックします。
- 棒グラフ、円グラフ、およびヒストグラムを作成するには [**図表**] をクリックします。
- 結果が表示される順序を指定するには [**書式**] をクリックします。

度数分布表の統計

図 13-2
[度数分布表: 統計] ダイアログ ボックス

パーセンタイル値。順序付けしたデータをグループに分割する量的変数の値で、指定されたパーセンテージのケースがその値を上回り、残りのパーセンテージのケースがその値を下回るような値。4 分位 (25、50、75 のパーセンタイル) は、観測をサイズの同じ 4 つのグループに分割します。必要な等サイズ グループの個数が 4 以外の場合は、**[等サイズの n グループに分割]** を選択します。個別のパーセンタイル (たとえば、95 パーセンタイル、観測の 95% が入る値) を指定することもできます。

中心傾向。分布の位置を記述する統計としては、平均値、中央値、最頻値およびすべての値の合計があります。

- **平均。**中心傾向の測定値。観測値の合計をケース数で割った算術平均。
- **中央値。**ケースの中央付近にある値です。50 パーセンタイルです。ケース数が偶数の場合、中央値は、昇順または降順に保存されたときの 2 つのまん中のケースの平均になります。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。
- **最頻値。**最も多く出現する値。複数の値が最高の頻度で出現する場合は、それぞれが最頻値となります。度数分析手続きは、それらのうちの最小の値だけを最頻値として報告します。
- **合計。**欠損値のないすべてのケースに対する変数の値の合計または全体。

散らばり。データの変動量または広がり量を測定する統計としては、標準偏差、分散、範囲、最小値、最大値、平均の標準誤差があります。

- **標準偏差。**平均の周辺のばらつき度。正規分布ではデータの 68% が平均 - SD と平均 + SD のなかに含まれ、データの 95% が平均 -2 SD と平均 +2 SD のなかに含まれます。たとえば、平均が 45 で、標準偏差が 10 である場合、正規分布ではデータの 95% が 25 と 65 の間に含まれます。
- **分散 (信頼性分析)。**平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。
- **OK (ファイルオープン時のオプション)。**数値型変数の最大値と最小値の差。
- **最小。**数値型変数がとる最も小さい値。
- **最大。**数値型変数の最大値。
- **平均値の標準誤差。**同一の分布から取り出したサンプル間で平均値がどの程度ばらついているかを測ったもの。観測された平均と仮説された値を比較するために使うことができます (すなわち、差と標準誤差の比率が -2 より小さいか +2 より大きい場合に、2 つの値は異なっていると結論付けることができます)。

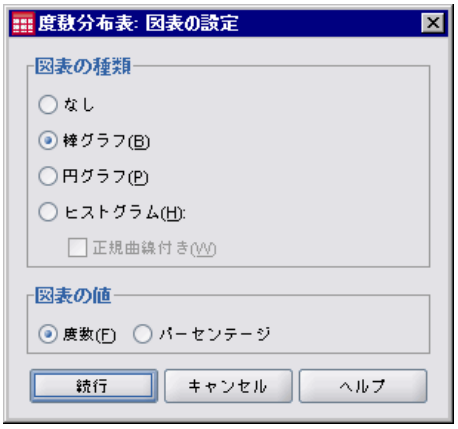
分布。尖度と歪度は、分布の形状や対称を示す統計量です。この統計量は標準誤差とともに表示されます。

- **歪度。**分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。
- **尖度。**観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、観測値の群がりが正規分布のものより大きく、長い裾を持つことを意味します。負の尖度は、観測値の群がりが正規分布のものより小さく、短い裾を持つことを意味します。

値はグループの中間点。データ内の値がグループの中間点にある場合 (たとえば、30 代の人すべての年齢が 35 としてコード化されている場合) グループ化される前の元データの中央値とパーセンタイルを推定するにはこのオプションを選択します。

度数分布表の図表

図 13-3
[度数分布表: 図表の設定] ダイアログ ボックス

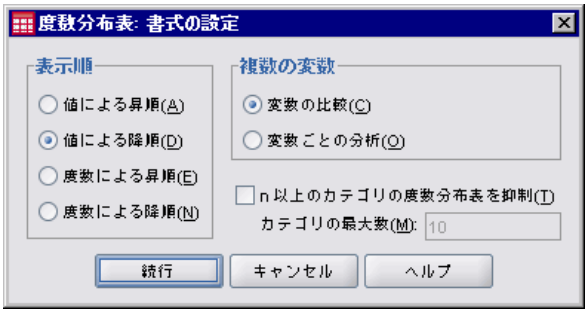


グラフの種類。円グラフは、全体に対する部分の割合を表示します。円グラフの各分割は、1 つのグループ化変数で定義されたグループに対応します。棒グラフは異なる値またはカテゴリの度数を別個の棒として表示するので、カテゴリを視覚的に比較することができます。ヒストグラムにも棒がありますが、こちらは等間隔のスケールに沿ってプロットされます。それぞれの棒の高さは区間内に入る量的変数の値の度数です。ヒストグラムは、分布の行列の形、中央、および広がりを示します。ヒストグラム上に重ね合わせた正規曲線を使用すると、データが正規に分布されているかどうかを判断することができます。

図表の値。棒グラフでは、スケール軸のラベルに度数またはパーセントを使用できます。

度数分布表の書式

図 13-4
[度数分布表: 書式の設定] ダイアログ ボックス



表示順。度数分析は、データ内の実際の値、または値の度数 (発生の度数) に従って、昇順または降順のいずれでも配置することができます。しかし、ヒストグラムまたはパーセンタイルを要求すると、SPSS は、変数が量的であるとみなしその値を昇順で表示します。

複数の変数。複数の変数の統計テーブルを作成する場合、1 つのテーブルにすべての変数を表示すること ([**変数の比較**]) も、変数ごとに統計テーブルを分けて表示すること ([**変数ごとの分析**]) もできます。

n 以上のカテゴリの度数分布表を抑制。このオプションは、指定した数以上の値がある度数分布表を表示しないようにします。

章
14

記述統計記述統計

記述統計手続きにより、複数の変数の 1 変量の要約統計量が 1 つの表に表示され、標準化された値 (*z* 得点) が計算されます。変数は、その平均値の大きさか (平均値の昇順または降順)、アルファベット順、または変数リスト順 (デフォルト) に配列することができます。

保存した *z* 得点はデータ エディタのデータに追加され、図表、データの一覧表表示または分析に使用することができます。変数が別々の単位で記録されている場合 (1 人当たりの国内総生産や識字率など)、*z* 得点変換を行えば、変数の尺度が統一され、変数を一目で比較できるようになります。

例。データの各ケースに各店員の 1 日ごとの売上の合計が、毎日の売上を数か月間分集計した形で含まれている場合 (たとえば、井上、加藤、山田にそれぞれ 1 ずつ入力)、記述統計手続きにより、各店員の一日の平均売上が計算され、平均値が最も大きい店員を先頭に最も小さい店員まで表示されます。

統計量。サンプル サイズ、平均値、最小値、最大値、標準偏差、分散、範囲、合計、平均値の標準誤差、および尖度と歪度とそれらの標準誤差。

データ。誤差、外れ値、および分布の異常を見つけるためにデータをグラフ表示して、修正した数値型変数を使用します。記述統計手続きは、大量のデータ ファイル (千や万単位のケース) を扱う場合に非常に便利です。

仮定。利用可能な統計量 (*z* 得点など) はそのほとんどが通常の理論に基づいており、対称型の分布を持つ数量変数 (間隔または比率尺度) に適しています。順序付けされていないカテゴリまたは非対称分布変数は避けます。*z* 得点の分布は元データと同じ形をしているため、*z* 得点の計算を行っても、データの問題点が解消されるわけではありません。

記述統計を行うには

▶ メニューから次の項目を選択します。

分析
記述統計
記述統計...

図 14-1
[記述統計] ダイアログ ボックス

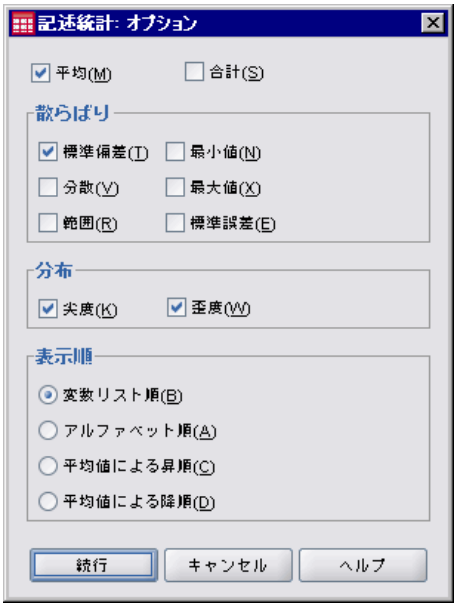
▶ 1 つ以上の変数を選択します。

オプションとして、次の選択が可能です。

- *z* 得点を新しい変数として保存するには、[**標準化された値を変数として保存**] チェック ボックスをオンにします。
- その他の統計量や表示順を選択する場合は、[**オプション**] をクリックします。

記述統計のオプション

図 14-2
[記述統計: オプション] ダイアログ ボックス



[平均値] と [合計]。デフォルトでは、平均値、つまり算術平均が表示されます。

散らばり。データの広がりまたは偏差を測定する統計には、標準偏差、分散、範囲、最小値、最大値、および平均値の標準誤差があります。

- **標準偏差。**平均の周辺のばらつき度。正規分布ではデータの 68% が平均 - SD と平均 + SD のなかに含まれ、データの 95% が平均 -2 SD と平均 +2 SD のなかに含まれます。たとえば、平均が 45 で、標準偏差が 10 である場合、正規分布ではデータの 95% が 25 と 65 の間に含まれます。
- **分散 (信頼性分析)。**平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。
- **OK (ファイルオープン時のオプション)。**数値型変数の最大値と最小値の差。
- **最小。**数値型変数がとる最も小さい値。
- **最大。**数値型変数の最大値。
- **平均値の S.E。**同一の分布から取り出したサンプル間で平均値がどの程度ばらついているかを測ったもの。観測された平均と仮説された値を比較するために使うことができます (すなわち、差と標準誤差の比率が -2 より小さいか +2 より大きい場合に、2 つの値は異なっていると結論付けることができます)。

分布。尖度と歪度は、分布の形状や対称を表す統計量です。この統計量は標準誤差とともに表示されます。

- **尖度。**観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、観測値の群がりが正規分布のものより大きく、長い裾を持つことを意味します。負の尖度は、観測値の群がりが正規分布のものより小さく、短い裾を持つことを意味します。
- **歪度。**分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。

表示順。デフォルトでは、変数は選択した変数リストの順に表示されます。[オプション] 機能で表示順を、アルファベット順、平均値による昇順または降順から選択することができます。

DESCRIPTIVES コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 一部の変数に対して (すべての変数ではありません)、標準化された得点 (*z* 得点) を保存 (VARIABLES サブコマンドを使用)。
- 標準化された得点を含める新しい変数の名前を指定 (VARIABLES サブコマンドを使用)。

- 変数の欠損値のあるケースを分析から除外（MISSING サブコマンドを使用）。
- 平均値だけでなく統計値も加えた順番に変数の表示を並べ替える（SORT サブコマンドを使用）。

シンタックスの詳細は、『*Command Syntax Reference*』を参照してください。

章
15

探索的分析

探索的分析手続きは、ケースのすべて、またはケースのグループごとについて、要約統計量と図形表示を作成します。探索的分析手続きを使用するには多くの理由があり、データ スクリーニング、外れ値の識別、記述統計、仮説の検定、および下位母集団 (ケースのグループ) 間での相違点の特徴付けといったことができます。データ スクリーニングでは、データが異常値、極値、データ内のギャップ、またはその他の特性を持っているかどうかがわかります。データに対して探索的分析を行うと、データ分析に対して考えている統計手法が適切なものであるかどうかを判断することできます。探索的分析によっては、正規分布を前提とする手法を行う際にデータ変換が必要であることが示されることもあります。または、ノンパラメトリック検定が必要であると判断することもあります。

例。4 種類の強化計画に基づいてネズミ用の迷路学習時間の分布を見てみましょう。4 つのグループそれぞれに対して、時間の分布が近似的に正規分布しているかどうか、かつ 4 つの分散が等しいかどうかを確認することができます。また、学習時間の最大の 5 つのケースおよび最小の 5 つのケースを識別することもできます。箱ひげ図と幹葉図は、グループそれぞれの学習時間の分散を図で示して要約します。

統計量と作図。平均値、5% トリム平均値、標準誤差、分散、標準偏差、最小値、最大値、範囲、4 分位範囲、歪度と尖度およびその標準誤差、平均値の信頼区間 (および指定した信頼係数)、パーセンタイル、Huber の M 推定量、Andrews のウェイブ推定量、Hampel の M 推定量、Tukey のバイウェイト推定量、5 つの最大値と 5 つの最小値、正規性を検定するための Lilliefors の有意確率、Kolmogorov-Smirnov の統計量、および Shapiro-Wilk の統計量。箱ひげ図、幹葉図、ヒストグラム、正規性プロット、および Levene 検定と変換による水準と広がりの図。

データ。探索的分析手続きは、量的変数 (区間または比尺度の測定) に使用することができます。因子変数 (データをケースのグループに分解するときに使用する) には、妥当な異なった値 (カテゴリ) がなければなりません。これらの値は、短い文字型または数値にすることができます。箱ひげ図の外れ値を示すために使用するケースのラベルの変数は、短い文字型、長い文字型 (最初の 15 バイト)、または数値にすることができます。

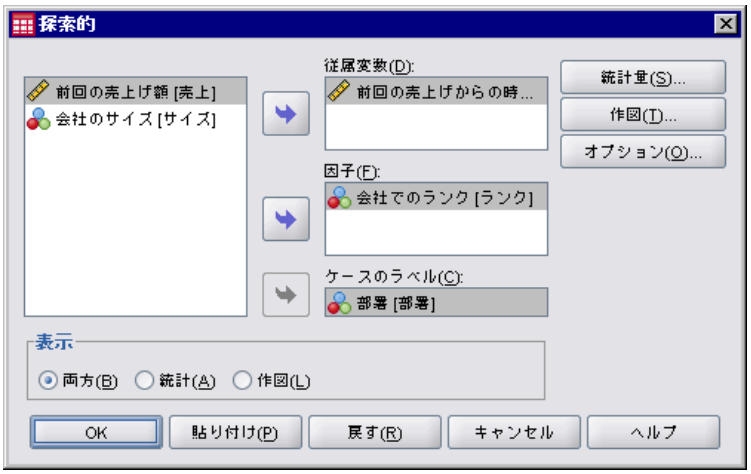
仮定。データの分布は、対称または正規である必要はありません。

データの探索的分析を行うには

▶ メニューから次の項目を選択します。

分析
記述統計
探索的...

図 15-1
[探索的分析] ダイアログ ボックス



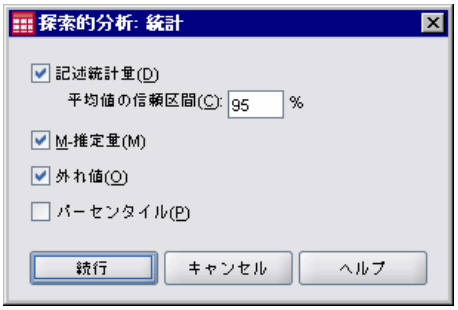
▶ 従属変数を 1 つ以上選択します。

オプションとして、次の選択が可能です。

- 値がケースのグループを定義する 1 つ以上の因子変数を選択できます。
- ケースにラベルを付けるための識別変数を選択できます。
- [統計] をクリックすると、M-推定量、外れ値、パーセンタイル、および記述統計量を使用できます。
- [作図] をクリックすると、ヒストグラム、正規性の検定とプロット、Levene の統計による水準と広がりの図を使用できます。
- [オプション] をクリックすると、欠損値の処理を行えます。

探索的分析の統計

図 15-2
[探索的分析: 統計] ダイアログ ボックス



記述統計。中心傾向と散らばりの測度は、デフォルトで表示されます。中心傾向の測度は分布の位置を表していて、それには平均値、中央値、および 5% トリム平均値が含まれます。散らばりの測度は値の非類似性を表し、標準誤差、分散、標準偏差、最小値、最大値、範囲、および 4 分位範囲が含まれます。記述統計には、分布の形状の測度も含まれ、歪度、および尖度はその標準誤差とともに表示されます。平均値の 95% 水準の信頼区間も表示されますし、任意の信頼水準を指定することもできます。

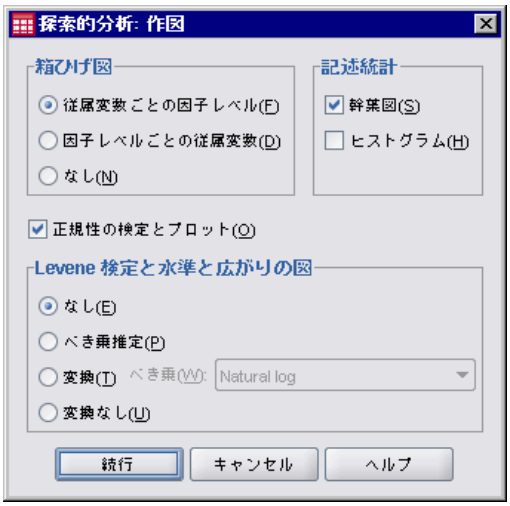
M-推定量。位置を推定するためのサンプル平均値と中央値の頑健な推定量。それぞれの推定量は、ケースに適用する重みで違いがあります。Huber の M-推定量、Andrews のウェイブ推定量、Hampel の M-推定量、および Tukey のバイウェイト推定量が表示されます。

外れ値。ケース ラベルにより 5 つの最大値と 5 つの最小値を表示します。

パーセンタイル。5、10、25、50、75、90、95 番目のパーセンタイルに値を表示します。

探索的分析の作図

図 15-3
[探索的分析: 作図] ダイアログ ボックス



箱ひげ図。複数の従属変数があるときには箱ひげ図の表示を制御します。[**従属変数ごとの因子レベル**] では、従属変数ごとに個別の図表が生成されます。1 つの図表内で、因子変数によって定義されたグループのそれぞれに箱ひげ図が作成されます。[**因子レベルごとの従属変数**] では、因子変数によって定義されたグループごとに個別の図表が生成されます。1 つの図表内で、各従属変数の箱ひげ図を並べて表示します。異なる時に測定した異なる変数が特定の特性を表す場合に、特に便利です。

記述統計量。[記述統計] グループを使用すると、幹葉図とヒストグラムを選択できます。

正規性の検定とプロット。正規確率と傾向化除去正規確率プロットを表示します。正規性を検定するための Lilliefors の有意確率と Kolmogorov-Smirnov の統計量も表示されます。整数でない重み付けが指定されている場合、重み付けサンプルサイズが 3 ～ 50 のときに Shapiro-Wilk 統計量が計算されます。重み付けが指定されていないか、または整数が指定されている場合は、重み付けサンプル サイズが 3 ～ 5,000 のときに Shapiro-Wilk 統計量が計算されます。

Levene 検定と水準と広がりの図。水準と広がりの図のデータ変換を制御します。水準と広がりの図のすべてに、回帰直線および等分散性の Levene の頑健な検定が表示されます。変換を選択すると、Levene の検定が変換データに基づいて実行されます。因子変数を選択しないと、水準と広がりの図は作成されません。[**べき乗推定**] では、セル内の分散が等しくなるようにべき乗変換の推定を行うだけではなく、すべてのセルの中央値の自然対数と 4 分位範囲の自然対数のプロットを作成します。水準と広がりの図を使うと、グループ全体で分散を安定させるため (より等しくするため) の変換のべき乗を決定することができます。[**変**

換] では、べき乗推定からの推奨に従って、べき乗の選択肢の 1 つを選択し、さらに変換データのプロットを作成することができます。4 分位範囲および変換データの中央値がプロットされます。[**変換なし**] では、生データのプロットが作成されます。これは、1 乗による変換と等しくなります。

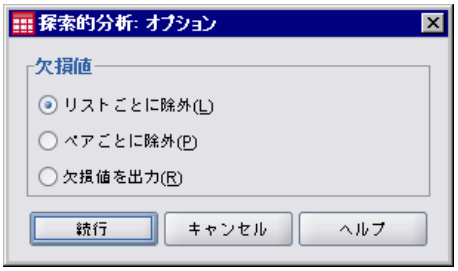
探索的分析のべき乗変換

これは、水準と広がりの図用のべき乗変換です。データを変換するには、変換用のべき乗を選択しなければなりません。次のオプションのどちらかを選択できます。

- **自然対数。**自然対数変換。これはデフォルトです。
- **平方根の逆数。**各データ値に対して、平方根の逆数が計算されます。
- **逆数。**各データ値の逆数が計算されます。
- **平方根。**各データ値の平方根が計算されます。
- **平方。**各データ値が 2 乗されます。
- **立方。**各データ値が 3 乗されます。

探索的分析のオプション

図 15-4
[探索的分析: オプション] ダイアログ ボックス



欠損値。欠損値の処理を管理します。

- **リストごとに除外。**従属または因子変数に対し欠損値のあるケースは、すべての分析から除外されます。これはデフォルトです。
- **ペアごとに除外。**グループ (セル) 内の変数に欠損値がないケースは、そのグループの分析に含まれます。ケースは、他のグループで使用する変数に欠損値がある場合も含みます。
- **欠損値を出力。**因子変数の欠損値は、別のカテゴリとして処理されます。すべての出力は、この追加カテゴリに作成されます。度数分布表には、欠損値のカテゴリが含まれます。因子変数の欠損値は分布には含まれますが、欠損として表示されます。

EXAMINE コマンドの追加機能

探索的分析手続きでは、EXAMINE コマンド シンタックスを使用します。コマンド シンタックスを使用すると、次の作業も実行できます。

- 因子変数で定義したグループの出力とプロットに加えて、合計の出力とプロットを要求 (TOTAL サブコマンドを使用)。
- 箱ひげ図のグループ用に共通尺度を指定 (SCALE サブコマンドを使用)。
- 因子変数の交互作用を指定 (VARIABLES サブコマンドを使用)。
- デフォルト以外のパーセンタイルを指定 (PERCENTILES サブコマンドを使用)。
- 5 種類のいずれかの方法に従ってパーセンタイルを計算 (PERCENTILES サブコマンドを使用)。
- 水準と広がりの図に使用する任意のべき乗変換を指定 (PLOT サブコマンドを使用)。
- 表示する極値の数を指定 (STATISTICS サブコマンドを使用)。
- 位置の M-推定量および頑健推定量のパラメータを指定 (MESTIMATORS サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
16
クロス集計表

クロス集計表手続きでは、2 元表および多次元表を作成し、さらに 2 元表の連関のさまざまな検定および測定を行うことができます。表の構造およびカテゴリが順序付けされているかどうかにより、使用される検定および測定が決定します。

クロス集計の統計および連関の測定が計算される対象は、2 次元表に限られます。行、列、および層（制御変数）を指定すると、[クロス集計表] 手続きは、層の各値（または 2 個以上の制御変数の値の組み合わせ）に対して 1 パネルの連関の統計量および測定方法を作成します。たとえば、**性別**が **人生観**（人生は楽しいか、日常的か、またはつまらないか）と **結婚経験**（はい、いいえ）のクロス表の層である場合、女性についての 2 次元表の結果は男性についてのものとは別々に計算され、交互に並んだパネルとして表示されます。

例。教育やコンサルティングなどのサービス業務において、中小企業の顧客は、大企業より収益を生む可能性が高いでしょうか。クロス集計より、この分野では、大企業（従業員 2,500 人以上）の収益が低く、中小企業（従業員 500 人未満）の収益が高いことを示しています。

統計量と連関の測定方法。Pearson のカイ 2 乗、尤度比カイ 2 乗、線型と線型による連関検定、Fisher の直接法、Yates の修正カイ 2 乗、Pearson の *r*、Spearman のロー、分割係数、ファイ、Cramer の *V*、対称および非対称ラムダ、Goodman と Kruskal のタウ、不確定性係数、ガンマ、Somers の *d*、Kendall の タウ *b*、Kendall のタウ *c*、イータ係数、Cohen のカッパ、相対リスク推定値、オッズ比、McNemar 検定、および Cochran 統計量と Mantel-Haenszel 統計量。

データ。各表変数のカテゴリを定義するには、数値型変数または短い文字型変数 (8 バイト以下) の値を使います。たとえば、**性別**の場合、1 と 2 または **男性**と **女性**のようにデータをコード化できます。

仮定。統計値および測定方法のなかには、順序付けされたカテゴリ（順位データ）または量的な値（区間データまたは比率データ）であることを前提とするものがあります（詳細は統計値についての章を参照）。また、表変数に順序付けされたカテゴリでないもの（名義データ）があるときに有効なものもあります。カイ 2 乗に基づく統計値（ファイ、Cramer の *V*、分割係数）の場合、データは多項分布から無作為に抽出されたサンプルとなります。

注: 順序変数は、カテゴリを表す数値型コード（たとえば、1 = **低**、2 = **中**、3 = **高**）または文字型値です。ただし、カテゴリの本当の順序を反映させるため、文字型値のアルファベット順を仮定しています。たとえば、値が **低**、**中**、**高**である文字型変数では、カテゴリの順序は **高**、**低**、**中**と解釈されますが、これは正しい順

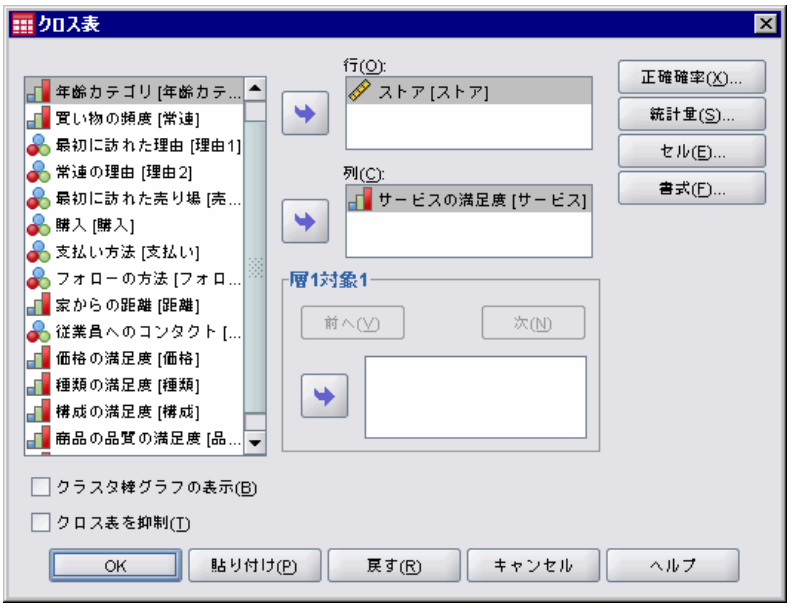
序ではありません。一般に、順序データを表す場合には、数値型コードを使用した方が信頼性が高いといえます。

クロス集計を行うには

▶ メニューから次の項目を選択します。

分析
  記述統計
    クロス集計表...

図 16-1
[クロス集計表] ダイアログ ボックス



▶ 1 つ以上の行変数および列変数を選択します。

オプションとして、次の選択が可能です。

- 1 つ以上の制御変数を選択する。
- [統計] をクリックして、2 次元表または副表の検定の統計値および連関の測定方法を選択する。
- [セル] をクリックして、[観測] および [期待] 値、[パーセンテージ]、[残差] を選択する。
- [書式] をクリックして、カテゴリの表示順を制御する。

クロス集計表の層

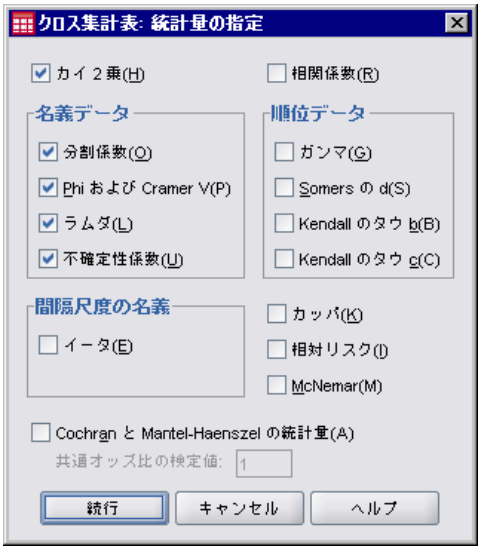
1 つ以上の層変数を選択すると、各層変数 (制御変数) のカテゴリごとに別々のクロス集計が作成されます。たとえば、行変数が 1 つ、列変数が 1 つで、2 つのカテゴリを持つ層変数が 1 つある場合、この層変数の各カテゴリに対してそれぞれ2 次元表が 1 つ得られます。別の制御変数の層を作成するには、[次] をクリックします。第 1 層変数のカテゴリと第 2 層変数との組み合わせに対してそれぞれ表が作成されます。統計量および連関の測定が要求されている場合は、2 次元表にだけ適用されます。

クロス集計表のクラスタ棒グラフ

クラスタ棒グラフの表示。クラスタ棒グラフを使用して、データをケースのグループに要約できます。[行] の一覧で指定した変数の各値を表す棒のクラスタが 1 つあります。各クラスタ内で棒を定義する変数は、[列] の一覧で指定した変数です。この変数の各値に対してそれぞれ違う色が付けられたり、または違うパターンの棒のグループが 1 つあります。列または行で 1 つ以上の変数を指定した場合は、行と列の変数のそれぞれの組み合わせに対してクラスタ棒グラフが作成されます。

クロス集計表の統計

図 16-2
[クロス集計表: 統計量の指定] ダイアログ ボックス



カイ 2 乗。2 つの行と 2 つの列を持つ表の場合は、[カイ 2 乗] チェック ボックスをオンにして、Pearson のカイ 2 乗、尤度比カイ 2 乗、Fisher 直接法、および Yates の修正カイ 2 乗 (連続性のための修正) を計算します。2 × 2 表の場合、大規模な表の欠損行または欠損列を持たない表で、セルの期待度数が 5 より少な

いときには、Fisher の直接法が計算されます。他のすべての 2 × 2 表の場合には、Yates の修正カイ 2 乗が計算されます。任意の数の行と列で構成される表の場合は、**[カイ 2 乗]** チェック ボックスをオンにして、Pearson のカイ 2 乗および尤度比のカイ 2 乗を計算します。表変数がどちらも量的変数である場合は、**カイ 2 乗**は線型と線型による連関検定になります。

相関係数。行および列の両方に順位の値が含まれている表の場合は、**[相関係数]** により Spearman の相関係数、ロー (数値データのみ) が得られます。Spearman のローは、ランク順間の関連度です。表変数 (因子) がどちらも量的変数である場合は、**[相関]** により Pearson の相関係数、*r*、変数間の線型による連関の測定が得られます。

名義。名義データ (カトリック、プロテスタント、ユダヤなどの非順序尺度) の場合には、**[ファイ]** (係数) と **[Cramer の V]**、**[分割係数]**、**[ラムダ]** (対称ならびに非対称ラムダおよび Goodman と Kruskal のタウ)、**[不確定性係数]** を選択できます。

- **分割係数。**カイ 2 乗に基づく連関度。値は 0 から 1 の範囲になります。値 0 は行変数と列変数の間に関連がないことを示し、1 に近い値は変数間に強い関連があることを示します。可能な最大値は、表の行と列の数によって決まります。
- **標準偏差 (グラフの集計関数)。**ファイは、カイ 2 乗に基づく関連度で、カイ 2 乗統計量をサンプル サイズで割り、その結果の平方根を取ります。Cramer の V は、カイ 2 乗に基づく関連度です。
- **ラムダ。**独立変数の値が従属変数の値を予測しようとするときの、誤差の減少を反映した予測連関指数です。値 1 は、独立変数が従属変数を完全に予測することを意味します。値 0 は、独立変数が従属変数の予測に役立たないことを意味します。
- **不確定性係数。**1 つの変数値がその他の変数値の予測に使われるとき、誤差内の予測連関指数を示す連関度。たとえば、値 0.83 は一方の変数がもう一方の変数の値を予測する際に、誤差を 83% 減らすという情報を示します。プログラムは、不確実性係数の対称版と非対称版の両方を計算します。

順序。行および列の両方に順序の値が含まれている場合は、**[ガンマ]** (2 次元表には 0 次で、3 次元～10 次元表には条件付)、**[Kendall のタウ b]**、および **[Kendall のタウ c]** を選択します。行カテゴリから列カテゴリを予測する場合は、**[Somers の d]** を選択します。

- **ガンマ (正規確率プロット 検定分布)。**2 つの順序変数間の対称な連関度で、−1 から 1 の範囲を取ります。絶対値 1 に近い値は、2 つの変数の間に強い関係があることを示します。値が 0 に近い場合は、関係が弱いかまったくないことを示します。2 次元表では、0 次ガンマが表示されます。3 次元表から n 次元表では、条件付きのガンマが表示されます。
- **Somers の d。**2 つの順序変数の間の関連度で、−1 から 1 の値をとります。絶対値が 1 に近い場合は 2 つの変数間に強い関係があることを示し、0 に近い場合は変数間の関係が弱いかまったくないことを示します。Somers の d は、

独立変数の同順位でないペアの数の差を取ることによって、ガンマ係数を非対称に拡張したものです。この統計量の対称版も計算されます。

- **Kendall のタウ b。**同順位を考慮する順序変数の、ノンパラメトリックな相関度。係数の符号はその関係の方向を示し、その絶対値は、より強い関係を示すより大きな絶対値によって強度を示します。-1 から 1 までの値を取りますが、-1 または +1 が得られるのは平方表からだけです。
- **Kendall のタウ c。**同順位を無視する順序変数の、ノンパラメトリックな関連度。係数の符号はその関係の方向を示し、その絶対値は、より強い関係を示すより大きな絶対値によって強度を示します。-1 から 1 までの値を取りますが、-1 または +1 が得られるのは平方表からだけです。

間隔尺度の名義。一方の変数がカテゴリ変数で、他の一方が量的変数であるとき、[**イータ**] を選択します。カテゴリ変数は数値でコード化されている必要があります。

- **イータ (グループの平均)。**0 から 1 までの範囲の連関度です。0 は行変数および列変数の間に連関がないことを示し、1 に近い値は連関の度合いが高いことを示します。イータは、間隔尺度で測定された従属変数 (収入など) とカテゴリの少ない独立変数 (性別など) の分析に適しています。2 つのイータ値が計算されます。1 つは、行変数を間隔変数として扱うもので、もう 1 つは、列変数を間隔変数として扱うものです。

カッパ (因子分析)。Cohen のカッパは、2 つの評価者が同じ対象を評価するときに、それらの評価の一致の度合を測定します。値 1 は完全な一致を表します。値 0 は、偶然以外の一致がないことを表します。カッパは、両方の変数が同じカテゴリ値を持ち、かつ両方の変数が同じ数のカテゴリを持つ表でだけ利用できます。

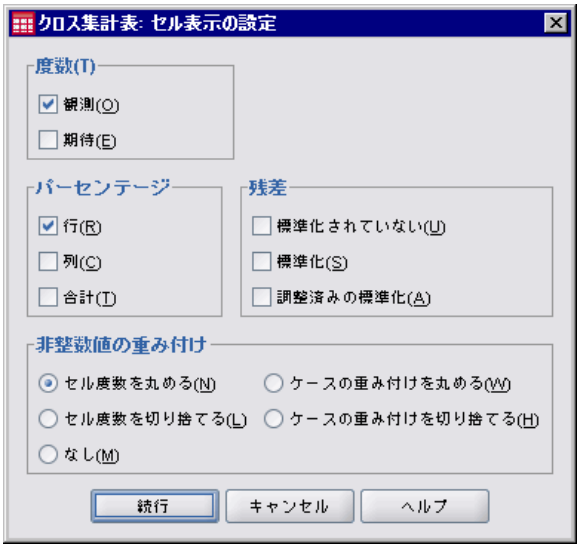
リスク。2 x 2 表の場合は、ある因子の存在とあるイベントの発生の関連の強さの測定値になります。統計量の信頼区間に 1 が含まれる場合は、その因子がそのイベントに関連していると仮定することはできません。因子の発生がまれなときには、オッズ比を推定値または相対リスクとして使うことができます。

McNemar。関連する 2 つの 2 分変数に対するノンパラメトリック検定。カイ 2 乗分布を使って応答の変化を検定します。設計の前後での実験的介入により、反応の変化の検出に便利です。大きな平方表では、McNemar-Bowker 対称検定が報告されます。

Cochran 統計量と Mantel-Haenszel 統計量。Cochran と Mantel-Haenszel 統計量は、1 つ以上の層 (制御) 変数によって定義された共変量パターンを条件として 2 値因子変数と 2 値応答変数の間の独立性を検定するために使われます。他の統計量は層ごとに計算されますが、Cochran と Mantel-Haenszel 統計量は、すべての層に対して一度に計算されます。

クロス集計表のセル表示の設定

図 16-3
[クロス集計表: セル表示の設定] ダイアログ ボックス



カイ 2 乗検定で有意になるデータのパターンを見つけやすくするために、[クロス集計表] 手続きにより、期待度数および観測度数と期待度数との差を測定する 3 種類の残差 (偏差) が表示されます。クロス表の各セルは、度数、パーセント、および残差から任意に選択して組み合わせることができます。

度数。行変数および列変数が相互に独立している場合は、実際の観測ケース数および期待ケース数。

パーセンテージ。パーセンテージは、行全体または列全体を合計できます。表 (1 つの層) に表示されるケースの合計数のパーセンテージも使えます。

残差。標準化されていない残差により、観測値および期待値の間の差分が示されます。標準化された残差および調整済みの標準化された残差も選択できます。

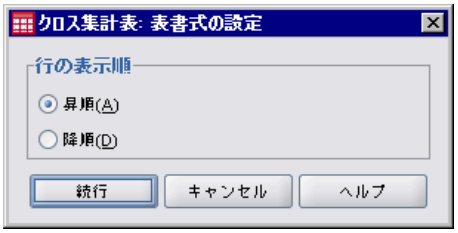
- **標準化されていない (判別分析)。**観測値と期待値の差。期待値は、2 つの変数の間に関係がないと想定した場合に期待されるセルのケース数です。正の残差は、行変数と列変数が独立であると想定される場合に、セルの期待されたケース数より実際のケース数が多いことを示します。
- **標準化残差。**残差を標準偏差の推定値で割った値。標準化残差は Pearson 残差とも呼ばれ、平均は 0 で、標準偏差 1 になります。
- **調整済みの標準化。**セルの残差 (観測度数 − 期待度数) をその標準誤差の推定値で割った値。結果の標準化残差は、標準偏差を単位として平均より上または下で表されます。

非整数値の重み付け。セル度数は、各セル内のケースの数を表すので、通常は整数値になります。ただし、小数値変数を含む重み付け変数 (たとえば 1.25) によって現在データ ファイルが重み付けられている場合、セル度数は小数値になります。セル度数の計算前または計算後に値の切り捨てや丸めを行ったり、小数値のセル度数をテーブル表示と統計計算の両方に使用することができます。

- **丸めセル度数。**ケース重みはそのままで使用されますが、セルの累積重みはすべての統計量が計算される前に丸められます。
- **切り捨てセル度数。**ケース重みはそのままで使用されますが、セルの累積重みはすべての統計量が計算される前に切り捨てられます。
- **丸めケース重み。**ケース重みは、使用前に丸められます。
- **切り捨てケース重み。**ケース重みは、使用前に切り捨てられます。
- **なし。**ケースの重み付けがそのまま使用され、小数値のセル度数が使用されます。ただし、正確検定統計量 ([正確確率検定] オプションが利用できるときのみ) が要求された場合、セルの累積重みは、正確確率検定統計量が計算される前に、丸められるか、切り捨てられます。

クロス集計表の表書式

図 16-4
[クロス集計表: 表書式の設定] ダイアログ ボックス



行は、行変数の値の昇順または降順に整列できます。

章
17

要約

ケースの要約手続きで、1 つ以上のグループ化変数から成るカテゴリ内の変数に対するサブグループ統計量を計算します。グループ化変数のすべてのレベルがクロス集計されます。統計量の表示順を選択できます。カテゴリ全体での各変数の要約統計量も表示されます。各カテゴリのデータ値をリスト表示したり、または表示を抑制できます。大きいデータセットでは、最初の *n* ケースだけを一覧表示できます。

例。地区および顧客業種別の平均製品売上高はどのようなものでしょうか。他地区に比べて西部地区の平均売上高はわずかに高いことがわかります。西部地区の企業顧客からは最高の平均売上高が得られているからです。

統計量。合計、ケースの数、平均値、中央値、グループの中央値、平均値の標準誤差、最小値、最大値、範囲、グループ化変数の最初のカテゴリの変数値、グループ化変数の最後のカテゴリの変数値、標準偏差、分散、尖度、尖度の標準誤差、歪度、歪度の標準誤差、総和のパーセント、総数のパーセント、グループ変数での合計のパーセント、グループ変数でのケース数のパーセント、幾何平均、調和平均。

データ。グループ化変数は、値が数値型または文字型のカテゴリ変数です。カテゴリの数はかなり少ないものでなくてはなりません。他の変数は、ランク付けできるものでなくてはなりません。

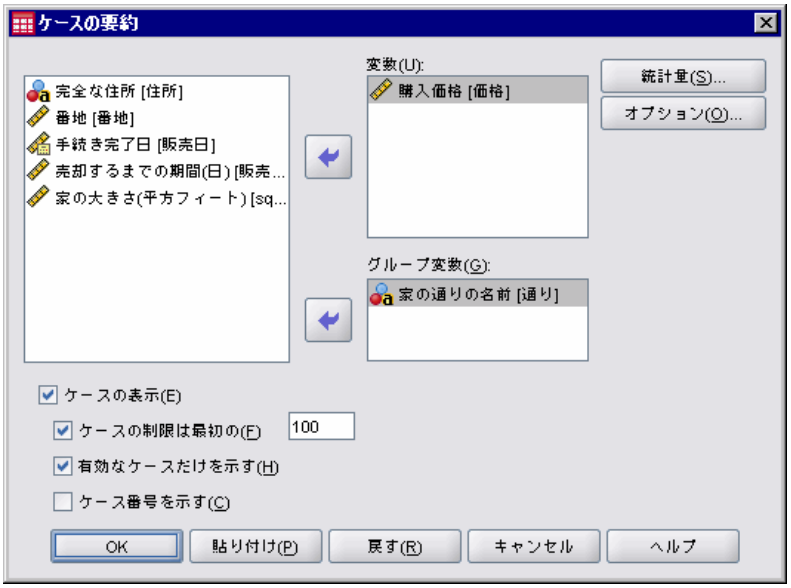
仮定。オプションのサブグループ統計量の中には、平均値や標準偏差などのように、通常の理論に基づいていて、対称的分布を持つ量的変数に適しているものがあります。中央値や範囲などの頑健な統計は、正規性の仮定に合う場合と合わない場合のある量的変数に適しています。

ケースの要約を行うには

▶ メニューから次の項目を選択します。

分析
報告書
ケースの要約...

図 17-1
[ケースの要約] ダイアログ ボックス



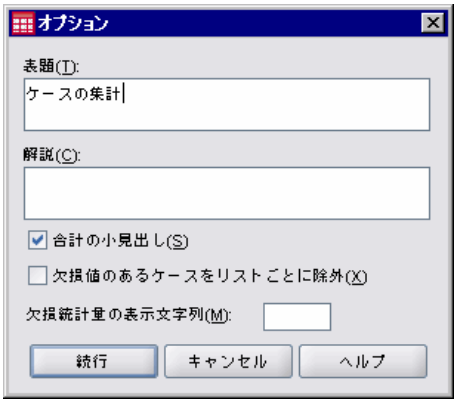
▶ 1 つ以上の変数を選択します。

オプションとして、次の選択が可能です。

- 1 つ以上のグループ化変数を選択して、データをサブグループに分割する。
- [**オプション**] をクリックして、出力表題を変更、解説を追加、または欠損値を持つケースを除外する。
- [**統計**] をクリックして、オプションの統計量にアクセスする。
- [**ケースの表示**] を選択して、各サブグループのケースを一覧で表示する。デフォルトでは、ファイル内の最初の 100 ケースだけが表示されます。[**ケース数を最初の n 個に制限**] の値を増やすか減らしたり、その項目を選択解除してすべてのケースを表示したりできます。

ケースの要約のオプション

図 17-2
[オプション] ダイアログ ボックス

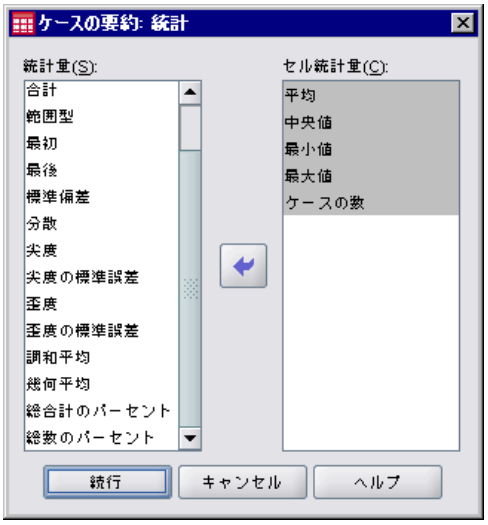


SPSS では、出力の表題を変更したり、出力テーブルの下に表示される解説を追加することができます。また、表題や解説で改行を指定することもできます。テキスト中の任意の位置で ¥n と入力すると、その位置で改行されます。

欠損値のケースは、ピリオドまたはアスタリスクを付けて出力することが望ましい場合がしばしばあります。欠損値の発生時に表示させたい文字、語句、またはコードを入力します。それ以外の場合は、出力時に欠損値のケースに対して特別な処理は行われません。

ケースの要約の統計

図 17-3
[ケースの要約: 統計] ダイアログ ボックス



各グループ化変数のカテゴリ内の変数に対して、サブグループ統計量である合計、ケース数、平均値、中央値、グループ中央値、平均値の標準誤差、最小値、最大値、範囲、グループ化変数の最初のカテゴリの変数値、グループ化変数の最後のカテゴリの変数値、標準偏差、分散、尖度、尖度の標準誤差、歪度、歪度の標準誤差、総和のパーセント、総数のパーセント、グループ化変数での合計のパーセント、グループ化変数でのケース数のパーセント、幾何平均、調和平均のうち 1 つ以上を選択できます。統計量が [セル統計量] リストに現れる順序は、出力に表示される順序です。カテゴリ全体で各変数に対する要約統計量も表示されます。

最初。データ ファイルで発生した最初のデータ値を表示します。

幾何平均。データの値の積の n 乗根です。n はケースの数を表します。

グループ中央値。グループに対してコード化されたデータについて計算された中央値。たとえば、年齢データで、30 代の各値が 35 にコード化され、40 代の各値が 45 にコード化されるとすると、グループ中央値はコード化されたデータから計算された中央値になります。

調和平均。グループ間でサンプルの大きさが等しくないときに、グループの大きさの平均を予測するために使われます。調和平均は、サンプルの合計をサンプルの大きさの逆数の和で割ったものです。

尖度。観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、観測値の群がりが正規分布のものより大きく、長い裾を持つことを意味します。負の尖度は、観測値の群がりが正規分布のものより小さく、短い裾を持つことを意味します。

最後。データ ファイルで発生した最後のデータ値を表示します。

最大。数値型変数の最大値。

平均。中心傾向の測定値。観測値の合計をケース数で割った算術平均。

中央値。ケースの中央付近にある値です。50 パーセンタイルです。ケース数が偶数の場合、中央値は、昇順または降順に保存されたときの 2 つのまん中のケースの平均になります。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。

最小。数値型変数がとる最も小さい値。

N。ケース (観測値やレコード) の数。

総数のパーセント。各カテゴリのケースの総数のパーセントです。

総合計のパーセント。各カテゴリの総和のパーセントです。

OK (ファイルオープン時のオプション)。数値型変数の最大値と最小値の差。

歪度。分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。

尖度の標準誤差。標準誤差に対する尖度の比率は、正規性の検定として使うことができます (すなわち、比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。尖度として大きな正の値は、分布の裾が正規分布の裾より長いことを示し、尖度として負の値は短い裾を示します (箱型の一様分布の裾のようになります)。

歪度の標準誤差。標準誤差に対する歪度の比率は、正規性の検定として使うことができます (すなわち、比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。歪度として大きな正の値は、右の裾が長いことを示し、極端な負の値は左の裾が長いことを示します。

合計。欠損値のないすべてのケースに対する変数の値の合計または全体。

分散 (信頼性分析)。平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。

章
18

平均値（信頼性分析）

[平均の比較] 手続きでは、1 つ以上の独立変数のカテゴリ内の従属変数を対象に、サブグループの平均と関連した 1 変量の統計量を計算します。オプションとして、一元配置分散分析、イータ、線型性の検定などを得ることができます。

例。異なる 3 種類の食用油について、その平均脂肪吸収量を測定し、一元配置分散分析によって平均値が異なるかどうかを確かめます。

統計量。合計、ケースの数、平均値、中央値、グループの中央値、平均値の標準誤差、最小値、最大値、範囲、グループ化変数の最初のカテゴリの変数値、グループ化変数の最後のカテゴリの変数値、標準偏差、分散、尖度、尖度の標準誤差、歪度、歪度の標準誤差、総和のパーセント、総数のパーセント、グループ変数での合計のパーセント、グループ変数でのケース数のパーセント、幾何平均、調和平均。オプションには、分散分析、イータ、イータ 2 乗、線型性 R および R^2 の検定があります。

データ。従属変数は量的で、独立変数はカテゴリです。カテゴリ変数の値は、数値または文字型です。

仮定。オプションのサブグループ統計量の中には、平均値や標準偏差などのように、通常の理論に基づいていて、対称的分布を持つ量的変数に適しているものがあります。頑健な統計量（中央値など）は、正規性の仮定に適合する量的変数にも、適合しない量的変数にも適しています。分散分析は正規性からの逸脱には頑健ですが、各セルのデータは対称でなければなりません。また各グループは、分散が等しい母集団からのものであると仮定します。この仮定を検定するには、[一元配置分散分析] 手続きで利用できる Levene の等分散性の検定を使用します。

サブグループの平均を求めるには

▶ メニューから次の項目を選択します。

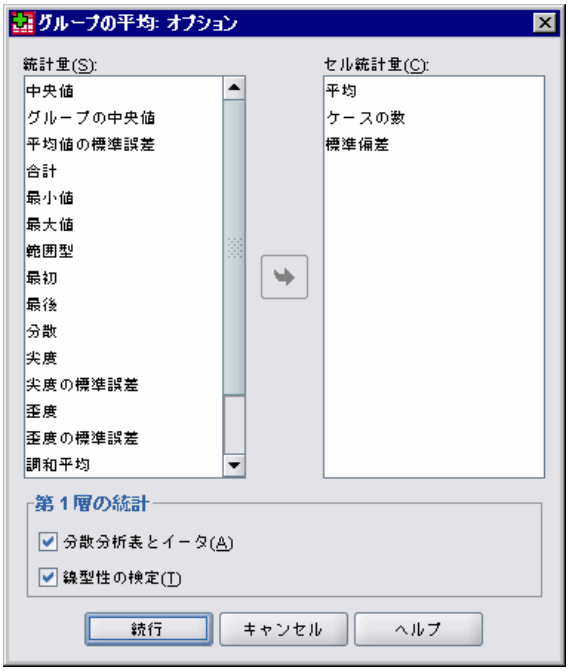
分析
平均の比較
グループの平均...

図 18-1
[平均値] ダイアログ ボックス

- 従属変数を 1 つ以上選択します。
- 以下の方法の 1 つを使って、カテゴリ独立変数を選択します。
  - 1 つ以上の独立変数を選択します。独立変数ごとに、結果が個別に表示されます。
  - 独立変数の層を 1 つ以上選択します。各層は、サンプルをさらに細分割します。層 1 と層 2 にある独立変数が 1 つの場合、その結果は 1 つのクロス表に表示され、独立変数ごとに別の表が作成されることはありません。
- オプションの統計量、分散分析表、イータ、イータの 2 乗、*R*、R^2 を使用する場合は、[**オプション**] をクリックします。

グループの平均のオプション

図 18-2
[グループの平均: オプション] ダイアログ ボックス



各グループ化変数のカテゴリ内の変数に対して、サブグループ統計量である合計、ケース数、平均値、中央値、グループ中央値、平均値の標準誤差、最小値、最大値、範囲、グループ化変数の最初のカテゴリの変数値、グループ化変数の最後のカテゴリの変数値、標準偏差、分散、尖度、尖度の標準誤差、歪度、歪度の標準誤差、総和のパーセント、総数のパーセント、グループ化変数での合計のパーセント、グループ化変数でのケース数のパーセント、幾何平均、調和平均のうち 1 つ以上を選択できます。サブグループの統計量は、表示順を変更できます。[セル統計量] リストに表示される統計量の表示順は、出力の表示順です。カテゴリ全体で各変数に対する要約統計量も表示されます。

最初。データ ファイルで発生した最初のデータ値を表示します。

幾何平均。データの値の積の n 乗根です。n はケースの数を表します。

グループ中央値。グループに対してコード化されたデータについて計算された中央値。たとえば、年齢データで、30 代の各値が 35 にコード化され、40 代の各値が 45 にコード化されるとすると、グループ中央値はコード化されたデータから計算された中央値になります。

調和平均。グループ間でサンプルの大きさが等しくないときに、グループの大きさの平均を予測するために使われます。調和平均は、サンプルの合計をサンプルの大きさの逆数の和で割ったものです。

尖度。観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、観測値の群がりが正規分布のものより大きく、長い裾を持つことを意味します。負の尖度は、観測値の群がりが正規分布のものより小さく、短い裾を持つことを意味します。

最後。データ ファイルで発生した最後のデータ値を表示します。

最大。数値型変数の最大値。

平均。中心傾向の測定値。観測値の合計をケース数で割った算術平均。

中央値。ケースの中央付近にある値です。50 パーセンタイルです。ケース数が偶数の場合、中央値は、昇順または降順に保存されたときの 2 つのまん中のケースの平均になります。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。

最小。数値型変数がとる最も小さい値。

N。ケース (観測値やレコード) の数。

総数のパーセント。各カテゴリのケースの総数のパーセントです。

総和のパーセント。各カテゴリの総和のパーセントです。

OK (ファイルオープン時のオプション)。数値型変数の最大値と最小値の差。

歪度。分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。

尖度の標準誤差。標準誤差に対する尖度の比率は、正規性の検定として使うことができます (すなわち、比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。尖度として大きな正の値は、分布の裾が正規分布の裾より長いことを示し、尖度として負の値は短い裾を示します (箱型の一様分布の裾のようになります)。

歪度の標準誤差。標準誤差に対する歪度の比率は、正規性の検定として使うことができます (すなわち、比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。歪度として大きな正の値は、右の裾が長いことを示し、極端な負の値は左の裾が長いことを示します。

合計。欠損値のないすべてのケースに対する変数の値の合計または全体。

分散 (信頼性分析)。平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。

第 1 層の統計

分散分析表とイータ。一元配置分散分析表を表示して、最初の層内にある各独立変数にイータおよびイータ 2 乗（連関度）を表示します。

線型性の検定。F 比、R および R2 乗以外に線型および非線型成分と関連する平方和、自由度および平均平方を計算します。独立変数が短い文字型の場合、線型性の検定は計算されません。

章
19

OLAP キューブ

[OLAP (Online Analytical Processing) キューブ] 手続きは、1 つ以上のカテゴリ グループ化変数のカテゴリ内で、連続集計変数の合計、平均値、その他の 1 変量の統計量を計算します。各グループ化変数のカテゴリごとに別々の層がテーブルに作成されます。

例。地域別売上げの合計と平均、地域内の製品群など。

統計量。選択対象のサブグループ統計量としては、合計、ケースの数、平均値、中央値、グループの中央値、平均値の標準誤差、最小値、最大値、範囲、グループ化変数の最初のカテゴリの変数値、グループ化変数の最後のカテゴリの変数値、標準偏差、分散、尖度、尖度の標準誤差、歪度、歪度の標準誤差、ケース合計のパーセント、集計合計のパーセント、グループ化変数内のケース合計のパーセント、グループ化変数内の集計合計のパーセント、幾何平均と調和平均があります。

データ。集計変数は量的変数 (区間尺度または比率尺度で測定される連続変数) で、グループ化変数はカテゴリ変数です。カテゴリ変数の値は、数値または文字型です。

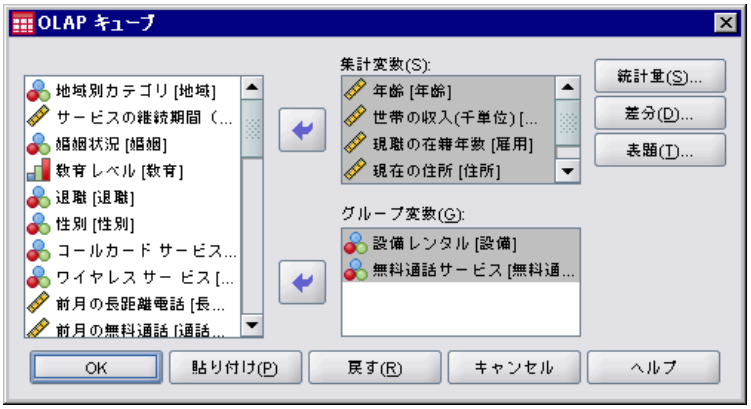
仮定。オプションのサブグループ統計量の中には、平均値や標準偏差などのように、通常の理論に基づいていて、対称的分布を持つ量的変数に適しているものがあります。頑健な統計量 (中央値や範囲など) は、正規性の仮定に適合する量的変数にも、適合しない量的変数にも適しています。

OLAP キューブを行うには

▶ メニューから次の項目を選択します。

分析
報告書
OLAP キューブ...

図 19-1
[OLAP キューブ] ダイアログ ボックス



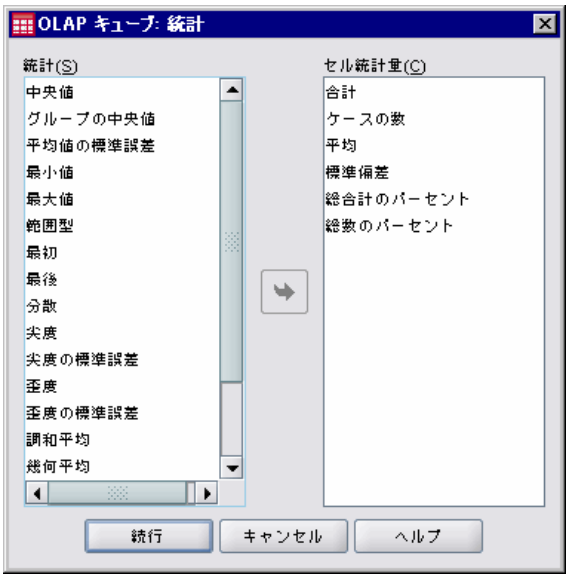
- 連続型の集計変数を 1 つ以上選択します。
- カテゴリ型のグループ化変数を 1 つ以上選択します。

次のオプションが選択できます。

- 別の要約統計量を選択する ([**統計**] をクリックします)。要約統計量を選択する前に、1 つ以上のグループ変数を選択します。
- 変数のペア、およびグループ化変数によって定義されるグループのペアの差分を計算する ([**差分**] をクリックします)。
- カスタム テーブルの表題を作成する ([**表題**] をクリックします)。

OLAP キューブの統計

図 19-2
[OLAP キューブ: 統計] ダイアログ ボックス



各グループ化変数のカテゴリ内の集計変数に対し、サブグループ統計量である合計、ケース数、平均値、中央値、グループ中央値、平均値の標準誤差、最小値、最大値、範囲、グループ化変数の最初のカテゴリの変数値、グループ化変数の最後のカテゴリの変数値、標準偏差、分散、尖度、尖度の標準誤差、歪度、歪度の標準誤差、総ケースのパーセント、総和のパーセント、グループ化変数内の総ケースのパーセント、グループ化変数内の総和のパーセント、幾何平均、調和平均のうち 1 つ以上を選択できます。

サブグループの統計量は、表示順を変更できます。[セル統計量] リストに表示される統計量の表示順は、出力の表示順です。カテゴリ全体で各変数に対する要約統計量も表示されます。

最初。データ ファイルで発生した最初のデータ値を表示します。

幾何平均。データの値の積の n 乗根です。n はケースの数を表します。

グループ中央値。グループに対してコード化されたデータについて計算された中央値。たとえば、年齢データで、30 代の各値が 35 にコード化され、40 代の各値が 45 にコード化されるとすると、グループ中央値はコード化されたデータから計算された中央値になります。

調和平均。グループ間でサンプルの大きさが等しくないときに、グループの大きさの平均を予測するために使われます。調和平均は、サンプルの合計をサンプルの大きさの逆数の和で割ったものです。

尖度。観測値が中心の周りに群がる度合いの測定値。正規分布の場合、尖度統計値は 0 です。正の尖度は、観測値の群がりが正規分布のものより大きく、長い裾を持つことを意味します。負の尖度は、観測値の群がりが正規分布のものより小さく、短い裾を持つことを意味します。

最後。データ ファイルで発生した最後のデータ値を表示します。

最大。数値型変数の最大値。

平均。中心傾向の測定値。観測値の合計をケース数で割った算術平均。

中央値。ケースの中央付近にある値です。50 パーセンタイルです。ケース数が偶数の場合、中央値は、昇順または降順に保存されたときの 2 つのまん中のケースの平均になります。中央値は、外れ値に対して敏感でない、中心化傾向の測定値です。それに対して平均値は、いくつかの極端に大きい、または小さい値に影響されます。

最小。数値型変数がとる最も小さい値。

N。ケース (観測値やレコード) の数。

グループ変数でのケース数のパーセント。他のグループ化変数のカテゴリ内における、指定されたグループ化変数のケース数のパーセント。グループ化変数が 1 つしかない場合は、ケースの総数のパーセントと同じになります。

合計のパーセント。他のグループ化変数のカテゴリ内における、指定されたグループ化変数の合計のパーセント。グループ化変数が 1 つしかない場合は、総和のパーセントと同じになります。

総数のパーセント。各カテゴリのケースの総数のパーセントです。

総合計のパーセント。各カテゴリの総和のパーセントです。

OK (ファイルオープン時のオプション)。数値型変数の最大値と最小値の差。

歪度。分布の非対称の測定値。正規分布は対称で、歪度は 0 となります。有意な正の歪度を持つ分布では、右の裾が長くなります。有意な負の歪度を持つ分布では、左の裾が長くなります。一般に、歪度がその標準誤差の 2 倍より大きい場合は、正規分布から逸脱していると考えられます。

尖度の標準誤差。標準誤差に対する尖度の比率は、正規性の検定として使うことができます (すなわち、比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。尖度として大きな正の値は、分布の裾が正規分布の裾より長いことを示し、尖度として負の値は短い裾を示します (箱型の一様分布の裾のようになります)。

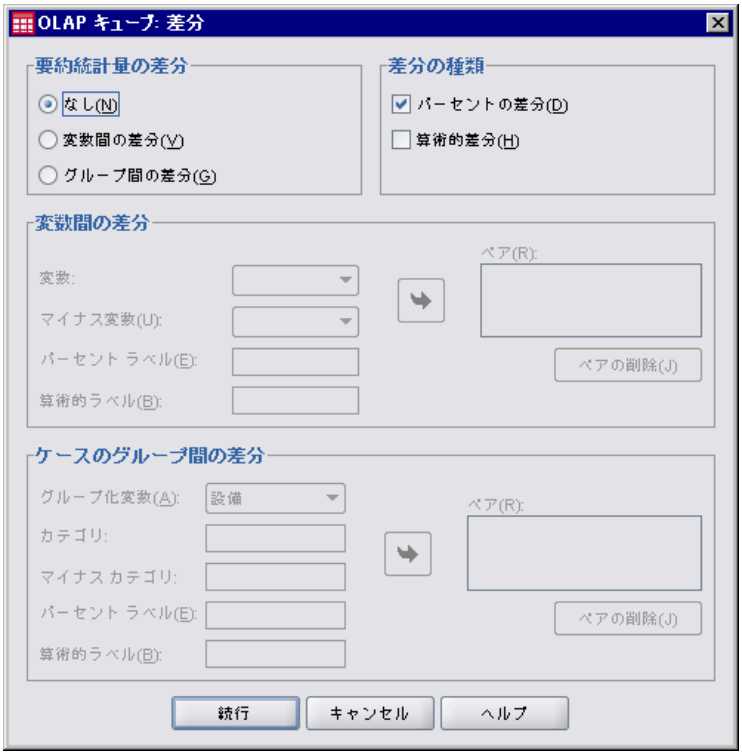
歪度の標準誤差。標準誤差に対する歪度の比率は、正規性の検定として使うことができます (すなわち、比率が -2 より小さいか +2 より大きい場合は、正規性を棄却することができます)。歪度として大きな正の値は、右の裾が長いことを示し、極端な負の値は左の裾が長いことを示します。

合計。欠損値のないすべてのケースに対する変数の値の合計または全体。

分散 (信頼性分析)。平均値のまわりの値の散らばりの程度。平均値からの偏差の平方和を、有効観測値の合計数から 1 を引いたもので割って求めます。分散の単位はその変数の単位の 2 乗です。

OLAP キューブの差分

図 19-3
[OLAP キューブ: 差分] ダイアログ ボックス



このダイアログ ボックスでは、集計変数間、またはグループ化変数によって定義されるグループ間の算術的な差分や割合を計算できます。差分は、[OLAP キューブ: 統計] ダイアログ ボックスで選択したすべての測定方法で計算されます。

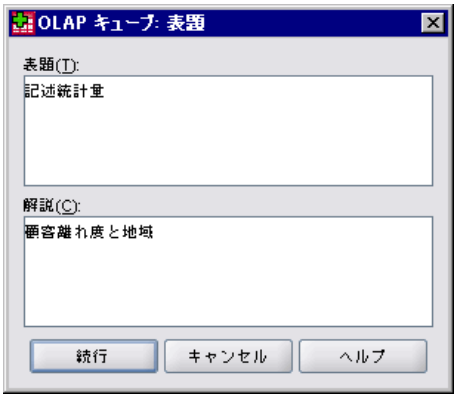
変数間の差分。変数のペア間の差分を計算します。各ペアにおいて、2 番目の変数 ([マイナス変数]) の要約統計量の値が 1 番目の変数の要約統計量の値から引かれます。パーセントの差を求める場合は、[マイナス変数] の集計変数の値が分母として使用されます。変数間の差分を指定する前に、メイン ダイアログ ボックスで少なくとも 2 つの集計変数を選択する必要があります。

ケースのグループ間の差分。グループ変数によって定義されたグループのペア間の差分を計算します。各ペアにおいて、2 番目のカテゴリ ([マイナス カテゴリ]) の要約統計量の値が 1 番目のカテゴリの要約統計量の値から引かれます。パーセントの差を求める場合は、[マイナス カテゴリ] の要約統計量の値が分母として使

用されます。グループ間の差分を指定する前に、メイン ダイアログ ボックスで 1 つ以上のグループ化変数を選択する必要があります。

OLAP キューブの表題

図 19-4
[OLAP キューブ: 表題] ダイアログ ボックス



出力の表題を変更したり、出力テーブルの下に解説を加えたりできます。また、表題や解説の改行を制御することもできます。テキスト内の必要な位置で「¥n」と入力すると、その位置で改行されます。

章
20

t 検定

t 検定には、次の 3 種類があります。

[独立したサンプルの t 検定]（2 サンプル t 検定）。2 グループのケースについて 1 つの変数の平均値を比較します。各グループの記述統計量と等分散性の Levene の検定の他に、分散が等しい場合および分散が等しくない場合の *t* 値と平均値の差の 95% 信頼区間が得られます。

[対応のある t 検定]（独立 t 検定）。1 つのグループについて 2 つの変数の平均値を比較します。この検定はまた、一致しているペアまたはケース コントロール研究の計画のための検定です。出力には、検定変数の記述統計量、変数間の相関係数、対応間の差の記述統計量、*t* 検定、および 95% 信頼区間が含まれています。

1 サンプルの t 検定。既知の値または仮説値を 1 変数の平均値と比較します。検定変数の記述統計量は、*t* 検定とともに表示されます。検定変数の平均値と仮説検定値の差の 95% 信頼区間は、デフォルト出力に含まれています。

独立したサンプルの t 検定

[独立サンプルの t 検定] 手続きでは、2 つのグループのケースによる平均値を比較します。理想的には、この検定の場合、被検者を 2 個のグループに対して無作為に割り当て、応答の差が他の要素によるものでなく、処置（または処置の欠如）によるものとします。男性および女性の平均収入を比較する場合、この検定は当てはまりません。ある被検者が男性または女性に無作為に割り当てられていません。そのような状況においては、他の要素における差が平均値の有意差を隠したり、または大きくしないようにします。平均収入の差は、教育水準などの要素により影響を受けるかもしれません（性別だけによる影響は受けないでしょう）。

例。高血圧の患者を偽薬グループと治療グループに無作為に指定します。偽薬の被検者には効き目のない錠剤を投与し、治療グループには血圧を下げる効力があると考えられる新薬を投与します。被検者の 2 か月間の治療後に、2 サンプルの *t* 検定を使用して、偽薬グループおよび治療グループの平均血圧を比較します。各患者はそれぞれ 1 回の測定を受け、1 つのグループに所属します。

統計量。各変数に対して: サンプル サイズ、平均値、標準偏差、および平均値の標準誤差。平均値の差に対して: 平均値、標準誤差、および信頼区間（信頼水準を指定できます）。検定: 等分散性の Levene 検定、および 2 つの母平均の差のプールされた分散ならびに等分散でないときの *t* 検定。

データ。任意の量的変数の値がデータ ファイルの 1 つの列にあります。この手続きでは、2 つの値を持ったグループ化変数を使用して当該ケースを 2 つのグループに分けます。グループ化変数は、数値型変数 (1 や 2、または 6.25 や 12.5) または短い文字型変数 (**はい**と **いいえ**など) のどちらかです。また別の方法として、量的変数 (**年齢**など) を使用して、分割値を指定することにより、ケースを 2 つのグループに分割することもできます (分割値を 21 にすると **年齢**は 21 未満のグループと 21 以上のグループに分割されます)。

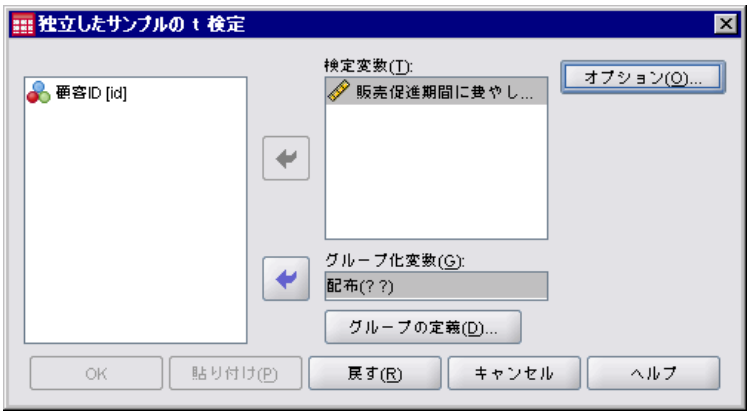
仮定。等分散の *t* 検定の場合、観測値は、同じ母集団分散を持つ正規分布からの独立した無作為サンプルでなくてはなりません。等分散でない *t* 検定の場合、観測値は、正規分布からの独立した無作為サンプルでなくてはなりません。2 サンプルの *t* 検定は、正規性からの逸脱に対して非常に頑健です。分布をグラフとしてチェックするときには、分布が対称的であり、さらに外れ値が皆無であることを確認します。

独立したサンプルの t 検定を行うには

▶ メニューから次の項目を選択します。

分析
平均の比較
独立したサンプルの T 検定

図 20-1
[独立したサンプルの t 検定] ダイアログ ボックス



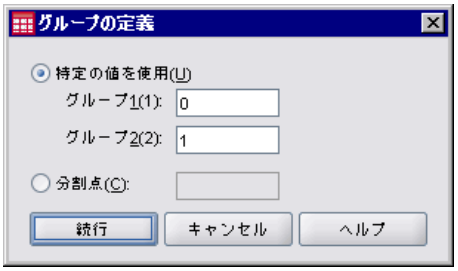
▶ 1 つ以上の量的検定変数を選択します。*t* 検定が変数ごとに行われます。

▶ 変数のリストから 1 つのグループ化変数を選択して [グループ化変数] ボックスに移動し、[**グループの定義**] をクリックして比較するグループの 2 つのコードを指定します。

▶ オプションとして、[**オプション**] をクリックして、欠損データの扱いと信頼区間の水準を指定します。

独立したサンプルの t 検定のグループの定義

図 20-2
数値型変数の [グループの定義] ダイアログ ボックス



数値型のグループ化変数では、2 つの値または分割値を指定して、*t* 検定を行う 2 つのグループを定義します。

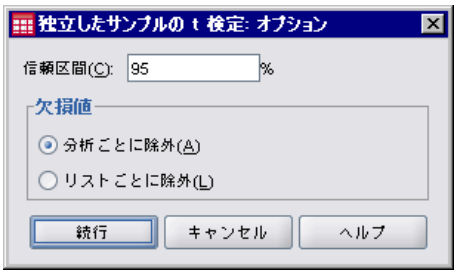
- **特定の値を使用。**[グループ 1] と [グループ 2] に 1 つずつ値を入力します。その他の値を持つケースは分析から除外されます。数字は整数でなくてもかまいません (たとえば、6.25 や 12.5 でも有効です)。
- **分割値。**グループ化変数の値を 2 つのグループに分割する数字を入力します。分割値未満の値のケースが一方のグループを形成し、分割値以上の値のケースは他の一方のグループを形成します。

図 20-3
文字型変数の [グループの定義] ダイアログ ボックス

短い文字型のグループ化変数では、[グループ 1] と [グループ 2] に 1 つずつ文字列 (**はい**と **いいえ**など) を入力します。他の文字列を持つケースは分析から除外されます。

独立したサンプルの t 検定のオプション

図 20-4
[独立したサンプルの t 検定: オプション] ダイアログ ボックス



信頼区間。デフォルトでは、平均値の差の 95% 信頼区間を表示します。1 から 99 の範囲の数値を入力して、別の信頼水準を表示することもできます。

欠損値。複数の変数を検定する場合で、1 つ以上の変数に対して欠損している場合、どのケースを含める (または除外する) かを指示できます。

- **分析ごとに除外。**各 *t* 検定では、検定する変数のデータが有効なケースをすべて使用します。したがって、サンプル サイズが検定ごとに変化します。
- **リストごとに除外。**各 *t* 検定では、要求された *t* 検定において使用されるすべての変数に対して有効なデータをもつケースのみを使用します。したがって、サンプル サイズが検定を通じて一定になります。

対応のある T 検定

[対応のあるサンプルの t 検定] 手続きでは、1 つのグループの 2 つの変数の平均を比較します。手続きは各ケースの 2 つの変数間の差を計算し、平均が 0 と異なるかどうかを検定します。

例。高血圧に関する調査で、すべての患者を調査の開始時に測定し、治療後に再度測定します。このように、各被験者には 2 つの測定値があり、多くの場合、**測定前**と**測定後**と呼ばれます。この検定が使用されるもう 1 つの計画は、一致するペアの研究またはケース コントロール研究です。ここで、データ ファイルの各記録には患者および一致する対照被験者の回答が含まれます。血圧の調査では患者と対照被験者を、年齢で (75 歳の患者と 75 歳の対照グループメンバーを) 一致させることもできます。

統計量。各変数に対して: 平均値、サンプル サイズ、標準偏差、および平均値の標準誤差。変数の各ペアについて: 相関係数、平均値の差の平均、*t* 検定、平均値の差に対する信頼区間 (信頼水準は指定可能)。標準誤差と平均値の差の標準誤差。

データ。各対応のある検定で 2 つの数量変数（区間尺度または比例尺度）を指定します。一致するペアの研究またはケース コントロール研究に対して、各検定の被験者と対になる対照被験者に対する応答は、データ ファイル内の同じケースにある必要があります。

仮定。ペアに対する観測は同じ条件で行われる必要があります。また、平均値の差は正規分布していなくてはなりません。各変数の分散は、等しい場合と異なる場合があります。

対応のあるサンプルの t 検定を行うには

▶ メニューから次の項目を選択します。

分析
  平均の比較
    対応のあるサンプルの T 検定...

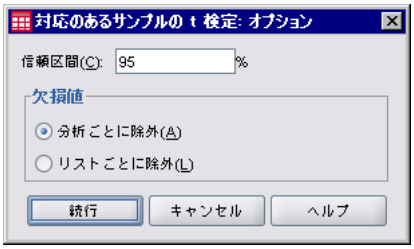
図 20-5
[対応のあるサンプルの t 検定] ダイアログ ボックス

▶ 1 つ以上の変数ペアを選択

▶ オプションとして、[**オプション**] をクリックして、欠損データの扱いと信頼区間の水準を指定します。

対応のあるサンプルの t 検定のオプション

図 20-6
[対応のあるサンプルの t 検定: オプション] ダイアログ ボックス



信頼区間。デフォルトでは、平均値の差の 95% 信頼区間を表示します。1 から 99 の範囲の数値を入力して、別の信頼水準を表示することもできます。

欠損値。複数の変数を検定する場合で、1 つ以上の変数に対して欠損している場合、どのケースを含める (または除外する) かを指示できます。

- **分析ごとに除外。**各 *t* 検定では、検定する変数のペアに対して有効なデータを持つすべてのケースを使用します。したがって、サンプル サイズが検定ごとに変化します。
- **リストごとに除外。**各 *t* 検定では、検定する変数のすべてのペアに対して有効なデータを持つケースだけを使用します。したがって、サンプル サイズが検定を通じて一定になります。

1 サンプルの t 検定

[1 サンプルの t 検定] 手続きでは、単一の変数の平均値が指定された定数と異なっているかどうかを調べます。

例。研究者は、学生のグループの平均 IQ 値が 100 と異なるかどうかを分析したいと考えています。または、シリアル製品の製造業者は、生産ラインからサンプルとして幾箱かを取り、サンプルの平均重量が 95% 信頼水準で 590g であるかどうかを調べます。

統計量。各検定変数: 平均値、標準偏差、平均値の標準誤差。各データ値と仮説検定値の差の平均、それが 0 であることを検定する *t* 検定とその信頼区間 (信頼水準を指定できます)。

データ。仮説検定値に対する量的変数の値を検定するには、量的変数と仮説検定値を選択します。

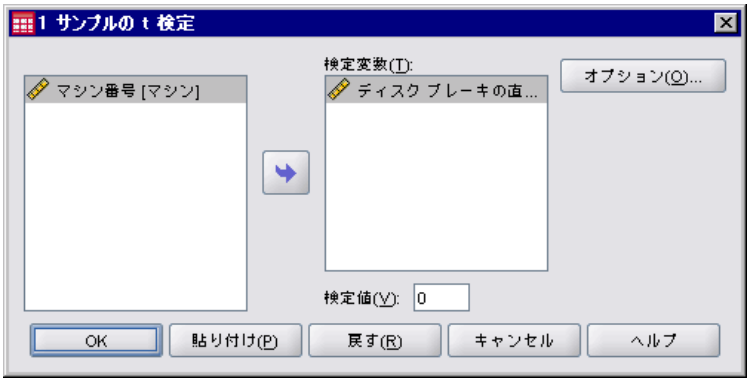
仮定。この検定では、データが正規分布しているものと仮定していますが、データが正規性から逸脱している場合にも、かなり頑健です。

1 サンプルの t 検定を行うには

▶ メニューから次の項目を選択します。

分析
  平均の比較
    1 サンプルの T 検定...

図 20-7
[1 サンプルの t 検定] ダイアログ ボックス



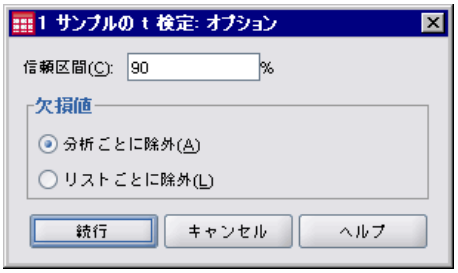
▶ 仮説値に対して検定する 1 つ以上の変数を選択します。

▶ 各サンプルの平均値と比較する数値を [検定値] ボックスに入力します。

▶ オプションとして、[**オプション**] をクリックして、欠損データの扱いと信頼区間の水準を指定します。

1 サンプルの t 検定のオプション

図 20-8
[1 サンプルの t 検定オプション] ダイアログ ボックス



信頼区間。デフォルトでは、平均値と仮説検定値との差の 95% 信頼区間を表示します。1 から 99 の範囲の数値を入力して、別の信頼水準を表示することもできます。

欠損値。複数の変数を検定する場合で、1 つ以上の変数に対して欠損している場合、どのケースを含める (または除外する) かを指示できます。

- **分析ごとに除外。**各 *t* 検定では、検定する変数のデータが有効なケースをすべて使用します。したがって、サンプル サイズが検定ごとに変化します。
- **リストごとに除外。**各 *t* 検定では、要求された *t* 検定で使用されるすべての変数のデータが有効なケースだけを使用します。したがって、サンプル サイズが検定を通じて一定になります。

t 検定コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 1 サンプルと独立サンプルの両 t 検定を、単一コマンドで実行することができます。
- リストに記載されている各変数に対する変数の検定を、対応のある t 検定で行なうことができます (PAIRS サブコマンド使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章

21

一元配置分散分析

[一元配置分散分析] 手続きは、量的従属変数に対して一元配置分散分析を一因子 (独立) 変数別に作成します。分散分析は、いくつかの平均値は等しいという仮説を検定するときに使用します。この手法は、2 サンプルの *t* 検定の拡張です。

平均値間に差があることを判断する以外に、どの平均値が違うかを知ることもできます。平均値を比較するための検定には [対比] および [その後の検定] という 2 種類があります。対比は実験を実行する**前**に設定された検定であり、その後の検定は実験が実行された**後**で実行される検定です。カテゴリ全体の傾向を検定することもできます。

例。ドーナツは料理をするとかなりの量で脂肪を吸収します。実験は 3 種類の脂肪、すなわち、ピーナッツ油、コーン油、およびラードに関して設定します。ピーナッツ油とコーン油は不飽和の脂肪で、ラードは飽和脂肪です。吸収された脂肪の量が使用した脂肪のタイプに関係しているどうかを判断することができ、さらに事前対比を設定すると脂肪の吸収量が飽和脂肪と不飽和脂肪では違うかどうかを判断することもできます。

統計量。各グループの、ケースの数、平均値、標準偏差、平均値の標準誤差、最小値、最大値、および平均値の 95% 信頼区間。Levene の等分散性の検定、各従属変数に対する分散分析表および平均値同等性の耐久検定、ユーザーが指定した事前対比、およびその後の範囲検定と多重比較。すなわち、Bonferroni、Sidak、Tukey の HSD、Hochberg の GT2、Gabriel の方法、Dunnett の方法、Ryan-Einot-Gabriel-Welsch の *F* 検定 (R-E-G-W *F*)、Ryan-Einot-Gabriel-Welsch の範囲検定 (R-E-G-W *Q*)、Tamhane の T2、Dunnett の T3、Games-Howell の方法、Dunnett の *C*、Duncan の多重範囲検定、Student-Newman-Keuls (S-N-K)、Tukey の *b*、Waller-Duncan の方法、Scheffé の方法、および最小有意差。

データ。因子変数の値は整数でなければならず、さらに従属変数は量的変数 (区間尺度の測定) でなければなりません。

仮定。各グループは、正規母集団から無作為に抽出された互いに独立したサンプルです。データは対称であるべきですが、分散分析は正規性からの逸脱に対し頑健です。グループは分散の等しい母集団から発生していなければなりません。この仮定を検定するには、Levene の等分散性の検定を使用します。

一元配置分散分析を行うには

▶ メニューから次の項目を選択します。

分析
平均の比較
一元配置分散分析...

図 21-1
[一元配置分散分析] ダイアログ ボックス

▶ 従属変数を 1 つ以上選択します。

▶ 変数のリストから 1 つの独立因子変数を選択して、[因子] ボックスに移動します。

一元配置分散分析の対比

図 21-2
[一元配置分散分析: 対比] ダイアログ ボックス

グループ間平方和をトレンド成分に分割したり、事前対比を指定できます。

多項式。グループ間平方和を傾向成分に分割します。因子変数の順序付けされた水準全体で従属変数の傾向を検定することができます。たとえば、得た最高の順位の水準で給与の線形傾向（増加または減少）を検定することができます。

- **次数。**1 次、2 次、3 次、4 次、または 5 次の直交多項式を選択できます。

係数。*t* 統計量で検定するユーザー指定の事前対比。因子変数の各グループ（カテゴリ）について係数を入力し、入力するごとに [**追加**] をクリックします。それぞれの新しい値は係数リストの下部に加えられます。対比の追加グループを指定するには、[**次**] をクリックします。[**次**] と [**前**] を使うと、対比の設定の間を移動できます。

係数の次数は、因子変数のカテゴリ値の昇順に対応するので重要です。リスト上の最初の係数は因子変数の最も小さいグループの値に対応し、最後の係数は最も大きい値に対応します。たとえば、因子変数のカテゴリが 6 つある場合、係数 –1、0、0、0、0.5、0.5 は最初のグループを 5 番目と 6 番目のグループと対比させます。ほとんどの場合、係数は合計して 0 になる必要があります。合計して 0 にならないような設定を使うこともできますが、警告メッセージが表示されます。

一元配置分散分析のその後の検定

図 21-3
[一元配置分散分析: その後の多重比較] ダイアログ ボックス

平均値の間に差があることが判明した後は、その後の範囲検定とペアごとの多重比較により、どの平均値が相異しているのかを決めることができます。範囲検定は、互いに平均値に差がない等質サブグループを識別します。ペアごとの多重比較はそれぞれのペアごとの平均値の差を検定して、5% 水準で有意な差があるグループの平均値には星印を付けます。

等分散が仮定されている

Tukey の HSD 検定、Hochberg の GT2、Gabriel の検定、および Scheffé の検定は、多重比較検定と範囲検定の両方です。利用できるその他の範囲検定としては、Tukey の *b* 検定、S-N-K (Student-Newman-Keuls)、Duncan、R-E-G-W *F* (Ryan-Einot-Gabriel-Welsch *F* 検定)、R-E-G-W *Q* (Ryan-Einot-Gabriel-Welsch 範囲検定)、および Waller-Duncan の方法があります。利用できる多重比較検定は、Bonferroni、Tukey の HSD 検定、Sidak、Gabriel、Hochberg、Dunnett の方法、Scheffé、および LSD (最小有意差) があります。

- **最小有意差。**t 検定を使って、グループ平均間のすべてのペアごとの比較を実行します。多重比較の誤差率は調整されません。
- **Bonferroni の方法。**t 検定を使ってグループ平均のペアごとの比較を行いますが、実験ごとの誤差率を総検定数で割った値に各検定の誤差率を設定することによって、全体の誤差率を制御します。したがって、有意確率は、多重比較がなされているとして調整されます。
- **Sidak。**t 検定に基づいたペアごとの多重比較検定。Sidak の方法は、多重比較の有意確率を調整して、Bonferroni の方法より厳しい限界を設定します。
- **Scheffe の検定。**平均値の可能なペアごと組み合せに対して、ペアごとの同時比較を実行します。F 分布を使います。ペアごとの比較だけでなく、グループ平均のすべての可能な線型結合を調べるために使うこともできます。
- **R-E-G-W の F。**F 検定に基づいた Ryan-Einot-Gabriel-Welsch の多重ステップダウン手続き。
- **R-E-G-W の Q。**スチューデント化された範囲に基づいた Ryan-Einot-Gabriel-Welsch の多重ステップダウン手続き。
- **S-N-K。**スチューデント化された範囲の分布を使って、平均値間のすべてのペアごとの比較を行います。サンプル サイズが等しい場合は、ステップワイズ法手続きを使って等質サブセット内の平均値のペアも比較します。平均値は最も高いものから最も低いものへと順序付けられ、極値の差が最初に検定されます。
- **Tukey。**スチューデント化された範囲の統計量を使ってグループ間のすべてのペアごとの比較を行います。すべてのペアごとの比較の集合に対する誤差率に、実験ごとの誤差率を設定します。
- **Tukey の b。**スチューデント化された範囲の分布を使ってグループ間のペアごとに比較を行います。臨界値は、Tukey の HSD 検定と Student-Newman-Klaus 検定に対応する値の平均です。
- **Duncan。**Student Newman Keuls 検定で使用される次数と同一の、ステップごとの比較次数を使ってペアごとに比較しますが、個々の検定に対する誤差率ではなく、検定の集合の誤差率に対する保護水準を設定します。スチューデント化された範囲の統計量を使います。
- **Hochberg の GT2。**スチューデント化された最大絶対値を使う多重比較と範囲検定。Tukey の HSD 検定に似ています。

- **Gabriel。**スチューデント化された最大法を使用したペアごとの比較検定。一般に、セルの大きさが等しくないときには Hochberg の GT2 より強力です。セルの大きさのばらつきが大きい場合には、Gabriel の検定の方が公平になることがあります。
- **Waller-Duncan。**T 統計量に基づいた多重比較検定。ベイズ派アプローチを使用します。
- **Dunnett。**1 組の実験群を単一の対照 (コントロール) と比較するペアごとの多重比較 T 検定。最後のカテゴリがデフォルトの対照カテゴリとなります。代わりに、最初のカテゴリを選択できます。**[両側]** は、因子の任意のレベルの平均値 (対照カテゴリを除く) が対照カテゴリの平均値と等しくないことを検定します。**[<対照カテゴリ]** は、因子の任意のレベルの平均値が対照カテゴリの平均値よりも小さいことを検定します。**[>対照カテゴリ]** は、因子の任意のレベルの平均値が対照カテゴリの平均値よりも大きいことを検定します。

等分散が仮定されていない

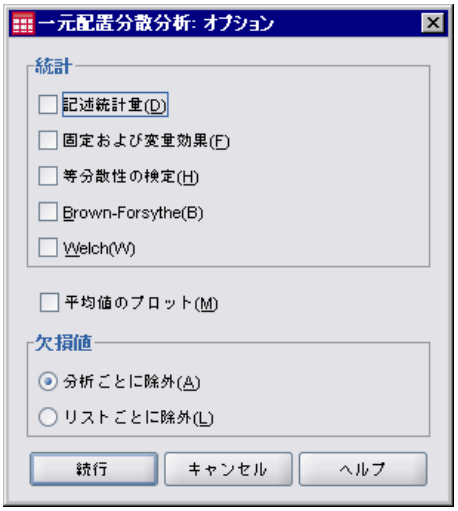
等分散を仮定しない多重比較検定は、Tamhane の T2、Dunnett の T3、Games-Howell、および Dunnett の *C* です。

- **Tamhane の T2。**t 検定に基づいた保守的なペアごとの比較検定。この検定は、分散が等しくないときに適しています。
- **Dunnett の T3。**スチューデント化された最大法に基づいたペアごとの比較検定。この検定は、分散が等しくないときに適しています。
- **Games-Howell。**時々公平なペアごとの比較検定。この検定は、分散が等しくないときに適しています。
- **Dunnett の C。**スチューデント化された範囲に基づいたペアごとの比較検定。この検定は、分散が等しくないときに適しています。

注: [テーブル プロパティ] ダイアログ ボックス (ピボット テーブルをアクティブにして、[書式] メニューの **[テーブル プロパティ]** を選択します) の [全般] タブで **[空白の行と列を隠す]** のチェックを外すと、その後の検定からの出力を解釈しやすくなる場合があります。

一元配置分散分析のオプション

図 21-4
[一元配置分散分析: オプション] ダイアログ ボックス



統計量。次の 1 つ以上を選択します。

- **記述統計量。**各グループに対してケースの数、平均値、標準偏差、平均値の標準誤差、最小値、最大値、および各従属変数の 95% 信頼区間を計算します。
- **固定および変量効果。**固定効果モデルの標準偏差、標準誤差、95% 信頼区間、および変量効果モデルの標準誤差、95% 信頼区間、成分間の推定分散を表示します。
- **等分散性の検定。**グループの分散の等質性を検定するため Levene の統計を計算します。この検定は、正規性の仮定に依存しません。
- **Brown-Forsythe。**グループ平均値の等質性を検定するため Brown-Forsythe の統計を計算します。等分散仮定が適用できない場合は、この統計量が *F* 統計量よりも適しています。
- **Welch。**グループ平均値の等質性を検定するため Welch の統計を計算します。等分散仮定が適用できない場合は、この統計量が *F* 統計量よりも適しています。

平均値のプロット。サブグループ平均値 (因子の値によって定義されたグループごとの平均値) をプロットするグラフを表示します。

欠損値。欠損値の処理を管理します。

- **分析ごとに除外。**ある分析での従属変数または因子変数のどちらかに欠損値があるケースは、その分析で使用されません。さらに、因子変数で指定した範囲外のケースも使用されません。
- **リストごとに除外。**因子変数またはメイン ダイアログ ボックスの従属変数リストに取り込まれた従属変数に欠損値があるケースは、すべての分析から除外されます。複数の従属変数を指定していない場合、影響はありません。

ONEWAY コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 固定効果および変量効果の統計を取得します。固定効果モデルの標準偏差、標準誤差、95% 信頼区間。変量効果モデルの標準誤差、95% 信頼区間、成分間の推定分散 (STATISTICS=EFFECTS 使用)。
- 最小有意差、Bonferroni、Duncan、Scheffé の各多重比較検定に対するアルファ レベルを指定します (RANGES サブコマンド使用)。
- 平均行列、標準偏差と度数の書き込み、または、平均行列、度数、プールされた分散とプールされた分散の自由度の読み込みを行ないます。これらの行列を生データの代わりに使用して、一元配置分散分析を取得します (MATRIX サブコマンド使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
22

GLM – 1 変量分散分析

[GLM – 1 変量分散分析] 手続きは、1 つの従属変数に対する回帰分析や分散分析を 1 つ以上の因子や変数を使用して行う手続きです。因子変数により、母集団をいくつかのグループに分けます。この [一般的な線型モデル] 手続きを使用すると、1 つの従属変数をさまざまに分けたグループの平均値に対する他の変数の効果について、帰無仮説を検定できます。因子間の交互作用や因子ごとの効果を調べることができ、その一部は任意にすることもできます。さらに、共変量の効果や共変量と因子の交互作用を含めることができます。回帰分析では、独立 (予測) 変数は共変量として指定します。

検定は、釣り合い型モデルと不釣り合い型モデルの両方に対して実行できます。モデル内の各セルに含まれているケース数が等しければ、その計画は釣り合っています。[GLM – 1 変量分散分析] 手続きでは、仮説の検定の他に、パラメータの推定値を生成します。

仮説を検定する際は、一般的に用いられている事前対比を利用できます。さらに、全体的な *F* 検定で有意確率が判明していれば、その後の検定を使用して、特定平均値間の差分を評価できます。推定周辺平均から、モデルに含まれるセルの予測平均値を推定できるとともに、こうした平均値のプロファイル プロット (交互作用プロット) を使用して関係の一部を簡単に視覚化できます。

残差、予測値、Cook の距離、てこ比の値は、データ ファイルに新変数として保存し、仮定の確認に利用できます。

WLS 重みで、測定方法ごとに異なる精度を補正するなどの目的で、観測値に重み付き最小 2 乗法 (WLS) 分析のためのさまざまな重みを付けるのに使用する変数を指定できます。

例。シカゴ マラソンの出場ランナーの個人データが数年分集められています。各ランナーの完走タイムが従属変数です。その他の因子には、天候 (寒い、快適、または暑い)、トレーニング月数、過去のマラソン出場回数、性別などが含まれています。年齢は共変量と見なします。ここでは、性別が有意の効果であること、そして性別と天候の交互作用が有意であるということが判明するかもしれません。

方法。異なる仮説を評価する場合は、タイプ I、タイプ II、タイプ III、タイプ IV 平方和を使用できます。デフォルトは タイプ III です。

統計量。その後の範囲検定と多重比較: 最小有意差、Bonferroni の方法、Sidak の方法、Scheffé の検定、Ryan-Einot-Gabriel-Welsch の多重 *F* 値、Ryan-Einot-Gabriel-Welsch の多重範囲、Student-Newman-Keuls の検定、Tukey の HSD 検定、Tukey の *b* 検定、Duncan の方法、Hochberg の GT2、Gabriel の方

法、Waller-Duncan の *t* 検定、Dunnett の方法 (片側と両側)、Tamhane の T2、Dunnett の T3、Games-Howell の方法、および Dunnett の *C*。記述統計: すべてのセルにおけるすべての従属変数の観測平均値、標準偏差、および度数。Levene の等分散性の検定。

作図。水準と広がりの図、残差、およびプロファイル (交互作用)。

データ。従属変数は量的変数です。因子も同じくカテゴリ型です。数値または最高 8 文字までの文字値を持つことできます。共変量は、従属変数に関連する量的変数です。

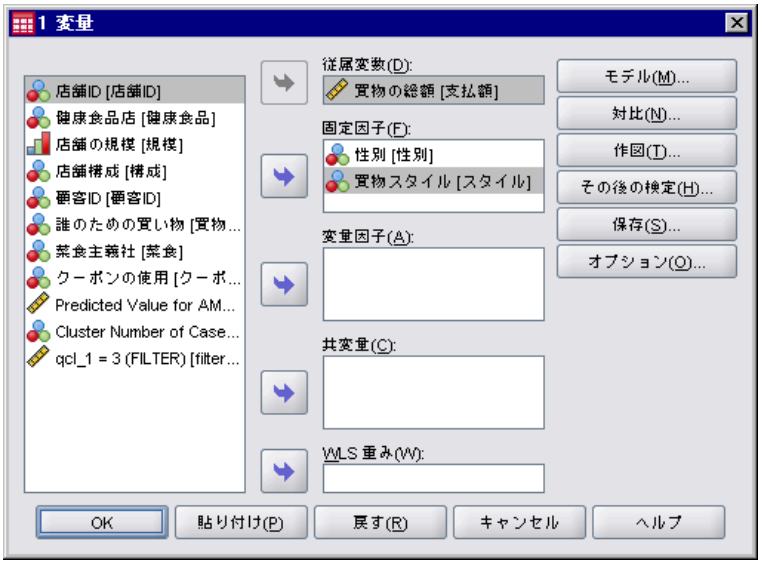
仮定。データは正規母集団からの任意のサンプルで、この母集団では、すべてセルの分散が同じです。データは対称であるべきですが、分散分析は正規性からの逸脱に対し頑健です。仮定の確認には、等分散性の検定や水準と広がりの図を使用できます。残差と残差プロットについて探索的分析を行うこともできます。

GLM - 1 変量分散分析テーブルを作成するには

▶ メニューから次の項目を選択します。

分析
一般線型モデル
1 変量...

図 22-1
[GLM 1 変量] ダイアログ ボックス



▶ [従属変数] ボックスに従属変数を選択します。

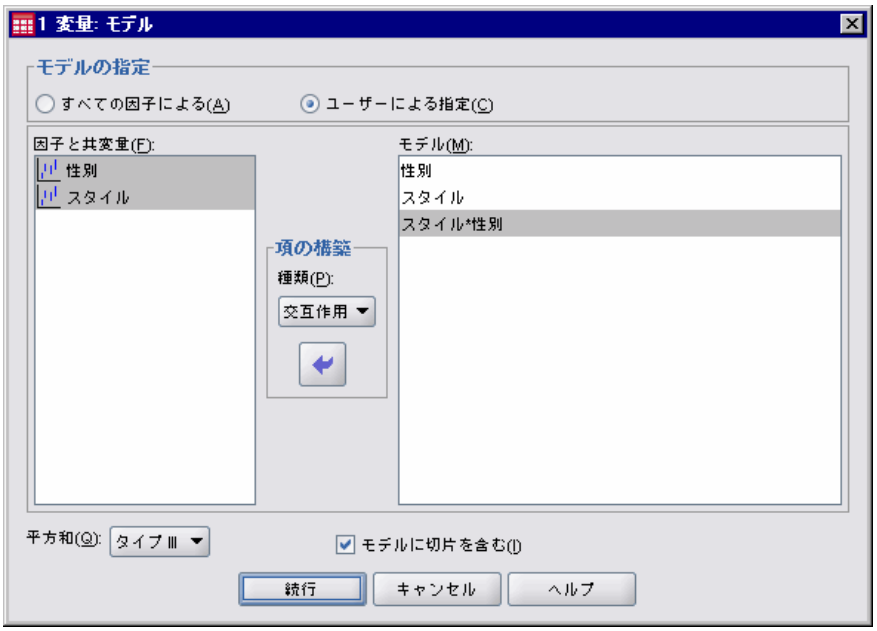
▶ データに合わせて、[固定因子]、[変量因子]、[共変量] の変数を選択します。

▶ オプションで、[WLS 重み] を使用して、重み付き最小 2 乗法分析のための重み付け変数を指定できます。重み付け変数の値がゼロ、負、または欠損の場合、

ケースは分析から除外されます。モデル内ですでに使用された変数は、重み付け変数として使うことはできません。

GLM モデル

図 22-2
[1 変量: モデル] ダイアログ ボックス



モデルの指定。すべての因子によるモデルには、因子の主効果、共変量の主効果、因子間の交互作用のすべてが含まれます。共変量の交互作用は含まれません。交互作用の一部だけを指定する場合や、因子と共変量の交互作用を指定する場合は、[**ユーザーの指定**] をクリックします。モデルに含める項目はすべて指示する必要があります。

因子と共変量。要素および共変量はリストされます。

モデル。モデルは、使用するデータの性質によって異なります。[**ユーザーの指定**] をクリックすれば、分析対象の主効果と交互作用を選択できます。

平方和。平方和の計算方法。釣り合い型モデルと不釣り合い型モデルで欠損セルがない場合は、タイプ III の平方和の方法が最もよく利用されます。

モデルに切片を含む。通常、モデルには定数項が含まれます。データが原点を通ると仮定できる場合は、切片を除外できます。

項の構築 (ロジット モデル)

選択した因子や共変量について、次の項を作成できます。

交互作用。選択した変数のすべてについて、最高水準の交互作用項を作成します。これはデフォルトです。

主効果。選択した変数のそれぞれに主効果の項を作成します。

2 次まで。選択した変数に 2 次までの交互作用を作成します。

3 次まで。選択した変数に 3 次までの交互作用を作成します。

4 次まで。選択した変数に 4 次までの交互作用を作成します。

5 次まで。選択した変数に 5 次までの交互作用を作成します。

平均平方

モデルには、平方和のタイプを選択できます。最も一般的に使われ、デフォルトでもあるのがタイプ III です。

タイプ I。この方法は、平方和の階層的分割法という名でも知られています。各項は、モデル内で先行する項目に対してだけ調整されます。タイプ I の平方和は、一般に次のような場合に使用します。

- 1 次の交互作用の効果より先になんらかの主効果が指定され、2 次の交互作用の効果より先になんらかの 1 次の交互作用の効果が指定されているといった具合の分散分析の釣り合い型モデル。
- 高次の項より先に低次の項目が指定されている多項式回帰モデル。
- 最初に指定された効果が 2 番目に指定された効果の中に入れ子になり、2 番目に指定された効果が 3 番目に指定された効果の中に入れ子になり、という具合に続く純粋な入れ子モデル。(入れ子の形式は、シンタックスを使用しなければ指定できません)。

タイプ II。この方法では、他のすべての「適当な」効果に対して調整されるモデルの効果の平方和が計算されます。適当な効果とは、調査中の効果を含んでいないすべての効果に対応するものです。タイプ II の平方和の方法は、一般に次のような場合に使用します。

- 分散分析の釣り合い型モデル。
- 因子の主効果しか持たないモデル。
- 回帰モデル。
- 純粋に入れ子になっている計画。(入れ子の形式は、シンタックスを使用して指定できます)

タイプ III。デフォルトです。この方法では、計画内にある効果の平方和を、その計画を含まないその他の効果、またはそうした効果に対して直交的な効果を対象に調整した平方和として計算します。タイプ III の平方和には、一般的な推定形式が一定の状態に保たれている限り、セル度数が変わらないという大きな利点があります。したがって、このタイプの平方和は、欠損セルがない不釣り合い型モデルに有用だと考えられます。欠損セルのない多因子計画でこの方法

に相当するのが、Yates の平均値の重み付き 2 乗法です。タイプ III の平方和の方法が一般的に使用されるのは次の場合です。

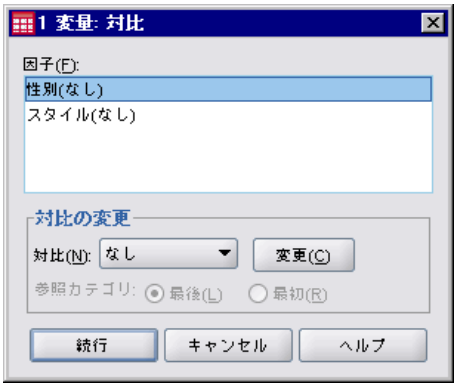
- タイプ I とタイプ II に記載されているモデル。
- 空白セルのない釣り合い型モデルまたは不釣り合い型モデル。

タイプ IV。この方法は、欠損セルが存在する状況を目的としたものです。計画における効果 *F* に対して、*F* が他のどの効果にも含まれていないとき、タイプ IV = タイプ III = タイプ II となります。*F* が他の効果に含まれているとき、タイプ IV は、*F* におけるパラメータ間で行われている対比を、より高いレベルの効果のすべてに等しく分配します。タイプ IV の平方和の方法は、一般に次のような場合に使用します。

- タイプ I とタイプ II に記載されているモデル。
- 空白セルを伴う釣り合い型モデルまたは不釣り合い型モデル。

GLM の対比

図 22-3
[1 変量: 対比] ダイアログ ボックス



1 つの因子の水準同士に差異があるかどうかを検定するときに用いられるのが対比です。対比は、モデルの因子ごと (反復測定モデルでは被験者間因子ごと) に指定できます。対比は、パラメータの線型結合を表します。

仮説の検定は、LB = 0 という帰無仮説に基づきます。この場合、L は対比係数行列で、B はパラメータのベクトルです。対比が指定されると、L 行列が作成されます。要素に対応する L 行列の列は、対比と一致します。残りの列は、L 行列が推定できるように調整されます。

出力には、対比のセットごとの *F* 統計量が含まれます。対比の差異については、スチューデントの *t* 分布に基づく Bonferroni の同時信頼区間も表示されます。

利用できる対比

利用できる対比には、偏差、単純、差分、Helmert、反復測定、多項式があります。偏差対比と単純対比については、参照カテゴリの最初か最終を選択できます。

対比の種類

全平均。各水準の平均（参照カテゴリは除きます）をすべての水準（全平均）の平均と比較します。因子の水準は任意の順序になります。

単純。各水準の平均と指定された水準の平均を比較します。この対比の型はコントロール群があるときに有効です。最初または最後のカテゴリを参照として選択できます。

差分。各レベル（最初は除く）の平均値を前のレベルの平均値と比較します。（「逆 Helmert 対比」と呼ばれることもあります）。

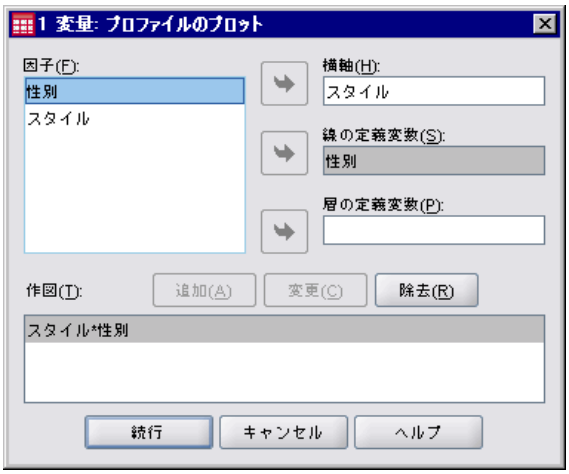
Helmert。因子の各水準（最終を除く）の平均をその後の水準の平均と比較します。

反復測定。各水準（最終を除く）の平均をそのすぐ後の水準の平均と比較します。

多項式。1 次効果、2 次効果、3 次効果などを比較します。第 1 自由度にはすべてのカテゴリを通じての 1 次効果が含まれ、第 2 自由度には 2 次効果が含まれるなど、それぞれが含まれます。このような対比は、多項式のトレンドを推定する場合に頻繁に使用します。

GLM のプロファイル プロット

図 22-4
[1 変量: プロファイルのプロット] ダイアログ ボックス

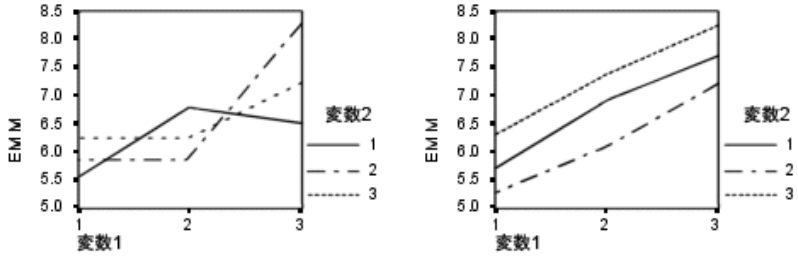


プロファイル プロット（交互作用プロット）は、モデル中の周辺平均を 比較するときに役立ちます。プロファイル プロットは、点を使って、因子の 1 つの水準で従属変数（共変量を対象に調整）の推定周辺平均を示す線のプロットです。第 2 因子の水準を使って、個別の線を作成できます。第 3 因子の各水準を使用して、層別のプロットを作成できます。固定因子と変量因子がある場合には、そのすべてをプロットに使用できます。多変量分析では、従属変数ごとにプロファイル プロットが作成されます。反復測定分析では、プロファイル プロットに被験者間因子と被験

者内因子の両方を使用できます。Advanced Models オプションをインストールしている場合にだけ、[GLM – 多変量] と [GLM – 反復測定] を使用できます。

1 つの因子についてのプロファイル プロットから、推定周辺平均が水準全体で増えているか減っているかがわかります。2 つ以上の因子では、因子間で交互作用がないと平行線で示されるので、1 つの因子に限ってそのレベルを調べることができます。平行ではない線は、交互作用を示しています。

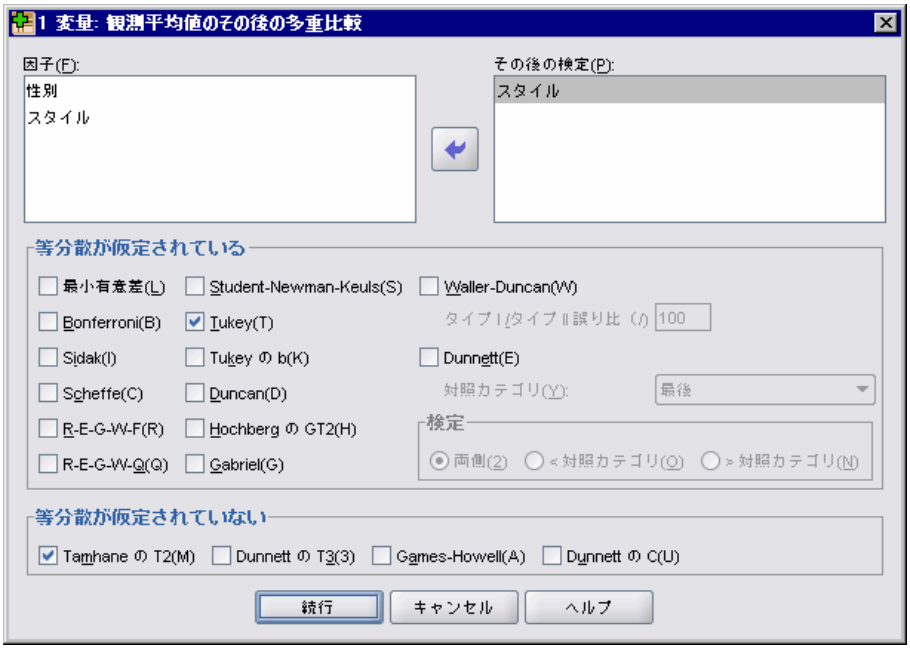
図 22-5
平行ではないプロット (左) と平行プロット (右)



水平軸の因子を選択してプロットを指定し、オプションとして線の定義変数と層の定義変数を選択したら、そのプロットは [作図] ボックスの一覧に追加しなければなりません。

GLM のその後の比較

図 22-6
[その後の検定] ダイアログ ボックス



その後の多重比較検定。平均値の間に差があることが判明した後は、その後の範囲検定とペアごとの多重比較により、どの平均値が相異しているのかを決めることができます。比較は調整済みでない値に基づいて行われます。このような検定は、固定被験者間因子の場合にだけ使用します。GLM – 反復測定では、被験者間因子がなければこの検定は利用できません。また、その後の多重比較検定は、被験者内因子の水準全体の平均に対して実行します。GLM– 多変量では、その後の検定は、それぞれの測定ごとに従属変数の被験者内での平均に対して実行します。Advanced Models オプションをインストールしている場合にだけ、[GLM – 多変量] と [GLM – 反復測定] を使用できます。

多重比較検定では、通常、Bonferroni 検定と Tukey の HSD 検定を使用します。**Bonferroni の検定**では、スチューデントの *t* 検定統計量に基づいて、多重比較が行われるという事実に対して有意水準が調整されます。**Sidak の t 検定**でも有意水準が調整され、Bonferroni の方法よりも厳しく限定されます。**Tukey の HSD 検定**では、スチューデント化された範囲統計量を使ってすべてのペアごとの比較をグループ間で行い、実験ごとの誤差率をすべてのペアごとの比較の集合の誤差率に設定します。大量の平均値のペアを検定する場合は、Tukey の HSD 検定の方が Bonferroni 検定より有効です。少量のペアの場合は Bonferroni の方法の方が有効です。

Hochberg の GT2 検定は Tukey の HSD 検定と類似していますが、ここで使用するのはスチューデント化された最大法です。一般的には、Tukey の検定の方が有効です。**Gabriel のペアごとの比較検定**も、スチューデント化された最大法を使用しますが、一般的に、セルの大きさが均等ではない場合は、Hochberg の GT2 より有効です。セルの大きさのばらつきが大きい場合には、Gabriel の検定の方が公平になることがあります。

Dunnett のペアごとの多重比較 t 検定では、処理のグループを単一の対照平均値と比較します。最後のカテゴリは、デフォルトの対照カテゴリです。代わりに、最初のカテゴリを選択できます。両側または片側の検定を選択することもできます。因子の任意のレベルの平均値 (対照カテゴリを除く) が対照カテゴリのそれと等しくないことを検定するには、両側の検定を使います。因子の任意のレベルの平均値が対照カテゴリの平均値よりも小さいことを検定するには、[< **対照カテゴリ**] を選択します。同様に、因子の任意のレベルの平均値が対照カテゴリの平均値より大きいことを検定するには、[> **対照カテゴリ**] を選択します。

Ryan と Einot、Gabriel、Welsch (R-E-G-W) は、2 種類のステップダウン多重範囲検定を開発しました。ステップダウン多重手続きでは、最初にすべての平均値が等しいかどうかを検定します。すべての平均値が等しくなければ、サブグループの平均値が等しいかどうかを検定します。**R-E-G-W の F値** は *F* 検定に基づき、**R-E-G-W の Q値**はスチューデント化した範囲に基づきます。この検定は、Duncan の多重範囲検定や Student-Newman-Keuls の検定 (これもステップダウン多重手続き) より有効ですが、セルの大きさが等しくない場合はお勧めできません。

分散が等しくない場合は、**Tamhane の T2** (*t* 検定に基づくペアごとの控えめな比較)、**Dunnett の T3** (スチューデント化した最大偏差に基づくペアごとの比較検定)、**Games-Howell の ペアごとの比較検定** (公平な場合もある)、または **Dunnett の C** (スチューデント化した範囲に基づくペアごとの比較検定) を使用します。

Duncan の多重範囲検定、Student-Newman-Keuls (**S-N-K**) の方法、および **Tukey の b 検定** はグループ平均を順位付け、範囲の値を計算する範囲検定です。これらの検定は先に述べた検定ほど頻繁には使われていません。

Waller-Duncan の t 検定では、Bayesian のアプローチが使われています。この範囲検定では、サンプル サイズが等しくない場合にサンプル サイズの調和平均が使われます。

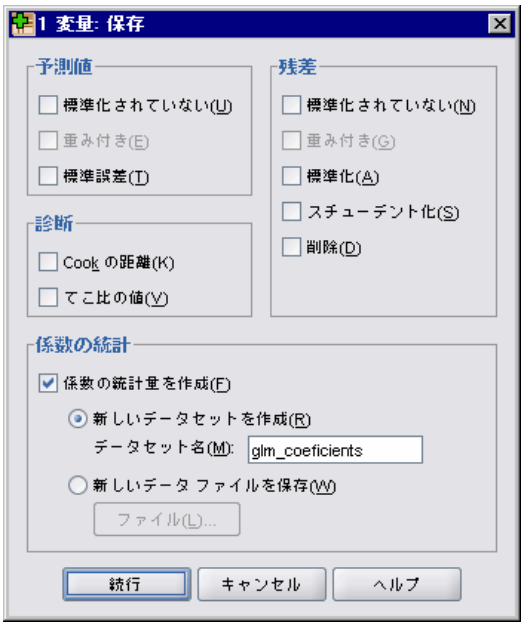
Scheffé の検定の有意水準は、この機能で利用できるペアごとの比較だけにとどまらず、グループ平均に可能なすべての線型結合を検定できる設計になっています。結果的に、Scheffé の検定は他の検定より控えめになってしまうことが多いため、有意確率を求める場合は、平均値間の差が大きい必要があります。

最小有意差 (**LSD**) のペアごとの多重比較検定は、グループのすべてのペア間ごとの多重 *t* 検定に相当します。この検定の欠点は、観測された有意水準を多重比較向けに調整する試みが一切行われないことです。

表示される検定。ペアごとの比較は、LSD、Sidak の方法、Bonferroni の方法、Games と Howell の方法、Tamhane の T2 と T3、Dunnett の *C* と Dunnett の T3 で使用できます。範囲検定の等質サブグループは、S-N-K、Tukey の *b*、Duncan、R-E-G-W の *F*、R-E-G-W の *Q* および Waller の方法で使用できます。Tukey の HSD 検定、Hochberg の GT2、Gabriel の検定、および Scheffé の検定は、多重比較検定と範囲検定の両方です。

GLM の保存

図 22-7
[保存] ダイアログ ボックス



モデルが予測した値、残差、関連測定値は、データ エディタで新変数として保存できます。これらの変数の多くは、データの仮定を調べるために利用できます。値を保存して別の SPSSセッションで利用するためには、現在のデータ ファイルを保存する必要があります。

予測値。モデルがケースごとに予測する値。

- **標準化されていない（判別分析）。**従属変数を予測するモデルの値。
- **重み付き。**重み付きの標準化されていない予測値。WLS 変数がすでに選択されている場合にのみ使用することができます。
- **標準誤差。**独立変数の同じ値を持つケースに対する従属変数の平均値の標準偏差の推定値。

診断。独立変数の一般的でない値の組み合せのケースとモデルに大きな影響を及ぼすケースを識別するための測定方法。

- **Cook の距離。**特定のケースが回帰係数の計算から除外された場合に、すべてのケースの残差がどのくらい変化するかを示す測定量。Cook の距離が大きいときは、回帰統計量の計算からケースを除外したことが係数を実質的に変化させたことを示しています。
- **てこ比の値。**非心てこ比の値。モデルの適合度に関する各観測の相対的な影響度。

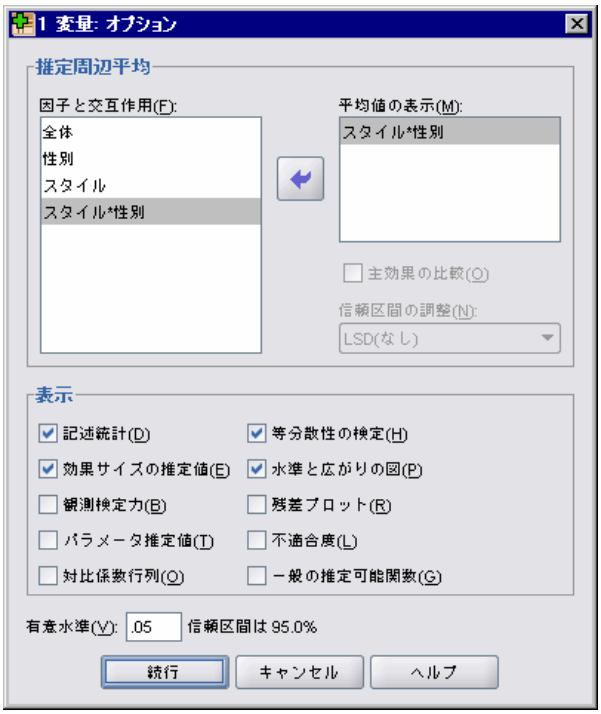
残差。標準化されていない残差は、従属変数の実際の値からモデルにより予測される値を引いたものです。標準化された残差、スチューデント化された残差、削除された残差も使用できます。WLS 変数を選択すると、重み付けした標準化されていない残差が使用できます。

- **標準化されていない（判別分析）。**観測値およびモデルで予測された値との差。
- **重み付き。**重み付きの標準化されていない残差。WLS 変数がすでに選択されている場合にのみ使用することができます。
- **標準化残差。**残差を標準偏差の推定値で割った値。標準化残差は Pearson 残差とも呼ばれ、平均は 0 で、標準偏差 1 になります。
- **スチューデント化された残差。**残差を、独立変数の平均値からの独立変数の各ケース値の距離に依存して、ケースごとに違う標準偏差の推定量で割ったもの。
- **除外（ケースの選択）。**あるケースが回帰係数の計算から除外されたときのケースの残差。従属変数と調整済み予測値の間の差です。

係数統計量。モデルにおけるパラメータ推定値の分散共分散行列を、現在のセッションにある新しいデータセット、または SPSS 形式の外部データ ファイルに出力します。また、それぞれの従属変数に対しても、パラメータ推定値の行、パラメータ推定値に対応する *t* 統計量の有意確率の行、および残差自由度の行が存在します。多変量モデルの場合は、各従属変数に同様の行があります。行列ファイルを読み込む別の手続きで、この行列ファイルを使用できます。

GLM のオプション

図 22-8
[オプション] ダイアログ ボックス



このダイアログ ボックスでオプションの統計を利用できます。統計量は、固定効果モデルを使用して計算されます。

推定周辺平均。セルにおける母周辺平均値について推定したい因子と交互作用を選択します。共変量が存在する場合、これらの平均値は、共変量に対して調整されます。

- **主効果の比較。**被験者間と被験者内因子の両方について、モデル内の主効果に対する推定周辺平均値間で、ペアごとに無補正の比較を行います。この項目は、[平均値の表示] リストで主効果を選択した場合にだけ選択できます。
- **信頼区間調整。**最小有意差 (LSD)、Bonferroni の方法、Sidak の方法の信頼区間と有意水準に対する調整を選択します。この項目は、[**主効果の比較**] を選択している場合にだけ選択できます。

表示。すべてのセルにおけるすべての従属変数の観測平均値、標準偏差、および度数を作るには、[**記述統計**] を選択します。[**効果サイズの推定値**] は、偏相関のイータの 2 乗の値をそれぞれの効果とそれぞれのパラメータ推定値に与えるものです。イータの 2 乗統計量は、因子に起因する総変動の比率を記述するものです。観測した値に基づいて対立仮説を立てるときに検定の検出力を得るには、[**観測検定力**] を選択します。パラメータ推定値、標準誤差、*t* 検定、信頼区間、および

それぞれの検定での観測検定力を創出するには、[**パラメータ推定値**] を選択します。L 行列を得るには、[**対比係数行列**] を選択します。

等分散性の検定は、被験者間因子の場合にだけ、各従属変数の分散の等質性分析について、被験者間因子のすべての水準の組み合せを通じて Levene の検定を創出するものです。水準と広がりの図および残差プロットのオプションは、データに関する仮定のチェックに役立ちます。この項目は、因子が存在しないときは無効になります。それぞれの従属変数について、標準化された予測による観測残差プロットを作成するには、[**残差プロット**] を選択します。このようなプロットは、分散が等しいという仮定の検証に役立ちます。[**不適合度**] を選択して、モデルにより従属変数と独立変数との関係を正確に記述できるかどうかをチェックします。**一般の推定可能関数** を使うと、一般推定可能関数に基づいてユーザー指定の仮説の検定を構築できます。任意の対比係数行列における行は、一般推定可能関数の一次結合です。

有意水準。信頼区間を構築するために、その後の検定や信頼係数で使われる有意水準を調整したい場合があります。指定された値は、検定の観測検定力の計算に使用できます。有意水準を指定するときは、信頼区間の関連水準がダイアログ ボックスに表示されます。

UNIANOVA コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 計画中の入れ子効果の指定 (DESIGN サブコマンドを使用)。
- 効果と効果や値の線型組み合せ検定の指定 (TEST サブコマンドを使用)。
- 多重対比の指定 (CONTRAST サブコマンドを使用)。
- ユーザー欠損値の包含 (MISSING サブコマンドを使用)。
- EPS 基準の指定 (CRITERIA サブコマンドを使用)。
- ユーザー定義の L 行列、M 行列、または K 行列の作成 (LMATRIX、MMATRIX、および KMATRIX サブコマンドを使用)。
- 全平均対比または単純対比での、中間参照カテゴリの指定 (CONTRAST サブコマンドを使用)。
- 多項式対比の計量の指定 (CONTRAST サブコマンドを使用)。
- その後の比較の誤差項の指定 (POSTHOC サブコマンドを使用)。
- 因子または因子の一覧にある因子間の交互作用の推定周辺平均値の計算 (EMMEANS サブコマンドを使用)。
- 仮の変数の名前の指定 (SAVE サブコマンドを使用)。
- 相関行列のデータ ファイルの作成 (OUTFILE サブコマンドを使用)。
- 被験者間因子の分散分析表の統計量を含む行列データ ファイルの作成 (OUTFILE サブコマンドを使用)。
- 計画行列の新しいデータ ファイルへの保存 (OUTFILE サブコマンドを使用)。

シンタックスの詳細は、『*Command Syntax Reference*』を参照してください。

章
23

2 変量の相関分析

有意水準のある Pearson の相関係数、Spearman のローおよび Kendall のタウ *b* を計算するには、[2 変量の相関分析] 手続きを使用します。変数またはランク順の関係を計算するには、相関分析を使用します。ただし、相関係数を計算する前に、2 変数間に外れ値 (これがあると誤った結果を出す可能性があります) および線型関係があるかどうかを確かめるためにグラフを描いてみてください。Pearson の相関係数は、線型連関の尺度になります。2 変数に強い関係があっても、その関係が直線的でない場合は、Pearson の相関係数は関連を測るのには適していません。

例。あるバスケットボール チームの勝数と 1 ゲームごとの平均得点の関係は相関しているでしょうか。散布図は直線関係があることを示しています。1994 年 ～ 1995 年の NBA シーズンにおけるデータ解析によると、Pearson の相関係数 (0.581) は 0.01 レベルで有意であることがわかります。シーズンあたりの勝数が多いほど、相手チームの得点が少ない (1 ゲームあたりの防御点が高い) ように見えます。この場合、変数は負 (–0.401) に相関し、その相関は 0.05 レベルで有意となります。

統計量。各変数に対して、非欠損値を持つケースの数、平均値および標準偏差。変数のペアごとに、有意水準のある Pearson の相関係数、Spearman のロー、Kendall のタウ *b*、交差積和、共分散。

データ。Pearson の相関係数には、正規分布している集団から抽出された量的変数を使用し、Spearman のローおよび Kendall のタウ *b* には、量的変数または順序付けされたカテゴリ変数を使用します。

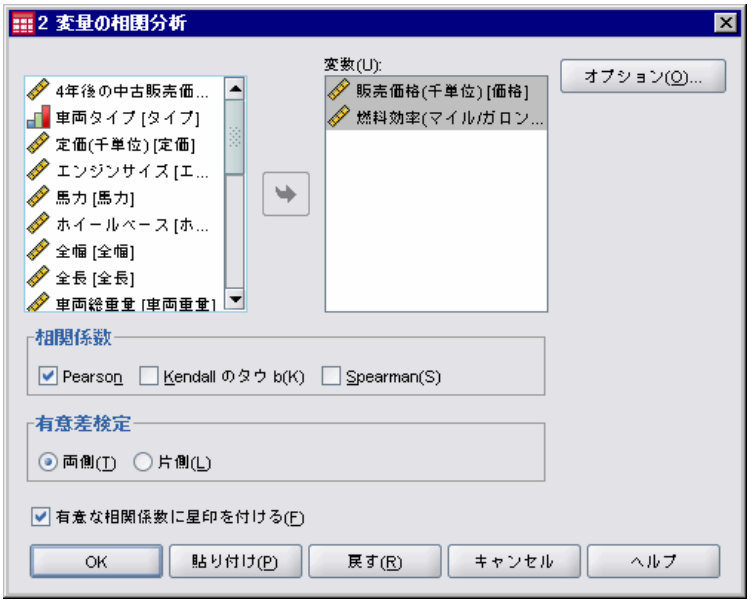
仮定。Pearson の相関係数は、それぞれの変数の組が 2 変量正規であると仮定します。

2 変量の相関分析を行うには

メニューから次の項目を選択します。

分析
相関
2 変量...

図 23-1
[2 変量の相関分析] ダイアログ ボックス



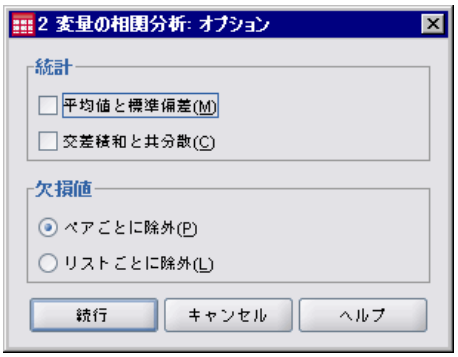
▶ 2 つ以上の数値型変数を選択します。

次のオプションも選択できます。

- **相関係数。**正規分布している集団から抽出された量的変数には、[Pearson] の相関係数を選択します。データが正規分布していないか、または順序付けされたカテゴリを持っている場合、[Kendall のタウ b] または [Spearman] を選択して、ランク順の間の連関を測ります。相関係数の範囲は、–1 (完全な負の相関) から +1 (完全な正の相関)までの値をとります。値 0 は線型関係がないことを示します。結果を解釈するとき、相関が有意であるからといって、因果関係があるという結論を出さないように注意する必要があります。
- **有意差検定。**両側確率または片側確率を選択できます。連関の方向がわかっている場合は、[**片側**] を選択します。不明の場合は、[**両側**] を選択します。
- **有意な相関係数に星印を付ける。**相関係数が 5% 水準で有意な場合は星印が 1 つ、1% 水準で有意な場合は星印が 2 つ付いた形で表示されます。

2 変量の相関分析のオプション

図 23-2
[2 変量の相関分析: オプション] ダイアログ ボックス



統計量。Pearson の相関分析については、次のどちらか一方または両方を選択できます。

- **平均値と標準偏差。**変数ごとに表示されます。また欠損値を除いた全ケース数も表示されます。欠損値の除外の仕方に関係なく、欠損値は変数単位で扱われます。
- **交差積和と共分散。**変数のペアごとに表示されます。交差積和は、平均値を修正した変数の積の和に等しくなります。これが、Pearson の相関係数の算出式の分子になります。共分散は 2 変数間の関係の標準化されていない尺度であり、交差積和を *N*–1 で割った値と等しくなります。

欠損値。次のオプションのうち 1 つを選択できます。

- **ペアごとに除外。**相関係数の変数の組み合せのうち、片方または両方が欠損値であるケースは分析から除外されます。各係数は、その組み合せの変数が両方とも欠損値でないすべてのケースから計算されるため、計算時には、利用できる最大の情報が使用されます。そのため、係数ごとに使用したケースの数が異なる場合があります。
- **リストごとに除外。**変数の欠損値のあるケースは、すべての相関から除外されます。

CORRELATIONS および NONPAR CORR コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- Pearson の相関係数では、相関行列を書き出して、生データの代わりに因子分析などの他の分析を実行 (MATRIX サブコマンドを使用)。
- 2 つの変数リスト間の各変数の相関を得る (VARIABLES サブコマンドのキーワード WITH を使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
24

偏相関

偏相関手続きは、1 つ以上の追加変数の効果をコントロールしながら 2 つの変数間の線形関係を表す偏相関係数を計算します。相関は、線形連関の測度です。2 つの変数は完全に関連していますが、その関係が線形ではない場合、相関係数は連関を測定するための適切な統計ではありません。

例。医療用資金と罹患率との間に関係はあるでしょうか?このような関係は負であるかのように思われるかもしれませんが、医療用資金が増えるのに応じて罹患率が上がるように見えるという**正の**相関をある学生が報告しています。しかし、医療サービス機関への訪問率を操作すると、観測された正の相関は実質的に排除されます。医療用資金と罹患率は、単に正の関係を持つように見えるだけです。なぜなら、資金が増えるほど、人々は医療を受けることができ、医者や病院からより多くの病気が報告されることになるからです。

統計量。各変数に対して、非欠損値を持つケースの数、平均値および標準偏差。自由度と有意確率を伴った偏相関行列および 0 次相関行列。

データ。対称的な量的変数を使用します。

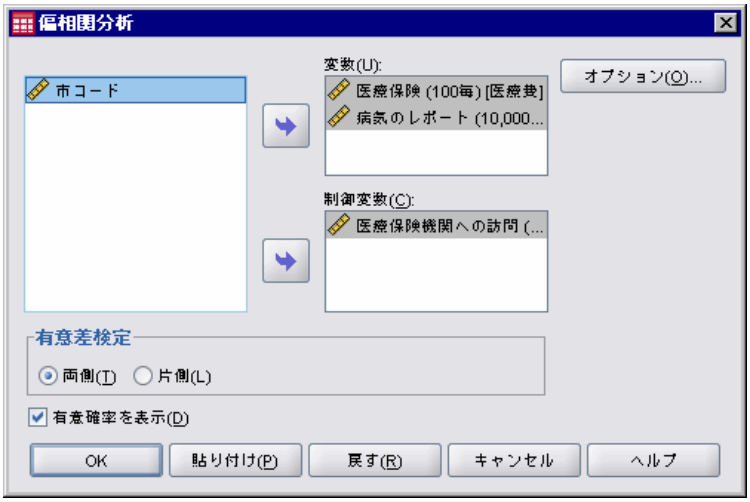
仮定。偏相関手続きは、各ペアの変数は 2 変量正規であると仮定しています。

偏相関を行うには

▶ メニューから次の項目を選択します。

分析
相関
偏相関...

図 24-1
[偏相関分析] ダイアログ ボックス



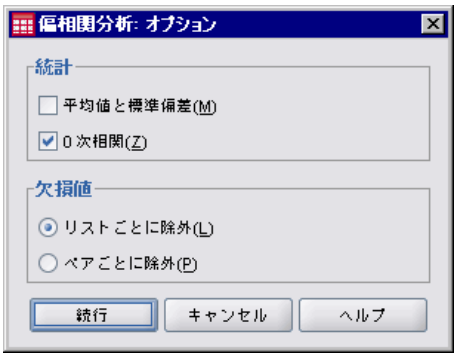
- 偏相関を計算する 2 つ以上の数値変数を選択します。
- 1 つ以上の数値型制御変数を選択します。

次のオプションも選択できます。

- **有意差検定。**両側確率または片側確率を選択できます。連関の方向がわかっている場合は、[**片側**] を選択します。不明の場合は、[**両側**] を選択します。
- **有意確率を表示。**デフォルトでは、確率と自由度は各相関係数に表示されます。この項目の選択を解除すると、5% 水準で有意な係数は 1 つの星印で識別され、1% 水準で有意な係数は 2 つの星印で識別され、自由度は抑制されます。この設定は、偏相関と 0 次相関両方に影響します。

偏相関のオプション

図 24-2
[偏相関分析: オプション] ダイアログ ボックス



統計量。次の中から 1 つまたは両方を選択することができます。

- **平均値と標準偏差。**変数ごとに表示されます。また欠損値を除いた全ケース数も表示されます。
- **0 次相関。**制御変数を含むすべての変数間の単純相関の行列が表示されます。

欠損値。次のオプションのどちらかを選択できます。

- **リストごとに除外。**制御変数を含め、変数に欠損値のあるケースはすべての計算から除外されます。
- **ペアごとに除外。**偏相関が基になっている 0 次相関の計算には、ペアになった変数の両方または 1 つに欠損値があるケースは使用されません。ペアごとの削除では、可能な限り多くのデータが使用されます。ただし、ケースの数は係数全体で異なります。ペアごとの削除が有効なとき、特定の部分係数の自由度は、すべての 0 次相関の計算で使用されるケースの最小数に基づいています。

PARTIAL CORR コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 0 次相関行列の読み込み、または偏相関行列の書き出し (MATRIX サブコマンド使用)。
- 2 つの変数リスト間の偏相関の取得 (VARIABLES サブコマンドで WITH キーワード使用)。
- 複数分析の取得 (複数の VARIABLES サブコマンド使用)。
- 制御変数が 2 つある場合の、要求する次数値の指定 (たとえば、1 次偏相関と 2 次偏相関) (VARIABLES サブコマンド使用)。
- 冗長な係数の非表示 (FORMAT サブコマンド使用)。
- 計算できない係数がある場合における、単純相関行列の表示 (STATISTICS サブコマンド使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
25

距離

この手続きでは、変数のペアまたはケースのペア間の類似度または非類似度（距離）を測定するさまざまな統計量を計算することができます。これらの類似度または距離は、因子分析、クラスタ分析、または多次元尺度法などの他の手続きで使用でき、複雑なデータセットの分析に役立ちます。

例。エンジンのサイズ、ガソリン 1 ガロン当たりの走行距離、馬力などの任意の特性変数をもとに、自動車のペア間の類似度を測定することは可能でしょうか。自動車の類似度を計算することによって、自動車が互いに類似しているか、類似していないかを知ることができます。より正式な分析を行うには、基本的構造を調べる階層クラスタ分析または多次元尺度法を使用することもできます。

統計量。非類似度（距離）の測度は、区間データの場合には、ユークリッド距離、平方ユークリッド距離、Chebychev、都市ブロック、Minkowski、またはカスタマイズ、度数データの場合には、カイ 2 乗またはファイ 2 乗、2 値データの場合には、ユークリッド距離、平方ユークリッド距離、サイズの差異、パターンの差異、分散、形、または Lance と Williams の距離です。間隔データの類似度の測度は、Pearson の相関係数またはコサインです。また 2 値データの類似度の測度は、Russel と Rao、単純マッチング、Jaccard、Dice、Rogers と Tanimoto、Sokal と Sneath 1、Sokal と Sneath 2、Sokal と Sneath 3、Kulczynski 1、Kulczynski 2、Sokal と Sneath 4、Hamann、ラムダ、Anderberg の *D*、Yule の *Y*、Yule の *Q*、落合、Sokal と Sneath 5、4 分点相関ファイ、または散らばりです。

距離行列を取得するには

▶ メニューから次の項目を選択します。

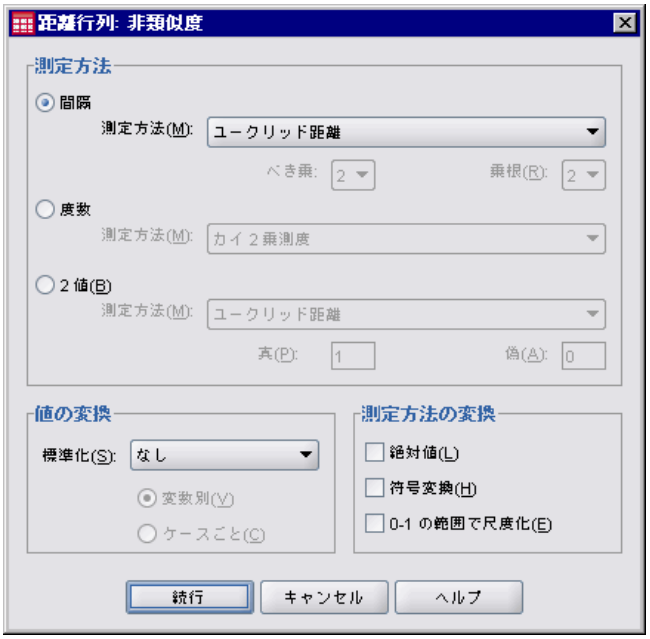
分析
相関
距離...

図 25-1
[距離行列] ダイアログ ボックス

- ケース間の距離を計算するには、[変数] ボックスに少なくとも 1 つの数値型変数を選択し、変数間の距離を計算するには、少なくとも 2 つの数値型変数を選択します。
- [距離の計算] グループの [ケース間] または [変数間] をクリックして、ケース間または変数間のどちらの距離を計算するか指定します。

距離行列の非類似度の測定方法

図 25-2
[距離行列: 非類似度] ダイアログ ボックス



[測定] グループでデータの種類 ([区間]、[度数]、または [2 値]) を選択します。次に、ドロップダウン リストからデータの種類に応じて測定方法を 1 つ選択します。データの種類別に使用できる測定方法は次のとおりです。

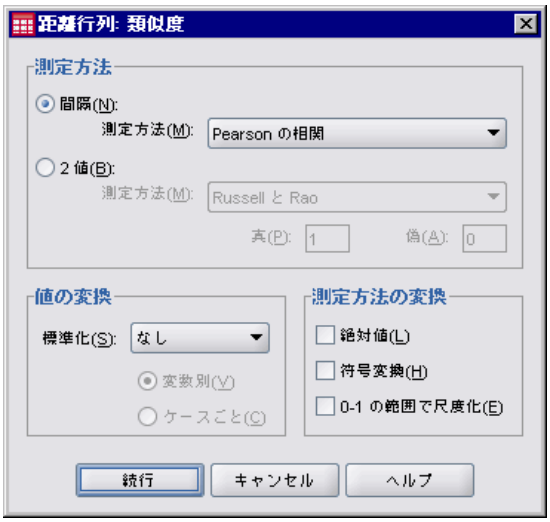
- **区間データ。**[ユークリッド距離]、[平方ユークリッド距離]、[Chebychev]、[都市ブロック]、[Minkowski]、または [カスタマイズ]。
- **度数データ。**[カイ 2 乗測度] または [ファイ 2 乗測度]。
- **2 値データ。**ユークリッド距離、平方ユークリッド距離、サイズの差異、パターンの差異、分散、形、Lance と Williams。([真] と [偽] の各ボックスに値を入力し、対象の 2 値を指定します。他の値は無視されます)。

[値の変換] グループでは、近接度を計算する**前**に、ケースまたは変数のデータの値を標準化することができます。2 値データの場合、[値の変換] は使用できません。使用できる標準化方法は、[*z* 得点]、[–1 から 1 の範囲]、[0 から 1 の範囲]、[最大値を 1]、[平均値を 1]、または [標準偏差を 1] です。

[測定方法の変換] グループでは、距離の測定方法によって生成された値を変換することができます。変換した値は、距離を計算した後で適用されます。使用できるオプションは、[絶対値]、[符号変換]、[0 ～ 1 の範囲で尺度化] です。

距離行列の類似度の測定方法

図 25-3
[距離行列: 類似度] ダイアログ ボックス



[測定] グループでデータの種類 ([区間] または [2 値]) を選択します。次に、ドロップダウン リストからデータの種類に応じて測定方法を 1 つ選択します。データの種類別に使用できる測定方法は次のとおりです。

- **区間データ。**Pearson の相関係数またはコサイン。
- **2 値データ。**Russell と Rao、単純マッチング、Jaccard、Dice、Rogers と Tanimoto、Sokal と Sneath 1、Sokal と Sneath 2、Sokal と Sneath 3、Kulczynski 1、Kulczynski 2、Sokal と Sneath 4、Hamann、ラムダ、Anderberg の *D*、Yule の *Y*、Yule の *Q*、落合、Sokal と Sneath 5、4 分点相関ファイ、散らばり。([真] と [偽] の各ボックスに値を入力し、対象の 2 値を指定します。他の値は無視されます)。

[値の変換] では、近接度を計算する前に、ケースまたは変数のデータの値を標準化することができます。2 値データの場合、[値の変換] は使用できません。使用できる標準化方法は、[*z* 得点]、[-1 から 1 の範囲]、[0 から 1 の範囲]、[最大値を 1]、[平均値を 1]、または [標準偏差を 1] です。

[測定方法の変換] グループでは、距離の測定方法によって生成された値を変換することができます。変換した値は、距離を計算した後で適用されます。使用できるオプションは、[絶対値]、[符号変換]、[0 ～ 1 の範囲で尺度化] です。

PROXIMITIESL コマンドの追加機能

距離行列手続きでは、PROXIMITIES コマンド シンタックスが使用されます。コマンド シンタックスを使用すると、次の作業も実行できます。

- Minkowski の距離のべき乗として任意の整数を指定。
- ユーザー指定の距離のべき乗および根として任意の整数を指定。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
26

線型回帰

線型回帰では、線型方程式の係数を推定します。線型方程式には 1 つ以上の独立変数があり、従属変数の値を最適に予測します。たとえば、営業部員の年間総売上高（従属変数）を、年齢、教育、経験年数などの独立変数から予測できます。

例。バスケットボール チームの 1 シーズン中の勝利ゲーム数は、ゲームごとのチーム得点の平均に関係しているでしょうか。散布図は、これらの変数には線型関係があることを示しています。勝利ゲームと相手チームの平均得点にも線型関係があります。これらの変数間には負の相関があります。つまり、勝利ゲームが増加すると、相手チームの平均得点が減少します。線型回帰では、変数間の関係をモデリングできます。適当なモデルを使用すると、チームの勝利ゲーム数を予測できます。

統計量。各変数: 有効ケース数、平均値および標準偏差。各モデル: 回帰係数、相関行列、部分相関および偏相関、多重**回答**、R^2、調整済み R^2、R^2 の変化量、推定値の標準誤差、分散分析表、予測値、および残差。また、各回帰係数の 95% 信頼区間、分散共分散行列、変動インフレーション因子、許容度、Durbin-Watson の検定、距離（Mahalanobis、Cook、てこ比の値）、DfBeta、DfFit、予測区間、ケースごとの診断。プロット: 散布図、偏残差の散布図、ヒストグラム、正規確率プロット。

データ。従属変数および独立変数は量的でなければなりません。宗教、専攻、居住地区などのカテゴリ変数は、2 値（ダミー）変数またはその他の種類の対比変数として再割り当てする必要があります。

仮定。独立変数の各値に対して、従属変数の分布は正規分布でなければなりません。従属変数の分布の分散は、独立変数のすべての値に対して一定でなければなりません。従属変数と各独立変数の関係は線型であり、すべての観測が独立している必要があります。

線型回帰分析を行うには

▶ メニューから次の項目を選択します。

分析
回帰
線型...

図 26-1
[線型回帰] ダイアログ ボックス

▶ [線型回帰] ダイアログ ボックスの変数のリストから数値型の従属変数を選び、[従属変数] ボックスに移動します。

▶ 変数のリストから 1 つ以上の数値型の独立変数を選び、[独立変数] ボックスに移動します。

オプションとして、次の選択が可能です。

- 独立変数を幾つかのブロックに分け、変数の組み合せを変えて投入方法を指定する。
- ケース選択変数を選んで、その変数の特定の値を持つケースだけに限定して分析を行う。
- ケースを識別する変数を選んで、プロット上の点を識別する。
- [WLS 重み] に数値型変数を選択して、重み付き最小 2 乗法分析を行う。

WLS。重み付き最小 2 乗法モデルを利用できます。データ点には、その変数の逆数で重みが付けられます。これは、分散の大きな観測は小さな分散に関連する観測値より分析に与える影響が小さいことを意味しています。重み付け変数の値がゼロ、負、または欠損の場合、ケースは分析から除外されます。

線型回帰の変数選択方法

[方法] ボックスでの選択により、独立変数を分析に投入する方法を指定できます。変数は同じ組み合せのまま方法を変えることで、さまざまな種類の回帰モデルを求めることができます。

- **強制投入法 (回帰)。**変数選択のための手順で、ブロック内のすべての変数は、1 つのステップで投入されます。
- **ステップワイズ法。**各ステップにおいて、回帰式にない独立変数の F 値の有意確率が十分小さければ、この変数は入力されます。すでに回帰式に含まれている独立変数でも、F 値確率が十分に大きくなると除去されます。入力や除去の対象の変数がなくなると、この方法は終わります。
- **削除 (基本的なテーブル表 統計量)。**ブロックの中の全ての変数は、変数の選択法によって、ある 1 つのステップで、強制除去されます。
- **変数減少法。**すべての変数を等式に入力してから順番に除去していく変数選択の手順。従属変数と最も小さい偏相関を持つ変数が、最初に除去する対象となります。その変数が除去するための基準を満たす場合は除去されます。最初の変数が削除されると、等式内に残っている変数のうち、最も小さい偏相関を持つ変数が次の対象となります。等式内に除去基準を満たす変数がなくなると、手順きは終了します。
- **変数増加法。**変数を順番にモデルに入力していく、ステップごとの変数選択の手順。方程式に最初に入力する変数は、従属変数との正または負の相関が最も大きいものです。この変数は、入力基準を満たしている場合に限って方程式に入力されます。最初の変数が入力されると、次は方程式に含まれていない独立変数のうち、最も大きい偏相関を持つものが対象となります。入力基準を満たす変数がないと、この手続きは中止します。

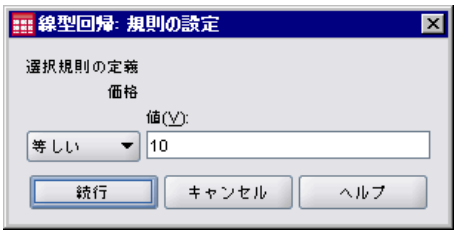
出力内の有意値は、1 つのモデルの適合に基づいたものです。したがって、ステップワイズの方法 (ステップワイズ法、変数増加法、変数減少法) を使用したときには、一般に、有意値は無効になります。

指定した投入方法に関係なく、変数を回帰式に投入するには、その変数が許容基準を満たしている必要があります。デフォルトの許容水準は 0.0001 です。また、ある変数を投入すると、すでにモデルに投入されている別の変数の許容度が許容基準より下がってしまう場合、その変数は投入されません。

選択されたすべての独立変数が、1 つの回帰モデルに追加されます。しかし、変数の組み合せを変えて、変数の投入方法を指定することもできます。たとえば、1 番目の変数のブロックをステップワイズ法で回帰モデルに投入し、2 番目のブロックを変数増加法で投入することもできます。2 番目のブロックを回帰モデルに追加するには、[次へ] をクリックします。

線型回帰の規則の設定

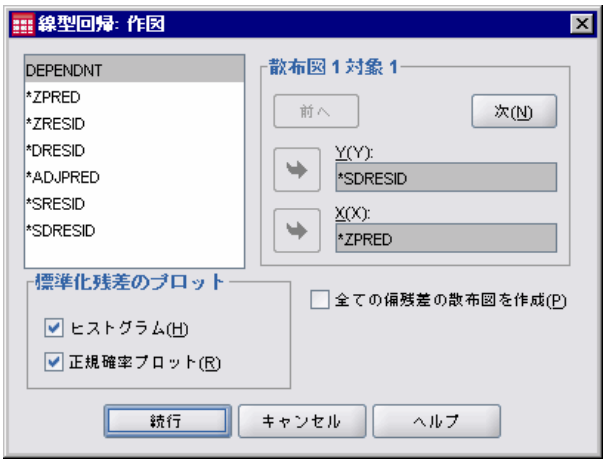
図 26-2
[線型回帰: 規則の設定] ダイアログ ボックス



選択規則により定義されたケースが分析に含まれます。たとえば、変数に対して [**等しい**] を選択し、[値] ボックスに **5** を入力すると、選択された変数の値に 5 を持つケースだけを分析で使用します。文字列も指定できます。

線型回帰の作図

図 26-3
[線型回帰: 作図] ダイアログ ボックス



作図は、正規性、線型性、分散の等質性に対する仮定の妥当性を確認するのに役立ちます。また、外れ値、異常な観測値、影響力の大きいケースを発見する上でも有用です。予測値、残差、その他の診断を新変数として保存すると、それらの変数をデータ エディタで使用できるようになり、独立変数を使用して作図できます。次の作図が利用できます。

散布図。従属変数、標準化予測値、標準化残差、削除ケース残差、調整済み予測値、スチューデント化された残差、スチューデント化された削除ケース残差の中から 2 つを選んで作図できます。標準化予測値に対して標準化残差を作図すると、線型性と等分散性を確認できます。

ソース変数リスト。従属変数 (DEPENDNT) と、標準化予測値 (*ZPRED)、標準化残差 (*ZRESID)、削除残差ケース (*DRESID)、調整済み予測値 (*ADJPRED)、スチューデント化された残差 (*SRESID)、スチューデント化された削除ケース残差 (*SDRESID) のそれぞれの予測変数と残差変数をリストします。

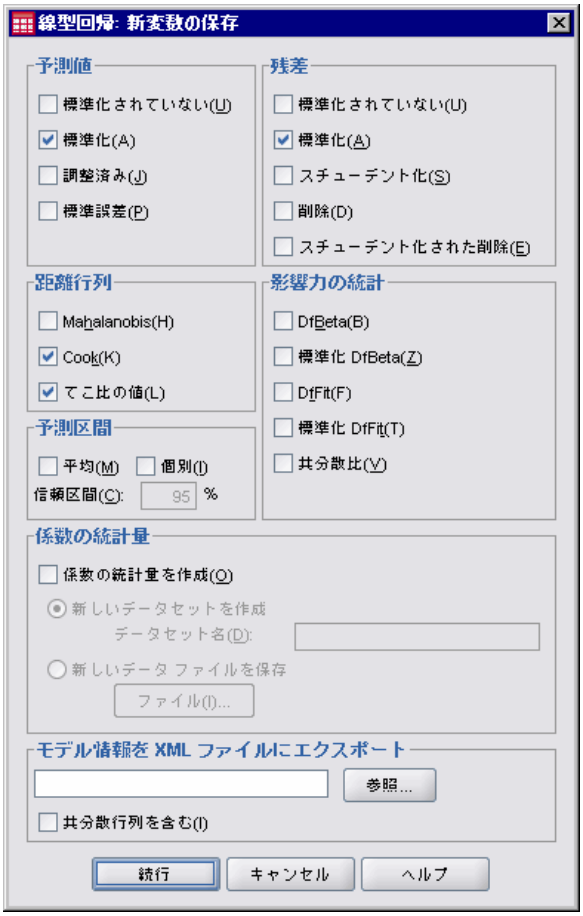
全ての偏残差の散布図を作成。独立変数と従属変数の両変数を残りの独立変数から離れて回帰させて、各独立変数の残差と従属変数の残差の散布図を表示します。偏残差プロットを作成するには、2 つ以上の独立変数をモデル内に投入する必要があります。

標準化残差のプロット。標準化残差の分布を正規分布と比較するために、ヒストグラムと正規確率プロットを作成できます。

どの作図を選んでも、標準化予測値と標準化残差 (*ZPRED と *ZRESID) の要約統計量が表示されます。

線型回帰の新変数の保存

図 26-4
[線型回帰: 新変数の保存] ダイアログ ボックス



予測値、残差、および診断に役立つその他の統計量を保存します。各統計量の選択により、1 つ以上の変数がアクティブなデータ ファイルに追加されます。

予測値。各ケースに対して回帰モデルが予測する値です。

- **標準化されていない (判別分析)。**従属変数を予測するモデルの値。
- **標準化残差。**各予測値を標準化された形式に変換したもの。すなわち、予測値から平均予測値を引き、その差を予測値の標準偏差で割ったものです。標準化予測値の平均は 0 で、標準偏差は 1 です。
- **調整済み。**回帰係数の計算に含まれないケースの予測値。
- **平均予測値の標準誤差。**予測値の標準誤差。独立変数の同じ値を持つケースに対する従属変数の平均値の標準偏差の推定値。

距離。回帰モデルに大きな影響をもたらす可能性がある、独立変数とケースの値の異常な組み合せを伴うケースを識別する測定です。

- **Mahalanobis。**独立変数のケースの値がケースの平均からどの程度離れているかを測るもの。大きい Mahalanobis の距離は、1 つ以上の独立変数に極値を持つケースを特定します。
- **Cook の統計量。**特定のケースが回帰係数の計算から除外された場合に、すべてのケースの残差がどのくらい変化するかを示す測定量。Cook の距離が大きいときは、回帰統計量の計算からケースを除外したことが係数を実質的に変化させたことを示しています。
- **てこ比の値。**回帰の適合性に対する 1 つの点の影響度を測定します。中心化てこ比の範囲は、0 (適合性への影響なし) から (N-1)/N までです。

予測区間。平均予測区間と個別予測区間の両方の上限と下限です。

- **平均。**平均予測応答の予測区間に対する下限と上限 (2 つの変数)。
- **個別 (散布図 オプション: 線の当てはめ)。**1 つのケースに対する従属変数の予測区間の下限と上限 (2 つの変数)。
- **信頼区間。**2 つの予測区間の値に対する信頼水準を指定するために、1 から 99.99 の値を入力します。この値を入力する前に、平均値または個別値が選択されている必要があります。一般的な信頼区間値は 90、95、および 99 です。

残差。従属変数の実際の値から回帰式で予測された値を引いた値です。

- **標準化されていない (判別分析)。**観測値およびモデルで予測された値との差。
- **標準化残差。**残差を標準偏差の推定値で割った値。標準化残差は Pearson 残差とも呼ばれ、平均は 0 で、標準偏差 1 になります。
- **スチューデント化された残差。**残差を、独立変数の平均値からの独立変数の各ケース値の距離に依存して、ケースごとに違う標準偏差の推定量で割ったもの。
- **除外 (ケースの選択)。**あるケースが回帰係数の計算から除外されたときのケースの残差。従属変数と調整済み予測値の間の差です。
- **スチューデント化された削除ケース残差。**ケースの削除済み残差をその標準誤差で割ったもの。スチューデント化された削除済み残差とその関連するスチューデント化された残差間の差は、それ自身の予測によって、ケースの削除によりどの程度の差が生じるかを示します。

影響力の統計。特定のケースを除外した場合の回帰係数の変化量 (DfBeta) と予測値の変化量 (DfFit) を使用できます。標準化 DfBeta 値および標準化 DfFit 値も共分散比とともに使用できます。

- **DfBeta(s)。**ベータ値の差は、特定のケースの除外から発生する回帰係数の変化です。値は、モデル内の各項に対して定数項を含めて計算されます。

- **標準化 DfBeta。**ベータ値の標準化された差。特定のケースの除去によって発生する回帰係数の変化。絶対値が 2 より大きいケースを N の平方根で割って調べることができます (N はケースの数)。値は、モデル内の各項に対して定数項を含めて計算されます。
- **DfFit。**当てはめ値の差は、特定ケースの除外から発生する予測値の変化です。
- **標準化 DfFit。**当てはめ値の標準化された差。特定ケースの除去によって発生する予測値の変化。2 を超える絶対値に p/N の平方根を掛けた値である標準化値を調べることができます (p はモデル内のパラメータの数、N はケースの数です)。
- **共分散比。**すべてのケースの分散共分散行列の行列式に対する、回帰係数の計算からあるケースが除外された場合の分散共分散行列の行列式の比。この比率が 1 に近い場合、そのケースは分散共分散行列に大きな影響を及ぼしていません。

係数統計量。回帰係数をデータセットやデータ ファイルに保存します。データセットは、同じセッションの今後で利用可能ですが、セッション終了前に明示的に保存しない限り、保存されません。データセット名は、変数命名規則に従う必要があります。 詳細は、85 ページの 5 章の変数名を参照してください。

モデル情報を XML ファイルにエクスポート。パラメータ推定とその共分散 (オプション) は指定されたファイルに XML (PMML) 形式でエクスポートされます。SmartScore、および SPSS (別製品) では、保存したモデル ファイルを使用して、得点付けのための他のデータ ファイルにモデル情報を適用できます。

線型回帰の統計

図 26-5
[統計] ダイアログ ボックス

次の統計を使用できます。

回帰係数。[**推定値**] は、回帰係数 *B*、*B* の標準誤差、標準化係数ベータ、*B* の *t* 値、および *t* の両側有意確率を表示します。[**信頼区間**] は、各回帰係数の 95% 信頼区間または分散共分散行列を表示します。[**分散共分散行列**] では、回帰係数の分散共分散行列を、共分散は対角線外に、分散は対角線上に表示します。相関行列も表示します。

モデルの適合度。モデルに投入された変数と除去された変数が一覧で表示され、次の適合度統計量、つまり多重**回答**、R^2、調整済み R^2、推定値の標準誤差、分散分析表を表示します。

R 2 乗の変化量。独立変数を加えたり取り除いたりすることによって生じる R^2 統計量の変化。ある独立変数についてこの R^2 統計量の変化が大きければ、その独立変数は従属変数をよく説明する予測変数であると見なすことができます。

記述統計。分析での各変数に対する有効ケース数、平均値、標準偏差を表示します。また、各相関係数に対する片側有意確率とケース数とともに、相関行列が表示されます。

偏相関。2 つの変数の間から他の変数の関係を取り除いた後で、それら 2 つの変数間に残った相関。従属変数と独立変数の両方からモデル内の他の独立変数の線型効果が取り除かれた後の、独立変数と従属変数の間の相関。

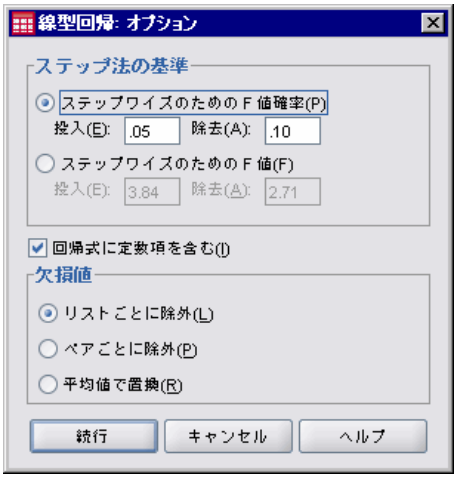
部分相関 (ピボットテーブル 回帰)。独立変数からモデル内の他の独立変数の線型効果が取り除かれた後の、独立変数と従属変数の間の相関。変数が方程式に追加されるときに 2 乗された R の変化と関連します。セミパーシャル相関とも呼ばれます。

共線性の診断。共線性 (または多重共線性) は、独立変数の 1 つが他の独立変数の線型関数であることを示す、望ましくない状況です。尺度化および非中心化された積和行列の固有値、条件指標、分散分解の比率が、変動インフレーション因子 (VIF) と個々の変数の許容度とともに表示されます。

残差。残差の系列相関に対する Durbin-Watson の検定と、選択基準に合うケースに対するケースごとの診断 (標準偏差 *n* 倍以上の外れ値) を表示します。

線型回帰のオプション

図 26-6
[線型回帰: オプション] ダイアログ ボックス



次のオプションを使用できます。

ステップ法の基準。このオプションは、変数選択の方法として変数増加法、変数減少法、またはステップワイズ法のいずれかが指定されている場合に適用できます。*F* 値 の有意確率または *F* 値自身のいずれかを、モデルへの変数の投入や除去に使用します。

- **ステップワイズのための F 値確率。**F 値の有意確率が [投入] の値よりも小さい場合、変数はモデルに入力され、[除去] の値よりも大きい場合除去されます。[投入] の値は [除去] の値より小さい値である必要があり、さらに両方の値は正である必要があります。さらに多くの変数をモデルに投入するには、[投入] 値を上げてください。さらに多くの変数をモデルから除去するには、[除去] 値を下げてください。
- **ステップワイズのための F 値。**変数は、その F 値が [投入] 値よりも大きい場合にモデルに投入され、[除去] よりも小さい場合に除去されます。[投入] の値は [除去] の値より大きい値である必要があり、さらに両方の値は正である必要があります。さらに多くの変数をモデルに投入するには、[投入] 値を下げてください。さらに多くの変数をモデルから除去するには、[除去] 値を上げてください。

回帰式に定数項を含む。デフォルトでは、回帰モデルに定数項が含まれます。このチェック ボックスをオフにすると、通常は使用しない原点を通る回帰になります。原点を通るようにした回帰と、定数を含むようにした回帰の結果は異なります。たとえば、R^2 を通常の方法で解釈できなくなります。

欠損値。次のオプションのうち 1 つを選択できます。

- **リストごとに除外。**分析で使うすべての変数が有効な値であるケースだけを使用します。

- **ペアごとに除外。**相関している変数のペアが両方とも完全なデータであるケースを使用して、回帰分析の基礎となる相関係数を計算します。自由度は、ペア単位の最小数 *N* が基礎になります。
- **平均値で置換。**欠損観測値を変数の平均値で置き換えて、すべてのケースを計算に使用します。

REGRESSION コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 相関行列の書き出しや、生データの代替として行列を読み込むことによる回帰分析の取得 (MATRIX サブコマンド使用)。
- 許容度の指定 (CRITERIA サブコマンド使用)。
- 同じ従属変数や、異なる従属変数に対する複数モデルの取得 (METHOD と DEPENDENT サブコマンド使用)。
- 追加の統計量の取得 (DESCRIPTIVES と STATISTICS サブコマンド使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
27

順序回帰

順序回帰分析では、一連の予測変数における、多分割順序応答の従属性をモデリングできます。予測変数は因子または共変量となります。順序回帰分析のデザインは、McCullagh (1980, 1998) の方法に基づいており、シンタックスでは、この手続きのことを PLUM と呼んでいます。

標準線型回帰分析では、応答 (従属) 変数と、予測 (独立) 変数の重み付き組み合せとの差を 2 乗したものの合計を最小化します。推定された係数は、予測変数の変化が応答変数にどのように影響するかを表します。応答は、応答の水準の変化が応答の範囲全体にわたって等しいという意味で、数値であると仮定されます。たとえば、身長 150 cm の人と、身長 140 cm の人の身長差は 10 cm です。これは、身長 210 cm の人と身長 200 cm の人の身長差と意味が同じです。このような関係は、順序変数に対しては必ずしも成り立たず、応答カテゴリの選択と数はきわめて恣意的です。

例。順序回帰分析を使用して、薬品投与に対する患者の反応を調べることができます。反応は、「**なし**」、「**穏やか**」、「**適度**」、「**激しい**」に分類できます。「穏やか」と「適度」の反応の違いは、数値化が困難、また不可能なため、知覚に基づいて判別されます。さらに言えば、「穏やか」と「適度」の応答の差は、「適度」と「激しい」の応答の差よりも大きい場合もあり、または逆に小さい場合もあります。

統計量と作図。観測度数、期待度数、累積度数、度数と累積度数の Pearson 残差、観測確率と期待確率、共変量パターンによる各応答カテゴリの観測累積確率と期待累積確率、パラメータ推定値の漸近相関行列および漸近分散共分散行列、Pearson のカイ 2 乗と尤度比カイ 2 乗、適合度統計量、反復の記述、平行線の仮定の検定、パラメータ推定値、標準誤差、信頼区間、Cox と Snell の R2 乗 統計量、Negalkerke の R2 乗 統計量、McFadden の R2 乗 統計量 。

データ。従属変数は、順序変数であると仮定され、数値または文字型のどちらかです。順序は、従属変数の値を昇順で並べ替えることにより決定されます。最低値によって、最初のカテゴリが定義されます。因子変数は、カテゴリ型と仮定されます。共変量変数は数値型でなければなりません。複数の連続共変量があると、非常に大きなセル確率テーブルが容易に作成されてしまうことに注意してください。

仮定。使用できる応答変数は 1 つだけで、必ず指定する必要があります。さらに、独立変数間の値の異なるパターンそれぞれについて、応答は独立多項分布変数であると仮定されます。

関連手続き。名義ロジスティック回帰分析では、名義従属変数に対して似たモデルを使用します。

順序回帰の作成

▶ メニューから次の項目を選択します。

分析
回帰
順序...

図 27-1
[順序回帰] ダイアログ ボックス

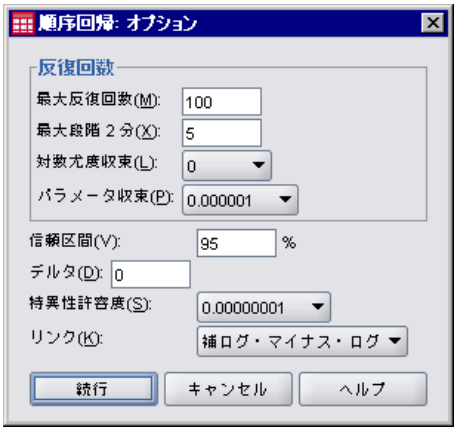
▶ 従属変数を 1 つ選択します。

▶ [OK] をクリックします。

順序回帰分析のオプション

[オプション] ダイアログ ボックスでは、反復推定アルゴリズムで使用されるパラメータの調整、パラメータ推定値の信頼水準の選択、リンク関数の選択を行うことができます。

図 27-2
[順序回帰: オプション] ダイアログ ボックス



反復回数。反復アルゴリズムをカスタマイズできます。

- **最大反復回数。**負でない整数を指定してください。0 を指定した場合、この手続きは初期推定値を返します。
- **最大段階 2 分。**正の整数を指定します。
- **対数尤度収束。**このアルゴリズムは、対数尤度の絶対変化または相対変化がこの値よりも小さければ停止します。0 を指定した場合、この収束基準は使用されません。
- **パラメータ収束。**このアルゴリズムは、それぞれのパラメータ推定値の絶対変化または相対変化がこの値よりも小さければ停止します。0 を指定した場合、この収束基準は使用されません。

信頼区間。0 以上、100 未満の値を指定します。

デルタ。0 のセル度数に加算される値。1 未満の負でない値を指定します。

特異性許容度。従属性の高い予測変数かどうかを調べる場合に使用します。オプションのリストから値を選択します。

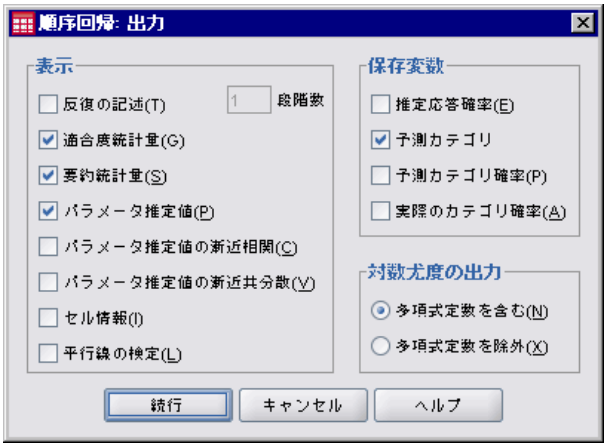
リンク関数。リンク関数とは、モデルの推定を行えるように累積確率を変換するためのものです。利用できるリンク関数は次の表に示す 5 種類です。

関数 (スクリプト ウィンドウ、新規手続き)	形式	代表的な適用例
ロジット	$\log(\xi / (1-\xi))$	分布が均一なカテゴリ
補ログ・マイナス・ログ	$\log(-\log(1-\xi))$	順序の高いカテゴリがより確率が高い
負ログ・マイナス・ログ	$-\log(-\log(\xi))$	順序の低いカテゴリがより確率が高い
プロビット	$\Phi^{-1}(\xi)$	潜在的変数が正規分布する
コーチット (コーシーの逆関数)	$\tan(\pi(\xi-0.5))$	潜在的変数に多数の外れ値が存在する

順序回帰分析の出力

[出力] ダイアログ ボックスでは、ビューアに表示するテーブルを作成し、変数を作業ファイルに保存できます。

図 27-3
[順序回帰: 出力] ダイアログ ボックス



表示。次のテーブルが作成されます。

- **反復の記述。**指定された出力反復頻度で、度数対数尤度およびパラメータ推定値を出力します。最初の反復と最後の反復は常に出力されます。
- **適合度統計量。**Pearson および尤度比カイ 2 乗統計量。変数リストで指定されている分類に基づいて計算されます。
- **要約統計量。**Cox と Snell の R2 乗 統計量、Nagelkerke の R2 乗 統計量、および McFadden の R2 乗 統計量 。
- **パラメータ推定値。**パラメータ推定値、標準誤差、および信頼区間。
- **パラメータ推定値の漸近相関。**パラメータ推定相関係数の行列。
- **パラメータ推定値の漸近共分散。**パラメータ推定共分散の行列。
- **セル情報。**観測度数、期待度数、累積度数、度数と累積度数に対する Pearson 残差、観測確率、期待確率、および共変量パターンによる各応答カテゴリの観測累積度数と期待累積度数。多数の共変量パターンを含むモデル (たとえば、連続共変量を含むモデル) では、このオプションにより、非常に大きく扱いにくいテーブルが生成される場合があることに注意してください。
- **平行線の検定。**位置パラメータが従属変数の水準にわたって等価であるという仮説の検定。これは、位置だけのモデルでしか使用できません。

保存変数。次の変数を作業ファイルに保存します。

- **推定応答確率。**因子/共変量パターンを応答カテゴリに分類するモデル推定確率。応答カテゴリの数だけ確率があります。
- **予測カテゴリ。**因子/共変量パターンに対して最大推定確率を持つ応答カテゴリ。

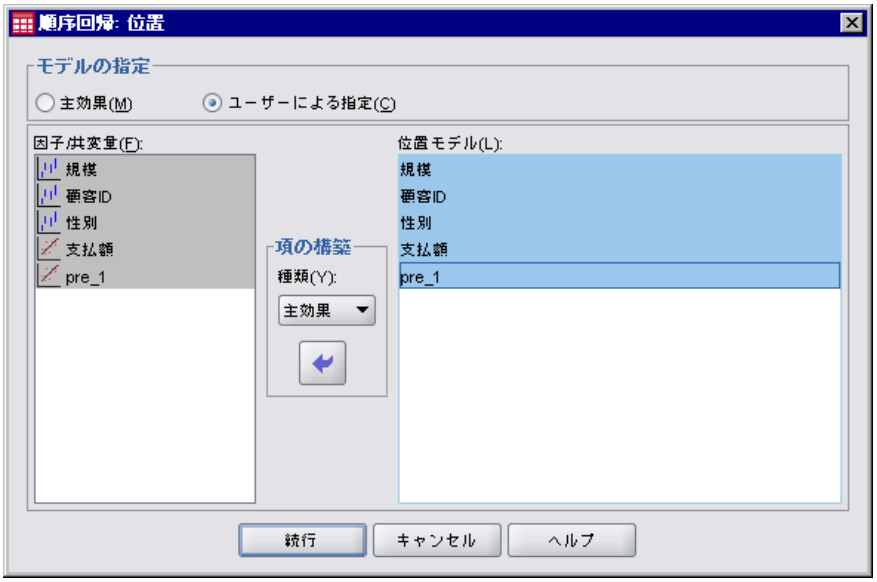
- **予測カテゴリ確率。**因子/共変量パターンを予測カテゴリに分類する推定確率。この確率は、因子/共変量パターンの推定確率の最大値でもあります。
- **実カテゴリ確率。**因子/共変量パターンを実際のカテゴリに分類する推定確率。

対数尤度を出力。対数尤度の表示を制御します。[**多項式定数を含む**] により、尤度の完全な値が求められます。その定数を含まない積について結果を比較するには、除外を選択します。

順序回帰分析の位置モデル

[位置] ダイアログ ボックスでは、分析する位置モデルを指定できます。

図 27-4
[順序回帰: 位置] ダイアログ ボックス



モデルの指定。主効果モデルには、共変量および因子主効果が含まれますが、交互作用効果は含まれません。ユーザーの指定によるモデルを作成して、因子交互作用または共変量交互作用のサブセットを指定できます。

因子/共変量。要素および共変量はリストされます。

位置モデル。モデルは、選択した主効果と交互作用効果によって異なります。

項の構築 (ロジット モデル)

選択した因子や共変量について、次の項を作成できます。

交互作用。選択した変数のすべてについて、最高水準の交互作用項を作成します。これはデフォルトです。

主効果。選択した変数のそれぞれに主効果の項を作成します。

2 次まで。選択した変数に 2 次までの交互作用を作成します。

3 次まで。選択した変数に 3 次までの交互作用を作成します。

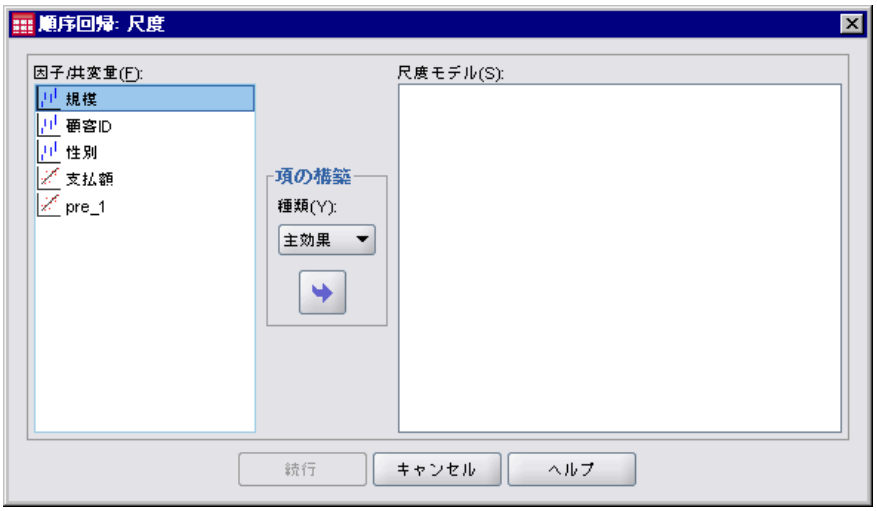
4 次まで。選択した変数に 4 次までの交互作用を作成します。

5 次まで。選択した変数に 5 次までの交互作用を作成します。

順序回帰分析の尺度モデル

[尺度] ダイアログ ボックスでは、分析する尺度モデルを指定できます。

図 27-5
[順序回帰: 尺度] ダイアログ ボックス



因子/共変量。要素および共変量はリストされます。

尺度モデル。モデルは、選択した主効果と交互作用効果によって異なります。

項の構築 (ロジット モデル)

選択した因子や共変量について、次の項を作成できます。

交互作用。選択した変数のすべてについて、最高水準の交互作用項を作成します。これはデフォルトです。

主効果。選択した変数のそれぞれに主効果の項を作成します。

2 次まで。選択した変数に 2 次までの交互作用を作成します。

3 次まで。選択した変数に 3 次までの交互作用を作成します。

4 次まで。選択した変数に 4 次までの交互作用を作成します。

5 次まで。選択した変数に 5 次までの交互作用を作成します。

PLUM コマンドの追加機能

選択内容をシンタックス ウィンドウに貼り付け、PLUM コマンド シンタックスを編集することで、順序回帰分析をカスタマイズできます。コマンド シンタックスを使用すると、次の作業も実行できます。

- 帰無仮説をパラメータの線型結合として指定することによって、カスタマイズした仮説の検定を作成。

シンタックスの詳細は、『*Command Syntax Reference*』を参照してください。

章
28

曲線推定

[曲線推定] 手続きは、曲線推定の回帰統計量と 11 種類の曲線推定の回帰モデルの関連プロットを作成します。個々のモデルは、従属変数ごとに作成されます。また、予測値、残差、および予測区間を新変数として保存することもできます。

例。あるインターネット サービス会社が、自社のネットワーク上を流れるウイルスに感染した電子メールの割合を長期にわたり追跡しています。散布図では、その関係が非線形であることが示されています。2 次モデルまたは 3 次モデルをデータに当てはめ、仮定の有効性とモデルの適合度を確認します。

統計量。各モデル: 回帰係数、多重**回答**、R^2、調整済み R^2、推定値の標準誤差、分散分析表、予測値、残差、および予測区間。モデル: 線型、対数、逆数、2 次、3 次、べき乗、複合成長、S 曲線、ロジスティック、成長、および指数。

データ。従属変数および独立変数は量的でなければなりません。独立変数としてアクティブなデータセットから (変数ではなく) **[時間]** を選択すると、[曲線推定] 手続きにより、ケース間の時間の長さが一定の時間変数が生成されます。**[時間]** を選択した場合は、従属変数が時系列でなければなりません。時系列分析に使用するデータ ファイルでは、各ケース (行) が個別の時間の一連の観測値を表し、さらにケース間の時間の長さが一定でなければなりません。

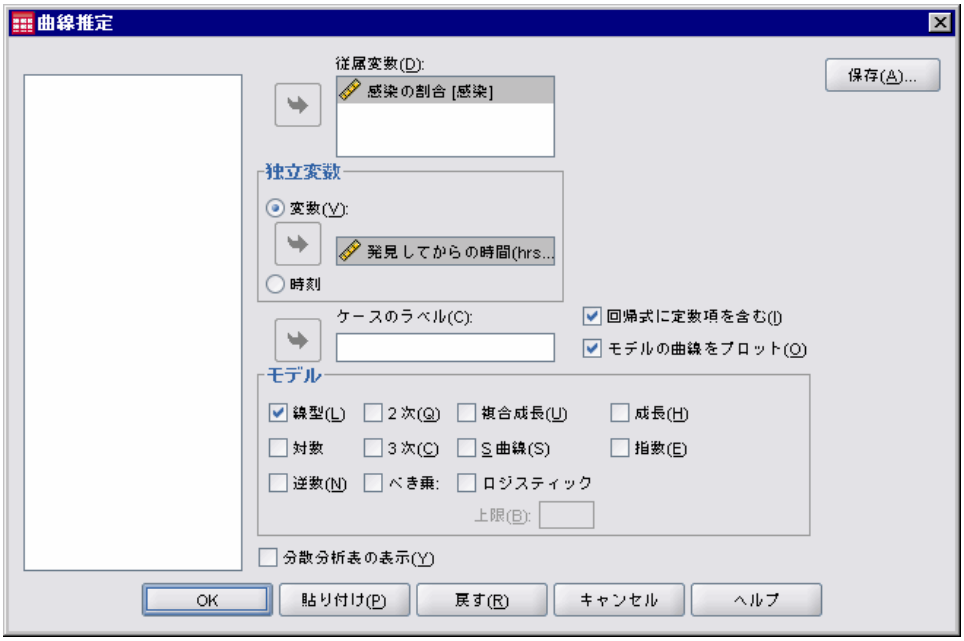
仮定。データをグラフ表示し、独立変数と従属変数の関係を判断します (線型、指数など)。理想的なモデルの残差は、ランダムに分布し、正規分布を示します。線型モデルを使用する場合は、次の 3 つの仮定を満たしている必要があります。すなわち、独立変数の各値に対して、従属変数の分布は正規分布でなければなりません。従属変数の分布の分散は、独立変数のすべての値に対して一定でなければなりません。従属変数と独立変数の関係は線型で、観測値はすべて独立でなければなりません。

曲線推定を行うには

▶ メニューから次の項目を選択します。

分析
回帰
曲線推定

図 28-1
[曲線推定] ダイアログ ボックス



- 従属変数を 1 つ以上選択します。個々のモデルは、従属変数ごとに作成されます。
- 独立変数として、アクティブなデータセットの変数を 1 つ選択するか、[時間] を選択します。
- 次のオプションが選択できます。
  - [ケースのラベル] ボックスに、散布図のケースにラベル付けする変数を 1 つ選択する。散布図で [点の識別] ツールを使用すると、各点に [ケースのラベル] ボックスの変数の値を表示できます。
  - [保存] をクリックして、予測値、残差、および予測区間を新変数として保存する。

次のオプションも選択できます。

- **回帰式に定数項を含む。**回帰式の定数項を推定します。デフォルトでは、このチェック ボックスはオンになっています。
- **モデルの曲線をプロット。**従属変数の値および選択された各モデルを独立変数に対して作図します。従属変数ごとに個別の図表が作成されます。
- **分散分析表の表示。**選択された各モデルに、要約された分散分析表を表示します。

曲線推定のモデル

1 つ以上の曲線推定の回帰モデルを選択できます。選択するモデルを決めるには、まずデータをプロットします。変数に線型関係があれば、単純な線型回帰モデルを使用します。線型関係がなければ、データを変換します。変換できない場合は、さらに複雑なモデルが必要となります。データの散布図を調べ、そのプロットがどれかの関数と似ていれば、データをそのモデルの型に当てはめます。たとえば、データが指数関数に似ている場合は、指数モデルを選択します。

線型 (ピボットテーブル GLM)。方程式が Y = b0 + (b1 * t) であるモデル。系列の値は時間の線型関数としてモデル化されます。

対数。次のような方程式をもつモデルです。Y = b0 + (b1 * ln(t))。

逆数。次のような方程式をもつモデルです。Y = b0 + (b1 / t)。

2 次 (ピボットテーブル GLM)。方程式が Y = b0 + (b1 * t) + (b2 * t**2) であるモデル。2 次モデルは、上昇する系列か下降する系列をモデル化するために使うことができます。

3 次 (ピボットテーブル GLM)。次のような方程式で定義されるようなモデルです。Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3)。

べき乗。次のような方程式をもつモデルです。Y = b0 * (t**b1) または ln(Y) = ln(b0) + (b1 * ln(t))。

複合成長。次のような方程式で表わされるモデルです。Y = b0 * (b1**t) または ln(Y) = ln(b0) + (ln(b1) * t)。

S 曲線。次のような方程式をもつモデルです。Y = e**(b0 + (b1/t)) または ln(Y) = b0 + (b1/t)。

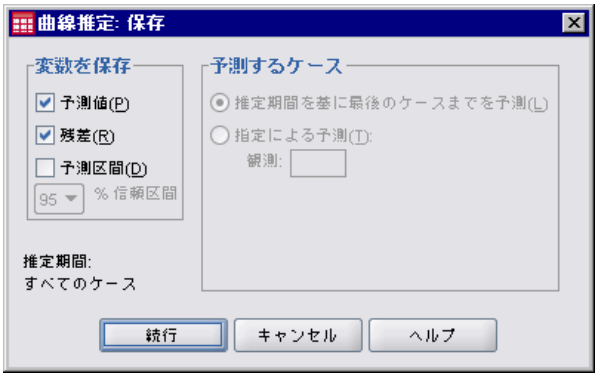
ロジスティック (正規確率プロット 検定分布)。次のような方程式をもつモデルです。Y = 1 / (1/u + (b0 * (b1**t))) または ln(1/y-1/u) = ln (b0) + (ln(b1) * t)。ここで u は、上限の値です。[ロジスティック] を選択したら、回帰式で使う上限の値を指定してください。この値は従属変数の最大値より大きい正の数である必要があります。

成長。次のような方程式をもつモデル。Y = e**(b0 + (b1 * t)) または ln(Y) = b0 + (b1 * t)。

指数。次のような方程式をもつモデルです。Y = b0 * (e**(b1 * t)) または ln(Y) = ln(b0) + (b1 * t)。

曲線推定の保存

図 28-2
[曲線推定: 保存] ダイアログ ボックス



変数を保存。選択されたモデルごとに予測値、残差 (従属変数の観測値からモデル予測値を引いた値)、および予測区間 (上限および下限) を保存できます。新しい変数名と記述的なラベルが出力ウィンドウのテーブルに表示されます。

予測するケース。アクティブなデータセット内で、独立変数として変数の代わりに [時間] を選択した場合は、時系列の最終点以降の予測期間を指定できます。次のオプションのどちらかを選択できます。

- **推定期間を基に最後のケースまでを予測。**推定期間内のケースに基づいてファイル内のすべてのケースに値を予測します。ダイアログ ボックスの下に表示される推定期間は、[データ] メニューの [ケースの選択] の [ケースの選択: 範囲の定義] ダイアログ ボックスで定義されたものです。推定期間が定義されていない場合は、すべてのケースを使用して値の予測が行われます。
- **指定による予測。**推定期間でのケースに基づいて指定した日付、時刻、または観測数までの値を推定します。この機能は、時系列の最後のケース以降の値を予測するために使うことができます。現在定義されている日付変数によって、推定期間の最終点を指定するボックスが異なります。日付変数が定義されていない場合は、[観測] ボックスで最終観測 (ケース) 数を指定します。

日付変数の作成には、[データ] メニューの [日付の定義] を使用します。

章
29

偏 2 段階最小 2 乗回帰

偏 2 段階最小 2 乗回帰手順は、偏 2 段階最小 2 乗 (PLS とは「潜在的構造投影方法 (projection to latent structure) のことです」) 回帰モデルを推定します。PLS は、通常最小 2 乗法 (OLS) 回帰、正準相関、構造方程式モデリングに対する代わりの予測技術であり、予測変数が相関しているとき、あるいは予測数がケース数を超えているときは特に役立ちます。

PLS は、主成分分析および多重回帰を組み合わせた機能です。最初に従属変数と独立変数の間の共分散をできるだけたくさん説明する潜在的要素を展開します。そして、回帰手順は独立変数の分解を使用して従属変数の値を予測します。

使用方法。PLS は拡張コマンドです。PLS を実行する予定のシステム上にインストールされるべき Python 拡張モジュール が必要です。PLS 拡張モジュールは別にインストールする必要があるため、インストーラーを *http://www.spss.com/devcentral* からダウンロードしてください。

注: PLS 拡張モジュールは、Python ソフトウェアとは関連がありません。SPSS は、Python ソフトウェアの所有者でもライセンス許諾者でもありません。Python のユーザーは、Python の Web サイトに掲載されている Python ライセンス契約の条件に同意する必要があります。SPSS は、Python プログラムの品質に関して一切言明しておりません。SPSS は、Python プログラムの使用に関して完全に免責されます。

テーブル。(潜在的因子により) 説明された分散の比率、潜在的因子の重み、因子負荷、投影の独立変数の重要度 (VIP)、回帰パラメータ推定値 (従属変数による) はすべてデフォルトにより作成されます。

図表。投影の変数の重要度 (VIP)、因子得点、3 つの潜在的因子の最初の因子の重み、モデルへの距離はすべて [オプション] タブから作成されます。

測定レベル。従属変数および独立変数 (予測) は尺度は、名義、順序、またはスケールのいずれかです。適切な尺度がすべてのジェン数に割り当てられたということを前提にする手順もありますが、ソース変数リストの変数を右クリックしてコンテキスト メニューから尺度を選択することで、その変数の尺度を一時的に変更できます。カテゴリ (名義または順序) 変数は手順により等しく扱われます。

カテゴリ変数のコード化。手順は、その手順の間 c コードの 1 つを使用してカテゴリ独立変数を記録します。変数の中に、c カテゴリがあるなら、変数は c ベクトルとして保存され、最初のカテゴリは (1,0,...,0) で示され、次のカテゴリは (0,1,0,...,0)、というようになり最後のカテゴリ (0,0,...,0,1) と示されます。カテゴリ

従属変数は参照カテゴリに対応する指標を単に省略した形式のダミー コードを使用して表示されます。

度数による重み付け。重み変数は使用前にもっとも近い全数に丸められます。欠損重みまたは、重みが 0.5 以下のケースは分析に使用されません。

欠損値。ユーザーおよびシステム欠損値は無効として取り扱われます。

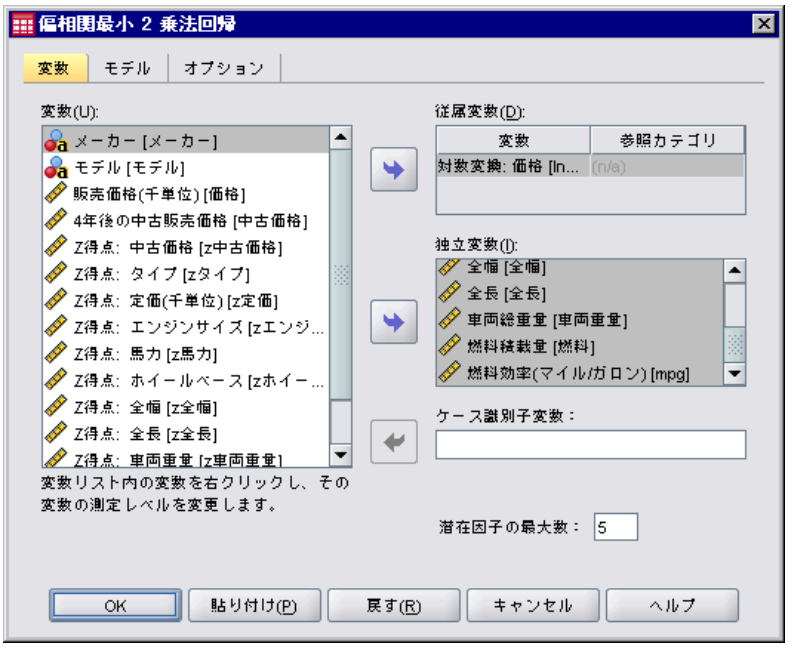
再調整。すべてのモデル変数は、カテゴリ変数を表す指示変数を含んで標準化および中心化されます。

偏 2 段階最小 2 乗回帰を取得するには次のようにします。

メニューから次の項目を選択します。

分析
回帰
偏 2 段階最小 2 乗...

図 29-1
[偏 2 段階最小 2 乗回帰変数] タブ



▶ 最低 1 つの従属変数を選択します。

▶ 最低 1 つの独立変数を選択します。

オプションとして、次の選択が可能です。

■ カテゴリ (名義、順位l) 従属変数の参照カテゴリを指定します。

- ケース主体の出力の一意の識別子として使用された、データセットを保存した変数を指定します。
- 表示される潜在的要素数の上限を指定します。

モデル

図 29-2
[偏 2 段階最小 2 乗回帰モデル] タブ

モデル効果を指定。主効果モデルには、共変量および因子主効果が含まれます。交互作用を指定するには、[**ユーザー指定**] を選択します。モデルに含める項目はすべて指示する必要があります。

因子と共変量。要素および共変量はリストされます。

モデル。モデルは、使用するデータの性質によって異なります。[**ユーザーの指定**] をクリックすれば、分析対象の主効果と交互作用を選択できます。

項の構築 (ロジット モデル)

選択した因子や共変量について、次の項を作成できます。

交互作用。選択した変数のすべてについて、最高水準の交互作用項を作成します。これはデフォルトです。

主効果。選択した変数のそれぞれに主効果の項を作成します。

2 次まで。選択した変数に 2 次までの交互作用を作成します。

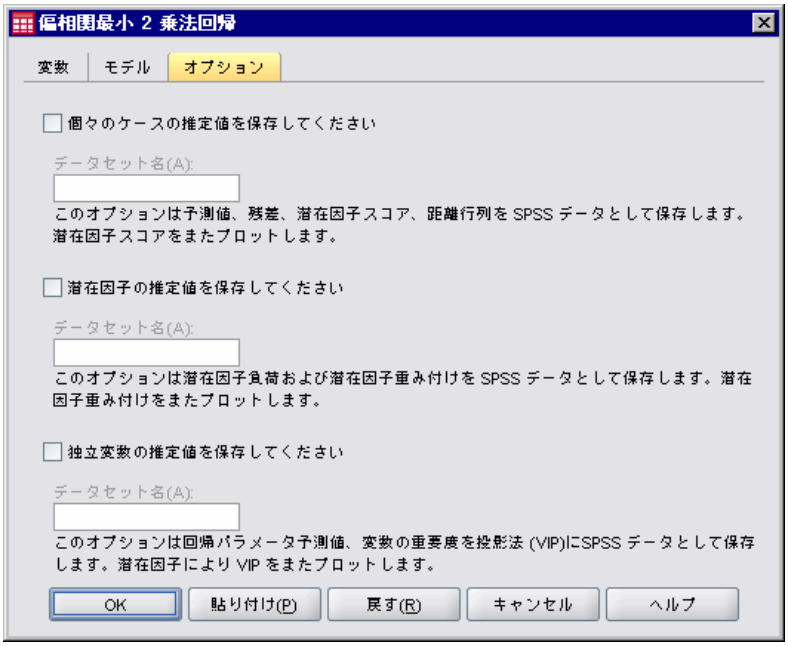
3 次まで。選択した変数に 3 次までの交互作用を作成します。

4 次まで。選択した変数に 4 次までの交互作用を作成します。

5 次まで。選択した変数に 5 次までの交互作用を作成します。

オプション

図 29-3
[偏 2 段階最小 2 乗回帰オプション] タブ



[オプション] タブを押すことにより、ユーザーが各ケース、潜在的因子、および予測変数を保存し、プロットできるようになります。

データの各型は、SPSS データセットの名前を指定します。データセット名は一意でなくてはいけません。既存のデータセットの名前を指定すると、内容は置き換えられ、新規データセットが作成されます。

- **各ケースの推定値を保存。**次のケースごとのモデル推定値、すなわち予測値、残差、潜在的要素モデルへの距離、潜在的因子得点を保存します。潜在的因子得点をプロットします。
- **潜在的因子の推定値を保存。**潜在的因子負荷と潜在的因子の重みを保存。潜在的因子の重みをプロットします。
- **独立変数の推定値を保存。**回帰パラメータ推定値と投影変数の重要度 (VIP) を保存。潜在的因子により VIP をプロットします。

章
30

判別分析

判別分析は所属グループの予測モデルを構築します。予測モデルは、グループ間で最良の判別を行う予測変数の線型結合をもとに、判別関数（または、複数のグループの場合、判別関数グループ）で構成されます。関数は所属グループがわかっているケースのサンプルから生成され、その関数は、所属グループがわからない予測変数の測定を含む新しいケースに適用することができます。

注: グループ化変数には 3 つ以上の値を設定できます。しかし、グループ化変数のコードは、整数でなければならず、さらに最小値と最大値を指定しなければなりません。この範囲外の値のケースは、分析から除外されます。

例。平均的に、温帯の国に住んでいる人は、熱帯に住んでいる人に比べて 1 日当たりのカロリー摂取が多く、さらに温帯の方が都市人口の比率が高くなっています。この 2 つのグループの国々をどの程度正確に判別できるかを確かめるため、研究者はこれらの情報を 1 つの関数にまとめたいと考えています。さらに、研究者は、人口や経済情報も重要であると考えています。判別分析を使って、多重線型回帰式の右側の項に似ている線型判別関数の係数を予測することができます。すなわち、係数 *a*、*b*、*c*、および *d* を使用した場合、式は次のようになります。

```
D = a * climate + b * urban + c * population + d
  * gross domestic product per capita
```

これらの変数が 2 つの気候地帯の判別に有効な場合、温帯の国と熱帯の国では、*D* の値が異なります。ステップワイズ変数選択法を使用すると、4 つの変数すべてを関数に取り入れる必要はないことがわかります。

統計量。各変数: 平均値、標準偏差、1 変量の分散分析。各分析: Box の *M*、グループ内相関行列、グループ内分散共分散行列、グループ別分散共分散行列、全分散共分散行列。各正準判別関数: 固有値、分散の割合、正準相関、Wilk のラムダ、カイ 2 乗。各ステップ: 事前確率、Fisher の分類関数の係数、標準化されていない分類関数の係数、各正準関数に対する Wilks のラムダ。

データ。グループ化変数には、整数でコード化された一定数のカテゴリがなければなりません。名義尺度の独立変数は、ダミー変数または対比変数として再割り当てしなければなりません。

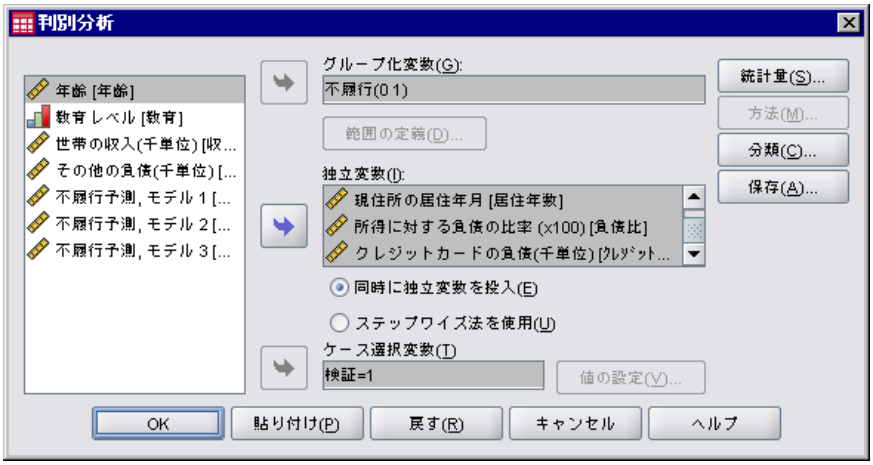
仮定。ケースは独立していなければなりません。予測変数は多変量正規分布を示していなければならず、さらにグループ内分散共分散行列はグループ全体で等しくなければなりません。各所属グループは重複せず (1 つのケースが複数のグループに所属していない)、グループ全体ですべてのケースを網羅している (全ケースがグループに所属している) と仮定します。この手続きは、所属グループがカテゴリ変数で構成されている場合に最も有効です。所属グループが連続変数の値 (たとえば、高い IQ 対低い IQ) に基づいて構成されている場合、連続変数そのものの豊富な情報を最大限に活用するために、線型回帰を使用するとよいでしょう。

判別分析を行うには

▶ メニューから次の項目を選択します。

分析
分類
判別分析...

図 30-1
[判別分析] ダイアログ ボックス



▶ 整数値をとるグループ化変数を選択し、[**範囲の定義**] をクリックして、目的のカテゴリを指定します。

▶ 独立変数つまり予測変数を選択します。(グループ化変数の値が整数でない場合、[変換] メニューの [値の再割り当て] で値が整数の変数を作成します)。

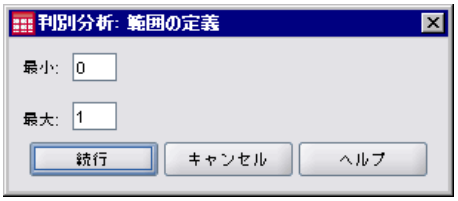
▶ 独立変数を投入する方法を選択します。

- **同時に独立変数を投入。**許容基準を満たすすべての独立変数が同時に投入されます。
- **ステップワイズ法を使用。**ステップワイズ法を使って変数の投入と削除を制御します。

▶ オプションとして、[選択] をクリックし、[ケース選択変数] ボックスに変数を選択することもできます。

判別分析: 範囲の定義

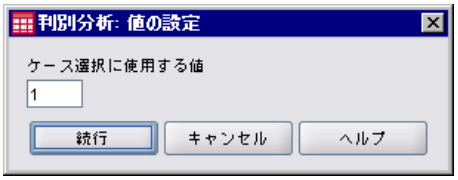
図 30-2
[判別分析: 範囲の定義] ダイアログ ボックス



グループ化変数の最小値と最大値を指定します。この範囲外の値のケースは、判別分析には使用されませんが、分析結果に基づいて既存のグループの 1 つに分類されます。最小値と最大値は整数でなければなりません。

判別分析: ケースの選択

図 30-3
[判別分析: 値の設定] ダイアログ ボックス



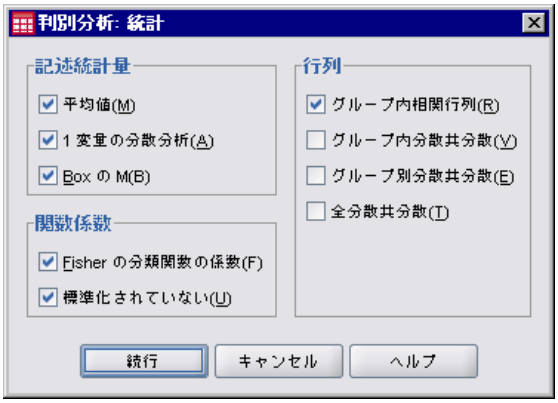
分析用にケースを選択するには、次のようにします。

- [判別分析] ダイアログ ボックスで、ケース選択変数を選択します。
- **[値の設定]** をクリックし、ケース選択に使用する整数の値を入力します。

判別関数の作成に使用されるのは、このボックスで指定したケース選択変数の値を持つケースのみです。統計および分類の結果は、選択されたケースと選択されていないケースの両方に対して生成されます。これを利用し、既存のデータに基づいて新しいケースを分類したり、データを学習用および検定用サブグループに分割し、作成されたモデルの有効性を確認したりできます。

判別分析: 統計

図 30-4
[判別分析: 統計] ダイアログ ボックス



記述統計。使用できるオプションは、[平均値] (標準偏差を含む)、[1 変量の分散分析]、および [Box の *M*] です。

- **平均値 (信頼性分析)。**独立変数に対して合計とグループ平均値、標準偏差を表示します。
- **1 変量の分散分析 (判別分析)。**各独立変数に対してグループ平均値の同等性を検定するため一元配置分散分析を実行します。
- **Box の M。**グループの分散共分散行列の同等性を調べる検定。サンプルが十分に大きい場合、有意でない p 値は、行列が異なるという証拠が不十分であることを意味します。この検定は、多変量正規性からの逸脱に対して敏感です。

関数係数。使用できるオプションは、[Fisher の分類関数の係数]、および [標準化されていない] です。

- **Fisher の分類関数の係数。**分類に直接使用できる Fisher の分類関数の係数を表示します。各グループについて係数のセットが求められ、ケースは最大判別得点を持つグループに割り当てられます。
- **標準化されていない (判別分析)。**標準化されていない判別関数の係数を表示します。

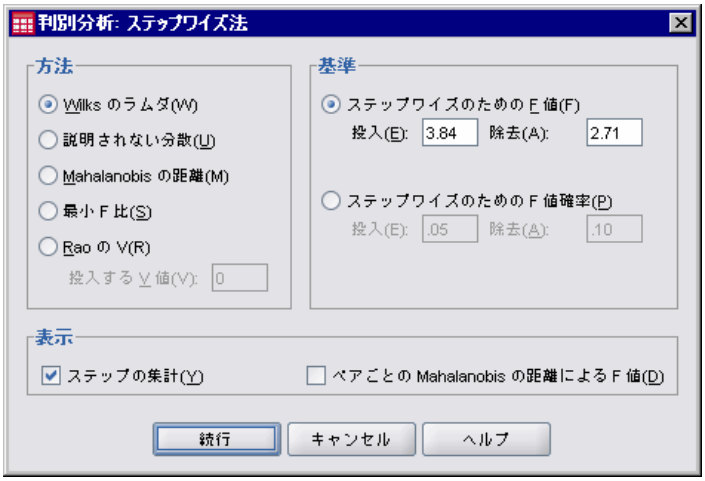
行列。独立変数に使用できる係数行列は、[グループ内相関行列]、[グループ内分散共分散]、[グループ別分散共分散]、および [全分散共分散] です。

- **グループ内相関行列。**相関を計算する前に、グループすべての個別の分散共分散行列を平均化することによって得られるプールされたグループ内相関行列を表示します。
- **グループ内分散共分散。**全分散共分散行列とは異なる場合もある、プールされたグループ内分散共分散行列を表示します。この行列は、すべてのグループの個別の分散共分散行列を平均化することによって得られます。

- **グループ別分散共分散。**各グループに個別の分散共分散行列を表示します。
- **全分散共分散。**1 つのサンプルから取り出したかのように、すべてのケースからの分散共分散行列を表示します。

判別分析: ステップワイズ法

図 30-5
[判別分析: ステップワイズ法] ダイアログ ボックス



方法。新しい変数を投入または除去するために使用する統計量を選択します。[Wilks のラムダ]、[説明されない分散]、[Mahalanobis の距離]、[最小 *F* 比]、および [Rao の *V*] を選択できます。[Rao の *V*] では、投入する変数に対して *V* の増加の最小値を指定できます。

- **Wilks のラムダ。**変数が Wilks のラムダを低下させる量を基に分析に入力する変数を選ぶ、ステップごとの判別分析における変数選択法。各ステップで、Wilks のラムダを最小化させる変数が入力されます。
- **説明されない分散。**各ステップで、グループ間の説明されない分散の合計を最小にする変数が投入されます。
- **Mahalanobis の距離。**独立変数のケースの値がケースの平均からどの程度離れているかを測るもの。大きい Mahalanobis の距離は、1 つ以上の独立変数に極値を持つケースを特定します。
- **最小 F 比。**グループ間の Mahalanobis の距離から計算した F 比の最大化に基づく、ステップワイズによる変数選択法。
- **Rao の V。**グループ平均間の差の測定値。Lawley-Hotelling のトレースとも呼ばれます。各ステップで、Rao の V における増加を最大化する値が入力されます。このオプションを選択したら、変数が分析に入るために必要な最小の値を入力してください。

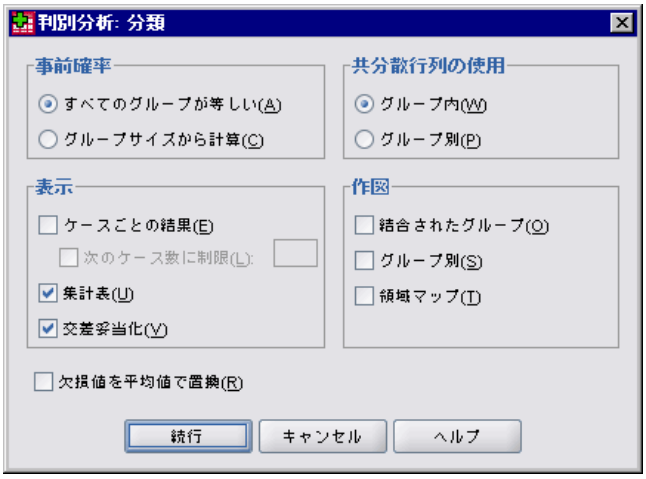
基準。利用できる選択肢は [**スワップワイズのための F 値**] と [**スワップワイズのための F 値確率**] です。変数の入力や削除に必要な値を指定します。

- **ステップワイズのための F 値。**変数は、その F 値が [投入] 値よりも大きい場合にモデルに投入され、[除去] よりも小さい場合に除去されます。[投入] の値は [除去] の値より大きい値である必要があり、さらに両方の値は正である必要があります。さらに多くの変数をモデルに投入するには、[投入] 値を下げてください。さらに多くの変数をモデルから除去するには、[除去] 値を上げてください。
- **ステップワイズのための F 値確率。**F 値の有意確率が [投入] の値よりも小さい場合、変数はモデルに入力され、[除去] の値よりも大きい場合除去されます。[投入] の値は [除去] の値より小さい値である必要があり、さらに両方の値は正である必要があります。さらに多くの変数をモデルに投入するには、[投入] 値を上げてください。さらに多くの変数をモデルから除去するには、[除去] 値を下げてください。

表示。[**ステップの集計**] をオンにすると、ステップごとにすべての変数の統計量が表示されます。[**ペアごとの Mahalanobis の距離による F 値**] をオンにすると、グループの各ペアに対してペアごとの *F* 比の行列が表示されます。

判別分析: 分類

図 30-6
[判別分析: 分類] ダイアログ ボックス



事前確率。このオプションは、分類係数を所属グループの事前知識に対して調整するかどうかを決定します。

- **すべてのグループが等しい。**すべてのグループに等しい事前確率が仮定されます。係数には影響がありません。
- **グループサイズから計算。**サンプルで観測されたグループサイズから、所属グループの事前確率を決定します。たとえば、分析に含まれる観測の 50% が 1 番目のグループに属し、25% が 2 番目に属し、25% が 3 番目に属す

場合、他の 2 つに対して 1 番目のグループに所属する尤度が大きくなるように、分類係数が調整されます。

表示。使用できる表示オプションは、[ケースごとの結果]、[集計表]、および[交差妥当化] です。

- **ケースごとの結果。**実際のグループ、予測グループ、事後確率、および判別得点のコードは、各ケースに表示されます。
- **集計表。**判別分析に基づいて各グループに正しくまたは誤って分けられたケースの数。「コンフュージョン行列 (Confusion Matrix)」と呼ばれることもあります。
- **交差妥当化。**分析における各ケースは、そのケース以外のすべてのケースから派生した関数で分類されます。これは、「U-方法」とも呼ばれます。

欠損値を平均値で置換。このオプションを選択すると、分類段階でのみ、欠損値が独立変数の平均値に置き換えられます。

共分散行列の使用。[グループ内] または [グループ別] のどちらかをクリックして、ケース分類に使用する共分散行列を選びます。

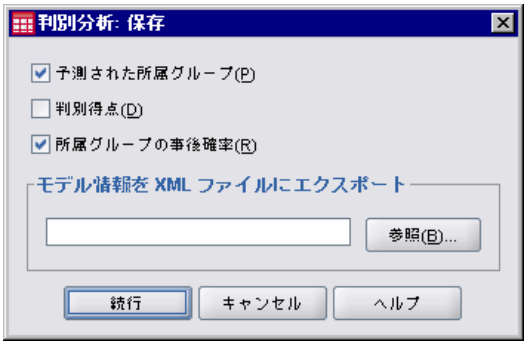
- **グループ内。**プールされたグループ内分散共分散行列は、ケースを分類するときに使用します。
- **グループ別。**グループ別分散共分散行列は、分類のために使われます。分類は元の変数ではなく判別関数に基づいているため、このオプションは 2 次の判別と等しくないときもあります。

作図。使用できるオプションは、[結合されたグループ]、[グループ別]、および [領域マップ] です。

- **結合されたグループ。**最初の 2 つの判別関数値のすべてのグループ散布図を作成します。関数が 1 つしかない場合は、代わりにヒストグラムが表示されます。
- **グループ別。**最初の 2 つの判別関数値のグループ別散布図を作成します。関数が 1 つしかない場合は、代わりにヒストグラムが表示されます。
- **領域マップ。**関数の値に基づいて、グループのケースを分類するための境界の作図です。これらの数字は、ケースが分類される先のグループに対応します。各グループの平均は、その境界内に星印で示されます。判別関数が 1 個の場合、このマップは表示されません。

判別分析: 保存

図 30-7
[判別分析: 保存] ダイアログ ボックス



作業中のデータ ファイルに新変数を追加することができます。使用できるオプションには、[予測された所属グループ] (単一変数)、[判別得点] (解の判別関数ごとに 1 つの変数)、および [所属グループの事後確率] (グループごとに 1 つの変数) です。

モデル情報を指定されたファイルに XML (PMML) 形式でエクスポートすることもできます。SmartScore、および SPSS (別製品) では、保存したモデル ファイルを使用して、得点付けのための他のデータ ファイルにモデル情報を適用できます。

DISCRIMINANT コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 複数の判別分析を実行し (1 つのコマンドを使用)、変数の投入順序を制御する (ANALYSIS サブコマンドを使用)。
- 分類に使用する事前確率を指定する (PRIORS を使用)。
- 回転後のパターン行列と構造行列を表示する (ROTATE サブコマンドを使用)。
- 抽出する判別関数の数を制限する (FUNCTIONS サブコマンドを使用)。
- 分析用に選択された (または選択されていない) ケースに対して分類を制限する (SELECT サブコマンドを使用)。
- 相関行列を読み込み、分析する (MATRIX サブコマンドを使用)。
- その後の分析で使用する相関行列を書き込む (MATRIX サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
31

因子分析

因子分析は、基礎となる変数、すなわち観測変数グループ内部の相関パターンを説明する **因子**を特定しようとする試みです。因子分析は、通常、データの分解の際に、多数の顕在変数で観測された分散のほとんどを説明する、少数の因子を識別するために使用します。因子分析は、原因のメカニズムに関する仮説を立てる場合、または次の分析に必要な変数を選別する（線型回帰分析を実行する前に共線性を識別するなど）場合にも使われます。

因子分析手続きは次のような柔軟性に富んだ機能を備えています。

- 7 種類の因子抽出方法があります。
- 非直交回転のためのプロマックスや直接オブリミン法など、5 種類の回転方法があります。
- 因子得点の計算方法は 3 種類あり、得点は変数として保存して、その後の分析で使用することができます。

例。政治についての調査質問に対する人々の回答内容は、どのような基本的姿勢によるものなのでしょうか。調査項目間の相関を調べると、税に関する質問が互いに相関していたり、軍事問題に関する質問が互いに相関しているなど、項目のさまざまなサブグループ間に有意な重複があることが明らかになります。因子分析では、基底因子の数を調べることができ、また多くの場合、因子が表す概念的な意味を識別することもできます。さらに、各回答者の因子得点を計算し、次の分析に利用することもできます。たとえば、因子得点をもとに、投票を予測するロジスティック回帰モデルを構築することもできます。

統計量。各変数: 有効ケース数、平均値および標準偏差。各因子分析: 有意確率、行列式、逆行列を含む変数の相関行列、反イメージなどの再生相関行列; 初期の解（共通性、固有値、および説明された分散のパーセント); サンプル抽出の適正さの Kaiser-Meyer-Olkin 測定および Bartlett の球面性検定; 因子負荷量、共通性、および固有値を含む回転前の解; 回転後のパターン行列や変換行列を含む回転後の解があります。オブリミン回転: 回転後のパターンと構造行列; 因子得点係数行列と因子共分散行列があります。固有値のスクリー プロットと最初の 2 因子または 3 因子の因子負荷プロットです。

データ。変数は、**区間**または**比例**尺度の量的変数である必要があります。カテゴリデータ（宗教、出生国など）は因子分析には適しません。Pearson の相関係数が正確に算出できるデータは因子分析に適しています。

仮定。データ中の変数の各ペアは 2 変量の正規分布に従い、観測値は独立している必要があります。因子分析モデルは、変数が共通因子 (モデルにより推測される因子) と独自因子 (観測変数間で重複しない因子) によって決定されること、すなわち計算された推測値は、すべての独自因子が互いに、また共通因子と相関しないという仮定に基づいていることを条件とします。

因子分析を行うには

- ▶ メニューから次の項目を選択します。

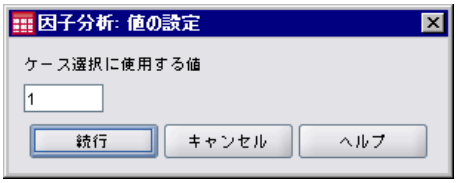
  分析
  データの分解
  因子分析...

- ▶ 因子分析のための変数を選択します。

図 31-1
[因子分析] ダイアログ ボックス

因子分析のケースの選択

図 31-2
[因子分析: 値の設定] ダイアログ ボックス



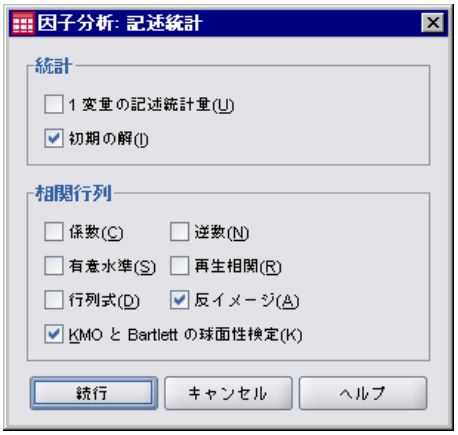
分析用にケースを選択するには、次のようにします。

- ▶ 選択変数を選択します。
- ▶ **[値の設定]** をクリックし、ケース選択に使用する整数の値を入力します。

  [ケース選択変数] に指定した値を持つケースのみが因子分析に使われます。

因子分析の記述統計

図 31-3
[因子分析: 記述統計] ダイアログ ボックス



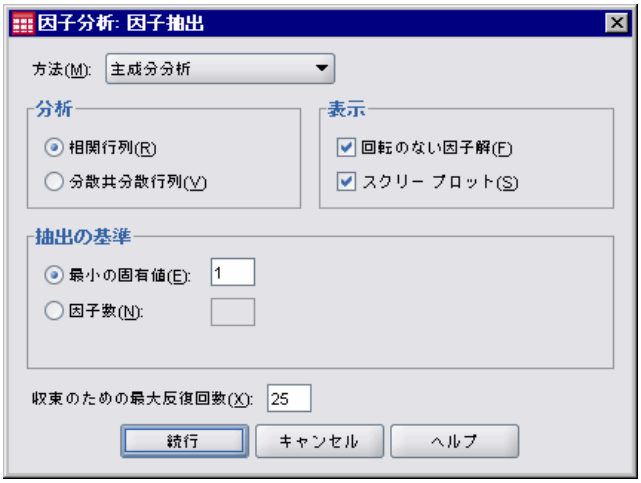
統計量。[**1 変量の記述統計**] をクリックすると、変数ごとの平均値、標準偏差、有効ケース数が表示されます。[**初期の解**] をクリックすると、初期の共通性と固有値、説明された分散のパーセントが表示されます。

相関行列。使用できるオプションは、係数、有意確率、行列式、KMO と Barlett の球面性検定、逆行列、再生相関、および反イメージです。

- **KMO と Bartlett の球面性検定。**変数間の偏相関が小さいかどうかを調べるサンプル適正検定の Kaiser-Meyer-Olkin 量。Bartlett の球面性の検定は、相関行列が単位行列かどうかを検定し、因子モデルが不適切かどうかを示します。
- **再生相関。**因子の解からの推定相関行列。残差 (推定された相関と観測された相関の差) も表示されます。
- **反イメージ。**反イメージ相関行列は偏相関係数の負の数を含み、反イメージ分散共分散行列はその偏共分散の負の数を含みます。よい因子モデルでは、対角線上にない要素のほとんどは小さい値となります。変数に対する抽出の妥当性は、反イメージ相関行列の対角線上に表示されます。

因子分析の因子抽出

図 31-4
[因子分析: 因子抽出] ダイアログ ボックス



方法。因子抽出の方法を指定できます。使用できる方法は、主成分分析、重み付けのない最小 2 乗法、一般化最小 2 乗法、最尤法、主因子法、アルファ因子法、イメージ因子法です。

- **主成分分析 (因子分析)。**観測変数の無相関線型結合を形成するために使われる因子抽出方法。第 1 主成分が最大の分散を持ちます。以降の成分は、その分散のより小さな部分を徐々に説明し、それらはすべて互いに相関しません。主成分分析は、初期因子解を得るために使われます。相関行列が特異であるときに使うことができます。
- **重み付けのない最小 2 乗法 (因子分析)。**対角成分を無視して、観測相関行列と再生相関行列間の差の 2 乗を最小化する因子抽出法です。
- **一般化最小 2 乗法 (因子分析)。**観測相関行列と再生相関行列の間の差の平方和を最小化する因子抽出法。相関は、一意性と逆に重み付けをするため、一意性の高い変数には一意性の低い変数より小さい重みが与えられます。
- **最尤法 (因子分析)。**サンプルが多変量正規分布から抽出されている場合に、観測相関行列を生成した可能性が最も高いパラメータ推定値を生成する因子抽出法。相関には、変数の一意性の逆数が重み付けされ、反復アルゴリズムが用いられます。
- **主因子法 (因子分析)。**共通性の初期推定値の対角線上に配置された重相関係数の 2 乗によって、元の相関行列から因子を抽出する方法。これらの因子負荷量は、対角線上にある古い共通性の推定値に置き換わる新しい共通性を推定するために使われます。反復による変化が抽出のための収束基準を満たすまで、反復は続きます。

- **アルファ (因子分析)。**分析中の変数を、ポテンシャル変数の母集団のサンプルと見なす因子抽出法。因子のアルファ信頼性を最大化します。
- **イメージ因子法 (因子分析)。**Guttman によってイメージ理論を基に開発された因子抽出法。偏イメージと呼ばれる変数の共通部分は、仮説的因子の関数としてではなく、残りの変数に関する線型回帰として定義されます。

分析。相関行列または共分散行列のどちらかを指定できます。

- **相関行列。**分析に含まれている変数を異なる尺度で測定する場合に便利です。
- **分散共分散行列。**各変数が異なる分散を持つ複数のグループに因子分析を適用する場合に便利です。

抽出の基準。固有値が指定した値を超えるすべての因子を保持するか、または保持する因子数を指定できます。

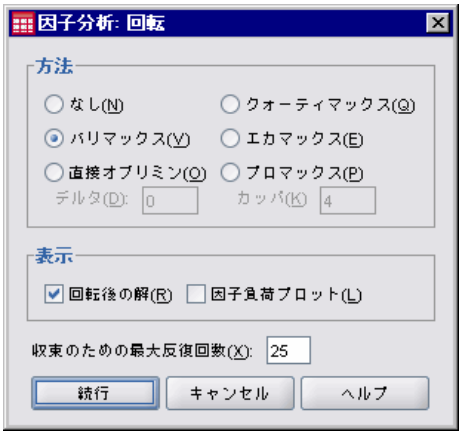
表示。回転前の因子解および固有値のスクリー プロットを指定できます。

- **回転のない因子解 (因子分析)。**因子解を求めるときに、回転を行わず求めた、因子負荷量、共通性、固有値を表示します。
- **スクリー プロット。**各因子と関連する分散のプロット。このプロットで、保持する因子の数を決めるために使われます。一般に、この作図は、大きい因子の急勾配と残りの因子の緩勾配の間に明瞭な区切りを示します。

収束のための最大反復回数。解を推定するためにアルゴリズムに含められるステップの最大数を指定できます。

因子分析の回転

図 31-5
[因子分析: 回転] ダイアログ ボックス



方法。因子回転の方法を選択することができます。使用できる方法は、バリマックス、直接オブリミン、コーティマックス、エカマックス、またはプロマックスです。

- **バリマックス法 (因子分析)。**各因子に高い負荷量を持つ変数の数を最小化する直交回転法。この方法は、因子の解釈を単純化します。

- **直接オブリミン法。**斜交 (非直交) 回転の方法。デルタが 0 (デフォルト) のとき、解は最も斜交します。デルタが負になるに従って、因子の斜交度は下ります。デフォルト値の 0 を無効にするには、0.8 以下の数を入力してください。
- **クォーティマックス法 (因子分析)。**各変数を説明するために必要な因子数を最小化する回転法。この方法は、観測変数の解釈を単純化します。
- **エカマックス法 (因子分析)。**因子を単純化するバリマックス法と、変数を単純化するクォーティマックス法を組み合せた回転法。因子に高い負荷を加える変数の数と、変数を説明するために必要な因子の数が最小化されます。
- **プロマックス回転 (因子分析)。**斜交回転であり、因子を相関させることができます。この回転は直接オブリミン回転よりも早く計算されるため、大きなデータセットを使用する場合に役立ちます。

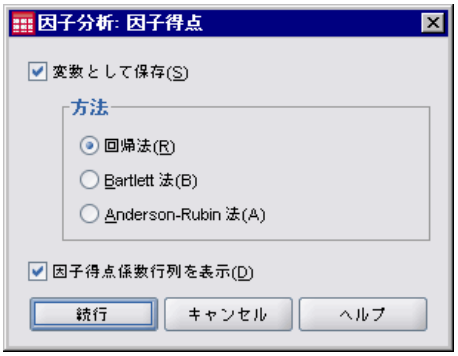
表示。最初の 2 因子または 3 因子の因子負荷プロットおよび回転後の解の出力を表示できます。

- **回転後の解 (因子分析)。**回転後の解を得るためには、回転方法が選択されている必要があります。直交回転では、回転されるパターン行列と因子変換行列が表示されます。斜交回転では、パターン、構成、および因子相関行列が表示されます。
- **因子負荷プロット。**最初の 3 つの因子の 3 次元の因子負荷プロット。2 因子解に対しては、2 次元の作図が表示されます。1 つの因子だけが抽出された場合、作図は表示されません。因子の回転が要求されると、回転した解を表示します。

収束のための最大反復回数。回転の実行のためにアルゴリズムに含められるステップの最大数を指定できます。

因子分析の因子得点

図 31-6
[因子分析: 因子得点] ダイアログ ボックス



変数として保存。最終解の各因子に対して新しい変数を 1 つ作成します。

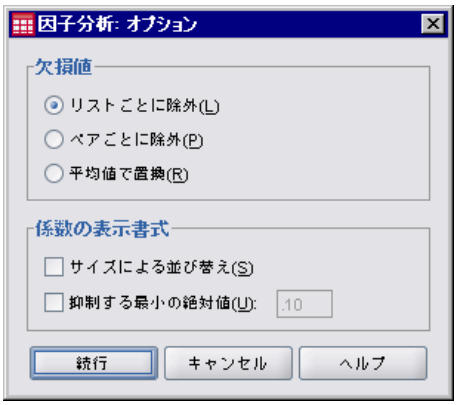
方法。因子得点を計算するための代わりの方法は、回帰、Bartlett 法または Anderson-Rubin 法です。

- **回帰法 (因子分析)。**因子得点係数を推定する方法。得られる得点では、平均値が 0 で、分散は推定因子得点と真の因子の値との重相関の 2 乗に等しくなります。得点は、因子が直交であっても相関することがあります。
- **Bartlett 得点。**因子得点係数を推定する方法。因子得点の平均値は 0 です。変数の範囲内にある独自の因子の平方和を最小化します。
- **Anderson-Rubin 法 (因子分析)。**因子得点係数の推定法。推定された因子の直交化を保証する Bartlett 法の修正版です。得られる得点では、平均が 0 で標準偏差が 1 となり、相関しません。

因子得点係数行列を表示。因子得点を計算するために変数に乗じる係数を表示します。因子得点間の相関も表示します。

因子分析オプション

図 31-7
[因子分析: オプション] ダイアログ ボックス



欠損値。欠損値の扱い方を指定できます。使用できる方法は、[**リストごとに**除外]、[**ペアごとに**除外]、または [平均値で置換] です。

係数の表示書式。出力行列の縦横比を制御できます。係数をサイズにより分類したり、絶対値が指定した値未満の係数を抑制できます。

FACTOR コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 抽出と回転における、反復の収束基準を指定します。
- 個別の回転因子のプロットを指定します。
- 保存する因子得点の個数を指定します。
- 主因子法の対角線要素の値を指定します。

- 相関行列や因子負荷行列を、今後の分析のためにディスクに書き込みます。
- 相関行列や因子負荷行列を読み込み、分析します。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
32
クラスタリングの手続きの選択

クラスタ分析は、TwoStep、階層、または大規模ファイルのクラスタ分析の各手続きを使用して実行できます。各手続きは、クラスタの作成に異なるアルゴリズムを採用しており、他の手続きでは使用できないオプションを備えています。

TwoStep クラスタ分析。[2 ステップ クラスタ分析] 手続きは、多くのアプリケーションで選択される方法です。この手続きには、次に示す独自の機能があります。

- クラスタ モデルを選択するだけではなく、最適なクラスタ数を自動的に選択します。
- カテゴリ変数や連続変数に基づいてクラスタ モデルを同時に作成できます。
- クラスタ モデルを外部の XML ファイルに保存したり、このファイルを読み込んで新しいデータでクラスタ モデルを更新できます。

さらに、[TwoStep クラスタ分析] 手続きは、大規模なデータ ファイルを分析することもできます。

階層クラスタ分析。[階層クラスタ分析] 手続きは、小規模なデータ ファイル (クラスタ化されるオブジェクトが数百個のもの) に限定されますが、次に示す独自の機能があります。

- ケースまたは変数をクラスタ化する機能。
- 可能な解の範囲を計算し、それらの解の所属クラスタを保存する機能。
- 複数の方法でクラスタの書式化、変数の変換、およびクラスタ間の非類似度の測定を実行。

すべての変数が同じ種類である場合は、[階層クラスタ分析] 手続きによって、間隔 (連続)、度数、または 2 値変数を分析できます。

大規模ファイルのクラスタ分析。[大規模ファイルのクラスタ分析] 手続きは、連続データに限られ、クラスタの個数を事前に指定しておく必要がありますが、次に示す独自の機能があります。

- 各オブジェクトについて、クラスタの中心からの距離を保存する機能。
- 外部 SPSS ファイルから初期クラスタ中心を読み込み、最後のクラスタ中心を外部 SPSS ファイルに保存する機能。

さらに、[大規模ファイルのクラスタ分析] 手続きは、大規模なデータ ファイルを分析することもできます。

章
33

TwoStep クラスタ分析

[TwoStep クラスタ分析] 手続きは、データセット内の自然なグループ (またはクラスタ) を明らかにすることを目的として設計された探索ツールです。この手続きで使用されるアルゴリズムには、従来のクラスタ分析技術とは異なるいくつかの優れた特徴があります。

- **カテゴリ変数と連続変数の処理。**変数が独立していると仮定することにより、カテゴリ変数および連続変数について、多項分布と正規分布を結合できます。
- **クラスタ数の自動選択。**異なるクラスタ解の間でモデル選択基準の値を比較することにより、手続きで、最適なクラスタ数を自動的に判断できます。
- **スケーラビリティ。**レコードを要約するクラスタ機能 (CF) ツリーを作成することにより、TwoStep アルゴリズムで大きなデータ ファイルを解析できます。

例。小売業者や消費財業者は、顧客の購買傾向、性別、年齢、収入などのデータにクラスタリング技術を日常的に応用しています。売上を伸ばし、ブランド ロイヤルティを構築するための各顧客グループに合せたマーケティング戦略および製品開発戦略を立てています。

統計量。この手続きにより、情報の基準 (AIC または BIC) が解のクラスタ数別に、最終クラスタ分析のクラスタ度数と記述統計量がクラスタ別に生成されます。

作図。この手続きにより、クラスタ度数の棒グラフ、クラスタ度数の円グラフ、および変数重要度グラフが作成されます。

図 33-1
[TwoStep クラスタ分析] ダイアログ ボックス

距離測度。2 つのクラスタ間の類似度を計算する方法を指定します。

- **対数尤度。**この尤度測定により、変数の確率分布を求めます。連続変数は正規分布しているものと仮定し、カテゴリ変数は多項分布しているものと仮定します。すべての変数は独立しているものと仮定します。
- **ユークリッド。**ユークリッド測定は、2 つのクラスタ間の「直線」距離です。この測定方法は、すべての変数が連続している場合にだけ使用できます。

クラスタ数。この項目を選択すると、クラスタの個数の判定方法を指定できます。

- **自動的に判定。**この手続きは、[クラスタ化の基準] グループで指定された基準を使用して、「最適な」クラスタの個数を自動的に判定します。オプションとして、手続きで考慮するクラスタの個数の最大値を正の整数で入力できます。
- **固定値を指定。**解に含まれるクラスタの個数を固定できます。正の整数を入力します。

連続変数の数。このグループは、[オプション] ダイアログ ボックスで指定された連続変数の標準化設定の集計を示します。詳細は、406 ページのTwoStep クラスタ分析のオプションを参照してください。

クラスタ化の基準。自動クラスタ化アルゴリズムで、クラスタの個数を判定する方法を指定します。ベイズの情報量基準 (BIC) または赤池情報量基準 (AIC) のどちらかを指定できます。

データ。この手続きは、連続変数とカテゴリ変数の両方に使用できます。ケースはクラスタ化されるオブジェクトを表し、変数はクラスタ化の基準となる属性を表します。

ケースの並び順。クラスタ機能ツリーと最終解は、ケースの並び順によって異なる可能性があることに注意してください。並び順の影響を最小限に抑えるには、ケースを無作為に並べます。異なる無作為な順序で並べ替えられたケースを使用していくつかの異なる解を得ることにより、特定の解の安定性を確認できます。ファイル サイズが非常に大きいことが原因でこの操作の実行が難しい場合は、異なる無作為な順序で並べ替えられたケースのサンプルを使用し、複数回に分けて実行することで代用します。

仮定。尤度距離測度は、クラスタ モデル内の変数は独立しているものと仮定します。さらに、各連続変数は正規分布であると仮定し、カテゴリ変数は多項分布であると仮定します。経験的内部検定は、この手続きが独立および分布仮定の違反に対して堅牢であることを示していますが、これらの仮定がどの程度満たされているか、把握するようにしてください。

2 つの連続変数の独立性を検定するには、[2 変量の相関分析] 手続きを使用します。2 つのカテゴリ変数の独立性を検定するには、[クロス集計表] 手続きを使用します。連続変数とカテゴリ変数間の独立性を検定するには、[グループの平均] 手続きを使用します。連続変数の正規性を検定するには、[探索的分析] 手続きを使用します。カテゴリ変数が、指定された多項分布になっているかどうかを検定するには、[カイ 2 乗検定] 手続きを使用します。

階層クラスタ分析を行うには

▶ メニューから次の項目を選択します。

分析
分類
2 ステップ クラスタ...

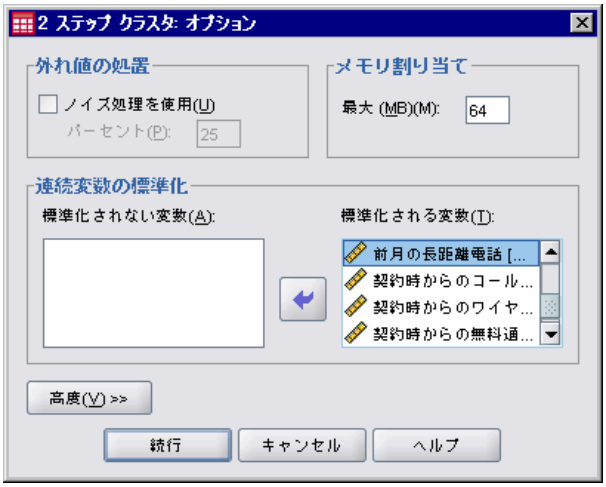
▶ 1 つ以上のカテゴリ変数または連続変数を選択します。

オプションとして、次の選択が可能です。

- クラスタの作成基準を調整する。
- ノイズ処理、メモリー割り当て、変数の標準化、およびクラスタ モデル投入の設定を選択する。
- オプションの表およびプロットをリクエストする。
- モデルの結果を作業ファイルまたは外部 XML ファイルに保存する。

TwoStep クラスタ分析のオプション

図 33-2
[TwoStep クラスタ オプション] ダイアログ ボックス



外れ値の処置。クラスタ機能 (CF) ツリーが満杯の場合は、このグループで、クラスタ化の間に外れ値を特別に処理できます。CF ツリーが葉ノードでこれ以上ケースを受け入れられない場合、および葉ノードが分割できない場合、その CF ツリーは満杯です。

- ノイズ処理を選択したが、CF ツリーが埋まっている場合は、空きがある葉の中にあるケースを「ノイズ」葉に配置した後に、CF ツリーが再生されます。ツリー内のケースが最大葉サイズに対する指定パーセントよりも少ない場合、そのリーフには、空きがあると見なされます。ツリーが再生された後、可能であれば外れ値が CF に配置されます。不可能であれば、外れ値は廃棄されます。
- ノイズ処理を選択せず、CF ツリーが満杯の場合は、さらに大きな距離変更しきい値を使用して、CF ツリーが再生されます。最終クラスタ化の後、クラスタに割り当てられなかった値は、外れ値としてラベル付けされます。外れ値クラスタには、–1 という ID 番号が与えられ、クラスタの個数に含まれません。

メモリ割り当て。このグループで、クラスタ アルゴリズムが使用する最大メモリー量をメガバイト (MB) で指定できます。手続きがこの最大値を超えると、ディスクを使用して、メモリー内に納まらない情報を保存します。4 以上の値を指定します。

- 自分のシステムで指定できる最大値については、システム管理者に問い合わせてください。
- この値が小さすぎると、アルゴリズムが正しいクラスタの個数、または望ましいクラスタの個数を検出できない場合があります。

変数の標準化。クラスタ化アルゴリズムは、標準化された連続変数で使用できます。標準化されていない連続変数は、[標準化される変数] リストの変数として残しておきます。コンピュータ操作に費やす時間と手間を省くには、すでに標準化してある連続変数を [標準化されない変数] リストの変数として選択します。

[高度] オプション

CF ツリーの調節基準。次のクラスタ化アルゴリズム設定は、クラスタ化の特徴 (CF) ツリーだけに適用されるもので、慎重に変更する必要があります。

- **距離の変化の初期のしきい値。**これは、CF ツリーの生成に使用される初期しきい値です。指定したケースを CF ツリーの葉に挿入したときにしきい値未満である場合、その葉は分割されません。このしきい値を超える場合は分割されます。
- **最大枝数 (葉ノードごと)。**葉ノードが持つことのできる子ノードの最大数。
- **ツリーの最大の深さ。**CF ツリーが持つことのできるレベルの最大数。
- **最大ノード数。**関数 $(b^{d+1} - 1) / (b - 1)$ に基づいて、手続きにより生成できる CF ツリー ノードの最大可能数を示します。b は最大枝数、d はツリーの最大の深さを表します。CF ツリーが大きすぎると、システム リソースが浪費され、手続きの動作にとって逆効果となる場合があることに注意してください。最低でも、各ノードに 16 バイトは必要です。

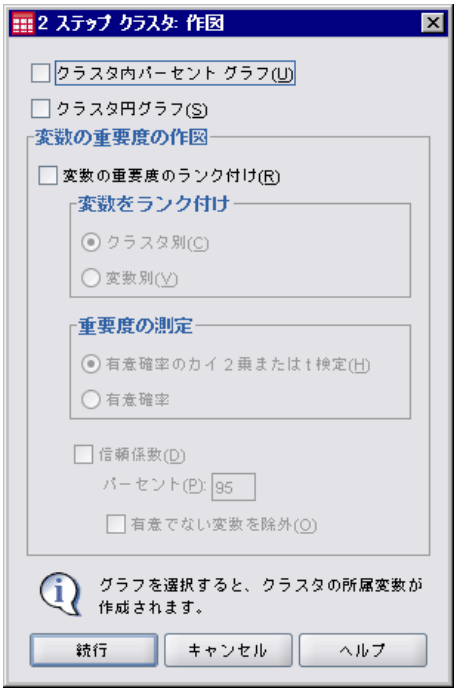
クラスタ モデルの更新。先に実行した分析で生成されたクラスタ モデルをインポートおよび更新できます。入力ファイルには、CF ツリーが XML 形式で格納されています。モデルは、アクティブ ファイルのデータで更新されます。メイン ダイアログ ボックスで、先の分析で指定したときと同じ順序で変数名を選択する必要があります。同じファイル名で新しいモデル情報を記述しない限り、XML ファイルは変更されません。詳細は、409 ページのTwoStep クラスタ分析の出力を参照してください。

クラスタ モデルの更新を指定すると、元のモデルに対して指定された CF ツリーの世代に関するオプションが使用されます。具体的には、保存されたモデルの距離測度、ノイズ処理、メモリー割り当て、または CF ツリーの基準調節設定が使用され、ダイアログ ボックス内のこれらのオプション設定はすべて無視されます。

注: クラスタ モデルの更新を行うとき、手続きでは、元のクラスタ モデルの作成に、アクティブなデータセット内の選択ケースは使用されなかったものと仮定します。また、モデルの更新に使用されるケースは、元のモデルの作成に使用されたケースと同じ分布のもの、つまり、連続変数の平均値と分散、およびカテゴリ変数のレベルは、どちらのケース グループについても同じであると仮定します。「新しい」ケースと「古い」ケースが不均質の分布のものである場合は、結合されたケース グループで [TwoStep クラスタ分析] 手続きを実行して、最良の結果を求めてください。

TwoStep クラスタ分析の作図

図 33-3
[2 ステップ クラスタ作図] ダイアログ ボックス



クラスタ内パーセント グラフ。各変数のクラスタ内変動を示す図表を表示します。それぞれのカテゴリ変数に対して、クラスタ ID ごとにカテゴリ度数を示すクラスタ棒グラフが作成されます。それぞれの連続変数に対して、クラスタ ID ごとにエラー バーを示すエラー バー グラフが作成されます。

クラスタ円グラフ。各クラスタ内の空セルのパーセントと度数を示す円グラフを表示します。

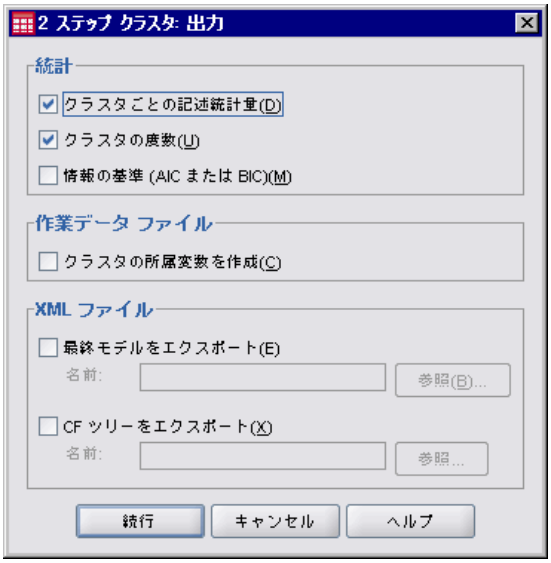
変数の重要度の作図。各クラスタ内の各変数の重要度を示すいくつかの異なる図表を表示します。この出力は、各変数の重要度順に並べ替えられます。

- **変数を順位付け。**このオプションで、クラスタごとにプロットを作成するか ([**クラスタ別**])、変数ごとにプロットを作成するか ([**変数別**]) を決定します。
- **重要度の測定。**このオプションで、作図する変数の重要度を測定する方法を選択できます。[**有意確率のカイ 2 乗または t 検定**] は、Pearson のカイ 2 乗統計量をカテゴリ変数の重要度として報告し、*t* 統計量を連続変数の重要度として報告します。[**有意確率**] は、連続変数について平均値の等質性を検定した *p* 値を 1 から引いた値、およびカテゴリ変数についてデータセット全体での期待度数を報告します。

- **信頼係数。**このオプションで、変数の全体分布に対して、クラスタ内の変数の分布の等質性検定を実行するときに適用する信頼水準を設定できます。50 以上 100 未満の数値を指定します。変数ごとに作図した場合、または変数の重要度を作図した場合、信頼水準の値が変数の重要度のプロットに垂直の線として表示されます。
- **有意でない変数を除外。**指定した有意確率において有意でない変数は、変数の重要度プロットに表示されません。

TwoStep クラスタ分析の出力

図 33-4
[2 ステップ クラスタ: 出力] ダイアログ ボックス



統計量。このグループで、クラスタ化の結果を示す表を表示するためのオプションを指定します。記述統計量とクラスタ度数は、最終クラスタ モデル用に生成され、情報の基準表に、クラスタ解の範囲の結果が表示されます。

- **クラスタごとの記述統計量。**各クラスタ内の変数について説明する 2 つの表を表示します。一方の表には、連続変数の平均値と標準偏差がクラスタ別に報告されます。もう一方の表には、カテゴリ変数の度数がクラスタ別に報告されます。
- **クラスタの度数。**各クラスタ内の観測数を報告する表を表示します。
- **情報の基準 (AIC または BIC)。**メイン ダイアログ ボックスで選択した基準に応じて、さまざまなクラスタの個数に関する AIC または BIC の値を含む表を表示します。この表は、クラスタの個数が自動的に判定されている場合にだけ表示されます。クラスタの個数が固定の場合、この設定は無視され、表は表示されません。

作業データ ファイル。このグループで、変数をアクティブなデータセットに保存します。

- **クラスタの所属変数を作成。**この変数には、各ケースのクラスタ ID 番号が含まれています。この変数の名前は tsc_n で、*n* は、所定のセッションでこの手順により完了するアクティブなデータセットの順序を示す正の整数です。

XML ファイル。XML 形式でエクスポートできる出力ファイルは、最終クラスタ モデルおよび CF ツリーです。

- **最終モデルをエクスポート。**最終クラスタ モデルが、指定されたファイルに XML (PMML) 形式でエクスポートされます。SmartScore、および SPSS (別製品) では、保存したモデル ファイルを使用して、得点付けのための他のデータ ファイルにモデル情報を適用できます。
- **CF ツリーをエクスポート。**このオプションで、クラスタ ツリーの現在の状態を保存し、後で新しいデータを使用して更新できます。

章
34

階層クラスタ分析

この手続きでは、個々のクラスタの各ケース（または変数）で始まるアルゴリズムを使用して、相対的に等質なケース（または変数）をグループの選択された特性に基づいて識別し、さらに、最後の 1 つになるまで他のすべてのクラスタを結合させます。元の変数を分析したり、さまざまな標準化変換を選択することもできます。距離と類似度は、[近接] 手続き（近接性分析）で生成します。統計量が段階ごとに表示されるので、最良の解を選択するのに役立ちます。

例。同じような視聴者層を持つテレビ番組のグループを識別することはできるでしょうか。階層クラスタ分析では、テレビ番組（ケース）を視聴者の特性に基づいて等質グループにクラスタ化できます。この分析結果は、コマーシャルの時間帯の特定に使用できます。また、都市（ケース）を等質グループにクラスタし、比較する都市を選択していろいろなマーケティング戦略を検定できます。

統計量。クラスタ凝集経過工程、距離（または類似）行列、および単一の解または解の範囲の所属クラスタ。プロット: デンドログラムおよびつららプロット。

データ。変数は数量、2 値、または度数データです。尺度の違いがクラスタの解に影響することがあるため、変数の尺度は重要です。変数の尺度がまったく異なる場合（たとえば、ある変数はドル単位で測定され、別の変数は年単位で測定される場合）、変数の標準化（[階層クラスタ分析] 手続きで自動的に実行可能）を行う必要があります。

ケースの並び順。同一の距離または類似度が入力データ内に存在するか、結合中に更新されたクラスタ内で発生した場合、導き出されるクラスタの解は、ファイル内のケースの並び順によって異なる可能性があります。異なる無作為な順序で並べ替えられたケースを使用していくつかの異なる解を得ることにより、特定の解の安定性を確認できます。

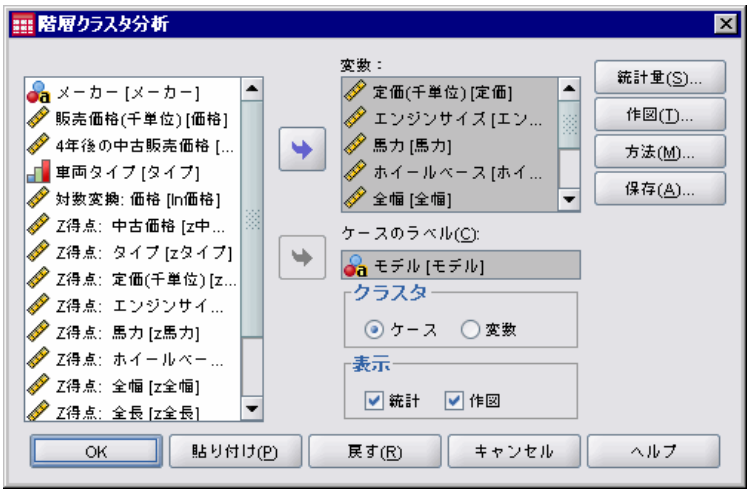
仮定。使用する距離、または類似度の測定方法は、分析データに適していなければなりません（距離と類似度の測定方法の選択については、[近接] 手続きを参照してください）。また、必要な変数をすべて分析に含める必要があります。重要な変数を省略すると、誤った解が生成される可能性があります。階層クラスタ分析は探索的手法なので、結果は独立サンプルで確認するまで仮説扱いとなります。

階層クラスタ分析を行うには

▶ メニューから次の項目を選択します。

分析
  分類
    階層クラスタ...

図 34-1
[階層クラスタ分析] ダイアログ ボックス

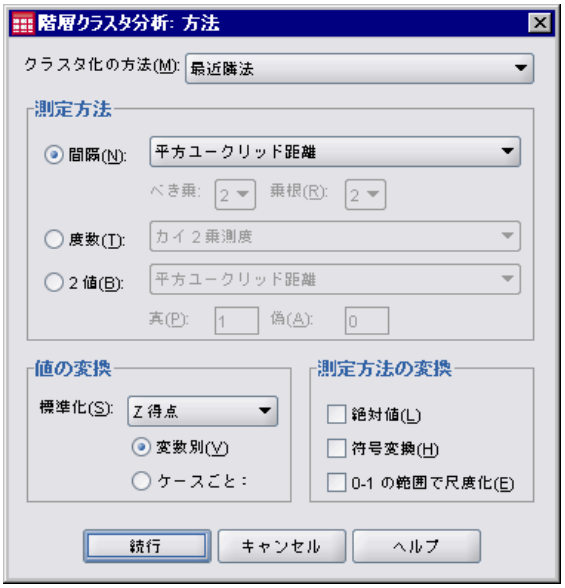


▶ ケースをクラスタ化するには、最低 1 つの数値型変数を選択します。変数をクラスタ化するには、最低 3 つの数値型変数を選択します。

オプションとして、ケースのラベル付けをする識別変数を選択することもできます。

階層クラスタ分析の方法

図 34-2
[階層クラスタ分析: 方法] ダイアログ ボックス



クラスタ化の方法。使用できるオプションは [グループ間平均連結法]、[グループ内平均連結法]、[最近隣法]、[最遠隣法]、[重心法]、[メディアン法]、または [Ward 法]です。

測定方法。距離や類似度の測定方法を指定して、クラスタ化に使用できます。データの種類とそれぞれの距離や類似度の測定方法を指定します。

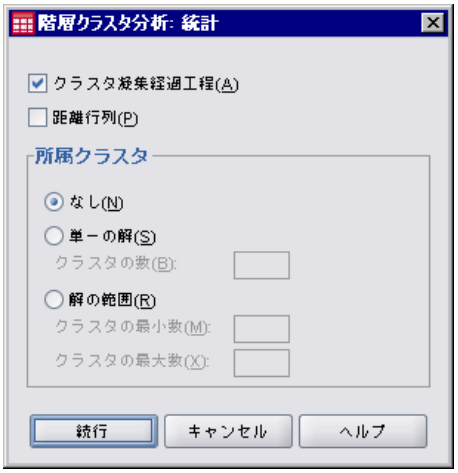
- **間隔。**使用できる測定方法は [ユークリッド距離]、[平方ユークリッド距離]、[コサイン]、[Pearson の相関]、[Chebychev]、[都市ブロック]、[Minkowski]、または [カスタマイズ] です。
- **度数。**使用できる測定方法は [カイ 2 乗測度] または [ファイ 2 乗測度] です。
- **2 値。**使用できる測定方法は [ユークリッド距離]、[平方ユークリッド距離]、[サイズの差異]、[パターンの差異]、[分散]、[散らばり]、[形]、[単純マッチング]、[ファイ 4 分点相関係数]、[ラムダ]、[Anderberg の *D*]、[dice]、[Hamann]、[Jaccard]、[Kulczynski 1]、[Kulczynski 2]、[Lance と Williams]、[落合]、[Rogers と Tanimoto]、[Russel と Rao]、[Sokal と Sneath 1]、[Sokal と Sneath 2]、[Sokal と Sneath 3]、[Sokal と Sneath 4]、[Sokal と Sneath 5]、[Yule *Y*]、または [Yule *Q*] です。

値の変換。近接度を計算する前に、ケースまたは値のデータを標準化します (2 値データには使用できません)。使用できる標準化方法は、[*z* 得点]、[−1 から 1 の範囲]、[0 から 1 の範囲]、[最大値を 1]、[平均値を 1]、または [標準偏差を 1] です。

測定方法の変換。距離測度によって生成された値を変換できます。変換した値は、距離を計算した後で適用されます。使用できるオプションは、[絶対値]、[符号変換]、および [0 ～ 1 の範囲で尺度化] です。

階層クラスタ分析の統計

図 34-3
[階層クラスタ分析: 統計] ダイアログ ボックス



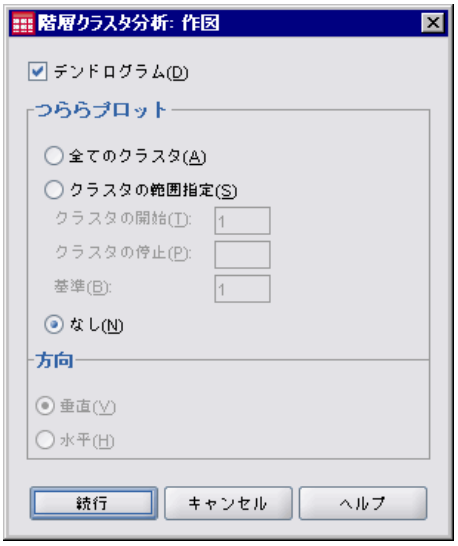
クラスタ凝集経過工程。各段階で結合されたケースまたはクラスタ、結合中のケースまたはクラスタ間の距離、およびケース (または変数) がクラスタに結合した最後のクラスタ レベルを表示します。

距離行列。項目間の距離または類似度を計算します。

所属クラスタ。クラスタを結合する 1 つ以上の段階で、各ケースが割り当てられているクラスタを表示します。使用できるオプションは、[単一の解] および [解の範囲] です。

階層クラスタ分析の作図

図 34-4
[階層クラスタ分析: 作図] ダイアログ ボックス

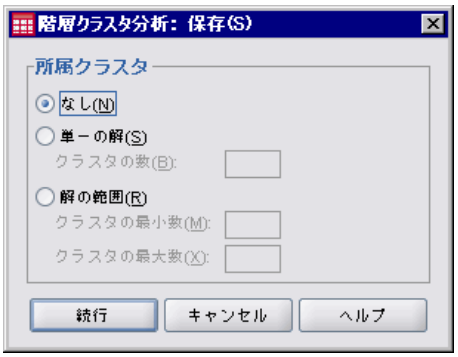


デンドログラム。デンドログラムを表示します。デンドログラムは、形成されたクラスタの結合性の評価に使用し、適当なクラスタ数を保持するための情報を提供します。

つららプロット。すべてのクラスタまたは指定された範囲のクラスタなどの **つららプロット**を表示します。つららプロットは、分析の反復ごとに、ケースのクラスタへの結合のしかたに関する情報を表示します。[方向] では、[垂直] または [水平] をクリックして、作図の方向を指定します。

階層クラスタ分析の新変数の保存

図 34-5
[階層クラスタ分析: 新変数の保存] ダイアログ ボックス



所属クラスタ。[単一の解] または [解の範囲] の所属クラスタを保存できます。保存変数は、その後の分析で、グループ間のその他の差の探索に使用できます。

CLUSTER コマンド シンタックスの追加機能

階層クラスタ手続きは、CLUSTER コマンドシンタックスを使用します。コマンド シンタックスを使用すると、次の作業も実行できます。

- 1つの分析にいくつかのクラスタの方法を使用。
- 近接行列の読み込み、および分析。
- 後で分析するために、近接行列をディスクに書き込み。
- ユーザー指定の([べき乗])距離速度に、べき乗と平方根の値を任意に指定。
- 保存変数名を指定。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
35
大規模ファイルのクラスタ分析

この手続きは、大量のケースを処理できるアルゴリズムを使用して、選んだ特性に基づくケースの中で相対的に等質のグループの識別を試みます。ただし、このアルゴリズムを使用するには、クラスタの個数を指定する必要があります。あらかじめわかっていれば、初期クラスタ中心を指定できます。ケースを分類するには、クラスタ中心を反復更新するか、分類するだけにするかの 2 つの方法のうち、どちらかを選択できます。保存できるのは、所属クラスタ、距離情報、クラスタ中心の最終値です。オプションとして、値をケースごとの出力のラベル付けに使用する変数を指定できます。また、分散分析の *F* 統計量も要求できます。この統計は便宜的なものですが (この手続きでは性質の違うグループを形成しようとするため)、統計量のサイズを比較すると、グループの分離に対する変数ごとの寄与率の情報を得ることができます。

例。各グループ内で同じような視聴者層を持つテレビ番組のグループで識別可能なのはどのようなものでしょうか。大規模ファイルのクラスタ分析では、視聴者の特性を基に、テレビ番組 (ケース) を *k* 個の等質グループにクラスタ化できます。この処理は、コマーシャルの時間帯の特定に使用できます。また、都市 (ケース) を等質グループにクラスタし、比較する都市を選択していろいろなマーケティング戦略を検定できます。

統計量。完全な解として初期クラスタ中心、分散分析表。ケースごとのクラスタ情報、クラスタ中心からの距離。

データ。変数は区間尺度または比例尺度の量的変数でなければなりません。変数が 2 値または度数の場合は、階層クラスタ分析手続きを使用します。

ケースと初期クラスタ中心の順序。初期クラスタ中心を選択するためのデフォルトのアルゴリズムは、ケースの順序によって変化します。[反復] ダイアログの **[移動平均を使用]** オプションを使用すると、初期クラスタ中心の選択方法に関係なく、解が潜在的にケースの順序に依存するようになります。これらの方法のいずれかを使用している場合、異なる無作為な順序で並べ替えられたケースを使用していくつかの異なる解を得ることにより、特定の解の安定性を確認できます。初期クラスタ中心を指定し、**[移動平均を使用]** オプションを使用しない場合は、ケースの順序に関連する問題を回避できます。ただし、ケースからクラスタ中心に同一の距離がある場合、初期クラスタ中心の順序は解に影響する場合があります。特定の解の安定性を評価するには、初期中心値の異なる転置の分析から得られた結果を比較します。

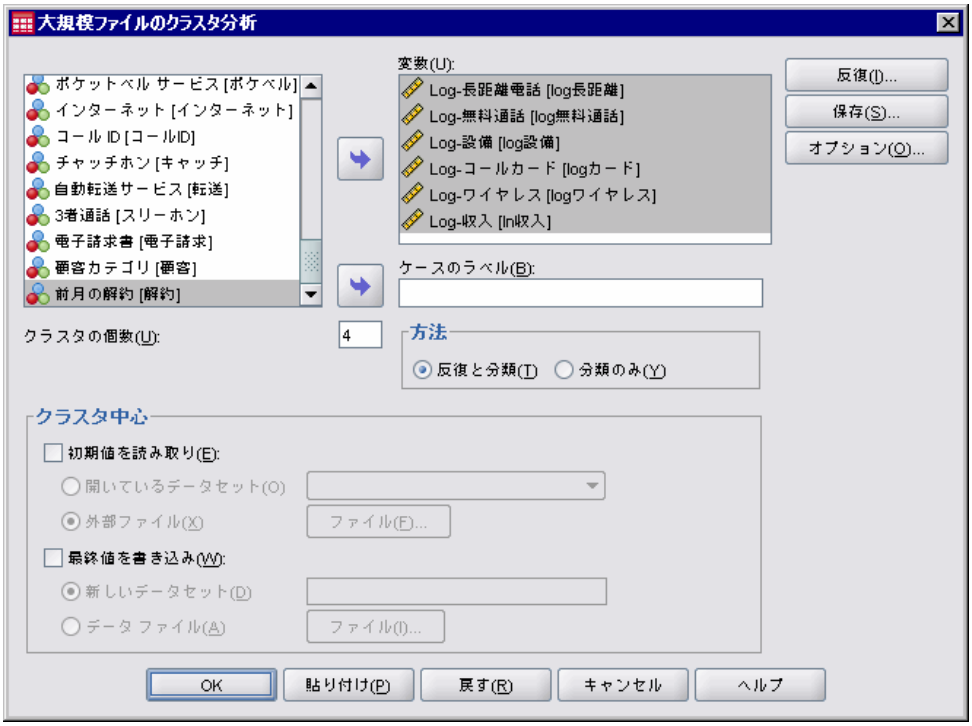
仮定。距離は、単純なユークリッド距離を使用して計算します。距離または類似度の別の測定方法を使用したい場合は、階層クラスタ分析手続きにします。変数の尺度は、重要な考慮事項です。変数が異なる尺度で測定されていると (ある変数はドル単位、別の変数は年単位など)、結果が疑わしくなります。このような場合には、大規模ファイルのクラスタ分析を実行する前に、まず変数を標準化する必要があります (このタスクは、[記述統計] で行うことができます)。この手続きでは、適切な個数のクラスタを選択し、必要な変数がすべて含まれていると仮定します。選んだクラスタの個数が不適切だったり、重要な変数が抜けてたりしていると、適切な結果を得られない場合があります。

大規模ファイルのクラスタ分析を行うには

▶ メニューから次の項目を選択します。

分析
分類
大規模ファイルのクラスタ...

図 35-1
[大規模ファイルのクラスタ分析] ダイアログ ボックス



▶ クラスタ分析で使用する変数を選びます。

▶ [クラスタの個数] ボックスでクラスタの個数を指定します。(クラスタの個数は、2 つ以上で、そのデータ ファイルのケース数以下でなければなりません。)

- ▶ [**反復と分類**] または [**分類のみ**] を選択します。
- ▶ オプションとして、ケースにラベルを付けるための識別変数を選択できます。

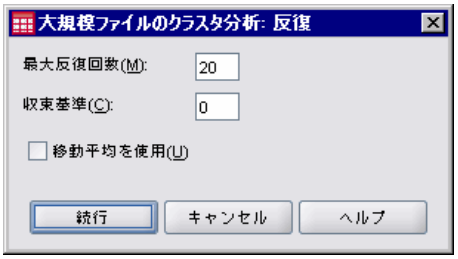
大規模ファイルのクラスタ分析の効率

大規模ファイルのクラスタ分析のコマンドは、[階層クラスタ分析] コマンドで採用するアルゴリズムを始めとする多くのクラスタ化の方法とは異なり、距離の計算でケースのすべてのペア間が対象とならないため、基本的には効率的です。

効率を最高にするには、ケースのサンプルを使用し、[**反復と分類**] をクリックして、クラスタ中心を決めます。[**最終値の書き出し**] を選択します。全体のデータ ファイルを読み込んでから、[**分類のみ**] を方法として選択、[**初期値の読み込み**] を選択し、サンプルから推定した中心を使ってファイル全体を分類します。ファイルやデータセットからの読み込み、およびそれらに対する書き出しを行なうことができます。データセットは、同じセッションの今後で利用可能ですが、セッション終了前に明示的に保存しない限り、保存されません。データセット名は、変数命名規則に従う必要があります。 詳細は、85 ページの 5 章の変数名を参照してください。

大規模ファイルのクラスタ分析の反復

図 35-2
[大規模ファイルのクラスタ分析: 反復] ダイアログ ボックス



注: これらのオプションが利用可能となるのは、[**反復と分類**] 方法を [大規模ファイルのクラスタ分析] ダイアログ ボックスで選択した場合のみです。

最大反復回数。大規模ファイルのクラスタ分析のアルゴリズムでの最大反復回数に上限を設けます。収束基準が満たされない場合でも、この回数の反復が終われば、それ以上の反復は行われません。入力する値は、1 から 999 の範囲内でなければなりません。

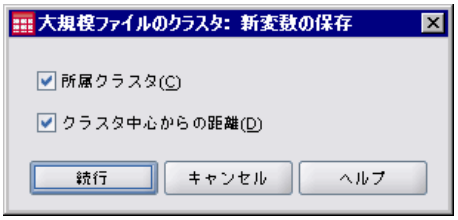
バージョン 5.0 以前の Quick Cluster コマンドで使用するアルゴリズムを再現するには、[**最大反復回数**] ボックスの値を 1 に設定します。

収束基準。反復停止のタイミングを決定します。初期クラスタ中心間の最小距離の比率を表しているので、0 よりも大きく、1 以下である必要があります。たとえば、この基準が 0.02 に等しければ、完全な反復を行ったときにクラスタ中心が初期クラスタ中心間の最小距離の 2% を超える距離だけ移動しない場合に反復が停止します。

移動平均を使用。ケースごとの割り当てが終わった後にクラスタ中心を更新するよう要求できます。このオプションを選ばないと、新しいクラスタ中心の計算は、すべてのケースの割り当てが終わった後に行われます。

大規模ファイルのクラスタ分析の保存

図 35-3
[大規模ファイルのクラスタ分析: 新変数の保存] ダイアログ ボックス



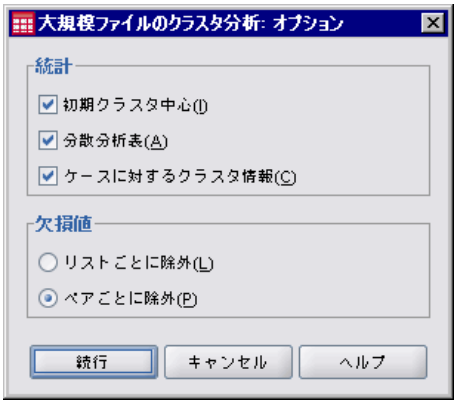
解に関する情報は、新変数として保存し、その後の分析に利用できます。

所属クラスタ。ケースごとの最終所属クラスタを示す新変数を作成します。新変数の値は、1 からクラスタの個数の範囲内です。

クラスタ中心からの距離。各ケースとその分類の中心との間のユークリッド距離を示す新変数を作成します。

大規模ファイルのクラスタ分析のオプション

図 35-4
[大規模ファイルのクラスタ分析: オプション] ダイアログ ボックス



統計量。使用できる統計量は、[初期クラスタ中心]、[分散分析表]、および [ケースに対するクラスタ情報] です。

- **初期クラスタ中心。**各クラスタに対する、変数の最初の平均値。デフォルトでは、クラスタの個数と同じ数の、十分に間隔のあるケースがデータから選択されます。初期クラスタ中心は分類の最初の一巡のために使われ、その後に更新されます。

- **分散分析表 (カテゴリ/カテゴリの回帰)。**各クラスタ変数に対して 1 変量 F 検定を含む分散分析表を表示します。F 検定は記述のみなので、結果として作成される確率を解釈することはできません。分散分析表は、すべてのケースが 1 つのクラスタに割り当てられない場合、表示されません。
- **ケースに対するクラスタ情報。**それぞれのケースについて、最終クラスタ割り当て、およびケースを分類するためのケースとクラスタ中心の間のユークリッド距離を表示します。さらに、最終クラスタ中心間のユークリッド距離も表示します。

欠損値。使用できるオプションは、[**リストごとに除外**] または [**ペアごとに除外**] です。

- **リストごとに除外。**クラスタ変数に対して欠損値のあるケースを分析から除外します。
- **ペアごとに除外。**非欠損値を持つすべての変数から計算された距離に基づいてケースがクラスタに割り当てられます。

QUICK CLUSTER コマンドの追加機能

大規模ファイルのクラスタ分析手続きでは、QUICK CLUSTER コマンド シンタックスを使用します。コマンド シンタックスを使用すると、次の作業も実行できます。

- 最初の *K*-ケース を初期クラスタ中心とします。これにより、推定にデータ パスの使用を避けることができます。
- 初期クラスタ中心をコマンド シンタックスの一部として指定します。
- 保存変数名を指定。

シンタックスの詳細は、『*Command Syntax Reference*』を参照してください。

章
36

ノンパラメトリック検定

[ノンパラメトリック検定] 手続きでは、基本的分布の形式に関する仮定を必要としない、数種類の検定を使用できます。

カイ 2 乗検定。変数がカテゴリに分類され、観測度数と期待度数の差に基づいて、カイ 2 乗統計量が計算されます。

2 項検定。2 分変数のカテゴリ内の観測度数と、2 項分布からの期待度数を比較します。

ラン検定。1 つの変数の 2 つの値の発生順序がランダムかどうかを検定します。

1 サンプルによる Kolmogorov-Smirnov 検定。変数の観測累積分布関数と、正規分布、一様分布、指数分布、またはポアソン分布のうち指定した理論分布を比較します。

2 個の独立サンプルの検定。1 つの変数で 2 つのグループに分けられたケースを比較します。使用できる検定は、Mann-Whitney の *U* 検定、2 サンプルの Kolmogorov-Smirnov 検定、Moses の外れ値反応の検定、および Wald-Wolfowitz のラン検定です。

2 個の対応サンプルの検定。2 個の変数の分布を比較します。使用できる検定は、Wilcoxon の符号付き順位検定、符号検定、および McNemar 検定です。

複数の独立サンプルの検定。1 つの変数で 2 つ以上のグループに分けられたケースを比較します。使用できる検定は、Kruskal-Wallis の検定、メディアン検定、および Jonckheere-Terpstra 検定です。

複数の対応サンプルの検定。2 個以上の変数の分布を比較します。使用できる検定は、Friedman の検定、Kendall の *W*、および Cochran の *Q* です。

上記すべての検定で、4 分位ならびに平均値、標準偏差、最小値、最大値、および欠損値を除いたケース数を計算できます。

カイ 2 乗検定

[カイ 2 乗検定] 手続きでは、変数をカテゴリに分類してカイ 2 乗統計量を計算します。この適合度検定では、それぞれのカテゴリの中の観測度数と期待度数を比較し、すべてのカテゴリに同じ比率の値が含まれているか、または各カテゴリに指定した比率の値が含まれているかを検定します。

例。カイ 2 乗検定を使用して、飴玉を入れた袋の中に青、茶、緑、橙、赤、黄色のそれぞれの飴玉が同じ比率で入っているかどうかを判別できます。また、それぞれの色の飴玉が、青 5%、茶 30%、緑 10%、橙 20%、赤 15%、黄色 15% の比率で入っているかどうかを調べることができます。

統計量。平均値、標準偏差、最小値、最大値、および 4 分位。欠損値を含まないケース数と欠損ケース数、およびパーセント; 各カテゴリの観測ケース数と期待ケース数; 残差; およびカイ 2 乗統計量。

データ。順序付けられている数値型カテゴリ変数、または順序付けられていない数値型カテゴリ変数を使用します (順序尺度または名義尺度)。文字型変数を数値型変数に変換するには、[変換] メニューの [値の再割り当て] を使用します。

仮定。ノンパラメトリック検定は、基本的な分布形状についての仮定を要求しません。データは、無作為のサンプルと仮定されます。それぞれのカテゴリの期待度数の最小値は 1 以上である必要があります。期待度数が 5未満のカテゴリは、全体の 20 %未満でなければなりません。

カイ 2 乗検定を行うには

▶ メニューから次の項目を選択します。

分析
ノンパラメトリック検定
カイ 2 乗...

図 36-1
[カイ 2 乗検定] ダイアログ ボックス

- 1 つ以上の検定変数を選択します。各変数は、個別の検定を作成します。
- 必要に応じて、[**オプション**] をクリックし、[統計量] で [記述統計量] および [4 分位]、[欠損値] で欠損値データの扱い方を指定できます。

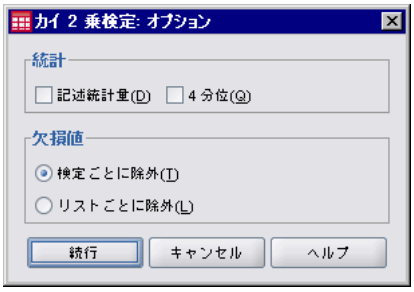
カイ 2 乗検定の期待範囲および期待度数

期待範囲。デフォルトでは、変数の各数値がカテゴリとして定義されます。指定された範囲内でカテゴリを定義するには、[**指定された範囲を使用**] をクリックして、[始] および [終] の各ボックスに整数を入力します。カテゴリは範囲内の各整数で定義され、範囲外の値のケースは除外されます。たとえば、[始] ボックスで 1、[終] ボックスで 4 を指定した場合、カイ 2 乗検定では、1 から 4 までの整数が使用されます。

期待値。デフォルトでは、すべてのカテゴリが同じ期待度数を持ちます。カテゴリは、ユーザー指定の期待比率を取ることができます。[**値**] をクリックし、検定変数の各カテゴリに 0 より大きい値を入力して [**追加**] をクリックします。値を追加するたびに、値リストの下端に表示されます。このボックスの値の順序は、検定変数のカテゴリ値の昇順に対応するので重要です。一覧の最初の値が検定変数のグループ値の最低値と対応し、最後の値が最高値と対応します。値の一覧の各値が合計され、値ごとにその合計値で割られ、対応するカテゴリの期待ケースの比率が計算されます。たとえば、値の一覧が 3、4、5、4 の場合、期待比率は、3/16、4/16、5/16、および 4/16 となります。

カイ 2 乗検定のオプション

図 36-2
[カイ 2 乗検定: オプション] ダイアログ ボックス



統計量。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **4 分位。**25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、すべての分析から除外されます。

NPAR TESTS コマンドの追加機能 (カイ 2 乗検定)

コマンド シンタックスを使用すると、次の作業も実行できます。

- 変数ごとに個別の最小値および最大値、または期待度数を指定 (CHISQUARE サブコマンドを使用)。
- 1 つの変数を異なる期待度数や異なる範囲で検定 (EXPECTED サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

2 項検定

[2 項検定] 手続きでは、2 分変数の 2 つのカテゴリの観測度数を、指定した確率パラメータの 2 項分布での期待度数と比較します。デフォルトでは、両方のグループの確率パラメータは 0.5 です。確率を変更するには、最初のグループの検定比率を入力します。2 番目のグループの確率は、1 から最初のグループに指定した比率を引いたものです。

例。10 セント硬貨を投げて裏か表かを決める場合、表の出る確率は 1/2 です。この仮説に基づき、硬貨を 40 回投げて、その結果を記録します (表か裏か)。2 項検定により、投げた回数の 3/4 が表であり、観測された有意水準が小さい (0.0027) ことがわかります。この結果は、表の出る確率が 1/2 でない、つまり、銀貨が歪んでいる可能性があることを示しています。

統計量。平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。

データ。検定変数は、数値型変数および 2 分変数でなければなりません。文字型変数を数値型変数に変換するには、[変換] メニューの [値の再割り当て] を使用します。**2 分変数**は、*Yes* または *No*、**真**または**偽**、0 または 1 のように、2 つの値だけを受け取る変数です。変数が 2 分変数でない場合は、分割点を指定する必要があります。分割点により、分割点を超過する値のケースが 1 つのグループに、残りのケースがもう 1 つのグループに割り当てられます。

仮定。ノンパラメトリック検定は、基本的な分布形状についての仮定を要求しません。データは、無作為のサンプルと仮定されます。

2 項検定を行うには

▶ メニューから次の項目を選択します。

分析
ノンパラメトリック検定
2 項...

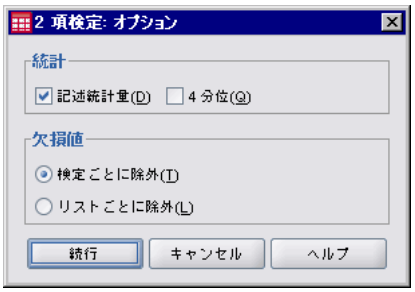
図 36-3
[2 項検定] ダイアログ ボックス

▶ 1 つ以上の数値検定変数を選択します。

▶ 必要に応じて、[**オプション**] をクリックし、[統計量] で [記述統計量] および [4 分位]、[欠損値] で欠損値データの扱い方を指定できます。

2 項検定のオプション

図 36-4
[2 項検定: オプション] ダイアログ ボックス



統計量。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **4 分位。**25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**検定される変数の欠損値のあるケースは、すべての分析から除外されます。

NPAR TESTS コマンドの追加機能 (2 項検定)

コマンド シンタックスを使用すると、次の作業も実行できます。

- 変数が 3 以上のカテゴリを持つ場合に、特定のグループを選択 (他の値を持つグループを除外) (BINOMIAL サブコマンドを使用)。
- 変数ごとに分割点または確率を指定 (BINOMIAL サブコマンドを使用)。
- 1 つの変数を異なる分割点または確率で検定 (EXPECTED サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

ラン検定

[ラン検定] 手続きは、変数の 2 つの値が現れる順序がランダムかどうかを検定します。ランは、同じ観測値の連続する長さです。ランが非常に多いかまたは非常に少ないサンプルは、そのサンプルがランダムではないことを示しています。

例。1 つの製品を購入するかどうかを確認するために、20 人を調査したとします。20 人のすべてが同じ性別であった場合は、仮定されたサンプルの無作為性が疑問になります。ラン検定を使用して、サンプルが無作為に抽出されたかどうかを判断できます。

統計量。平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。

データ。変数は数値型でなければなりません。文字型変数を数値型変数に変換するには、[変換] メニューの [値の再割り当て] を使用します。

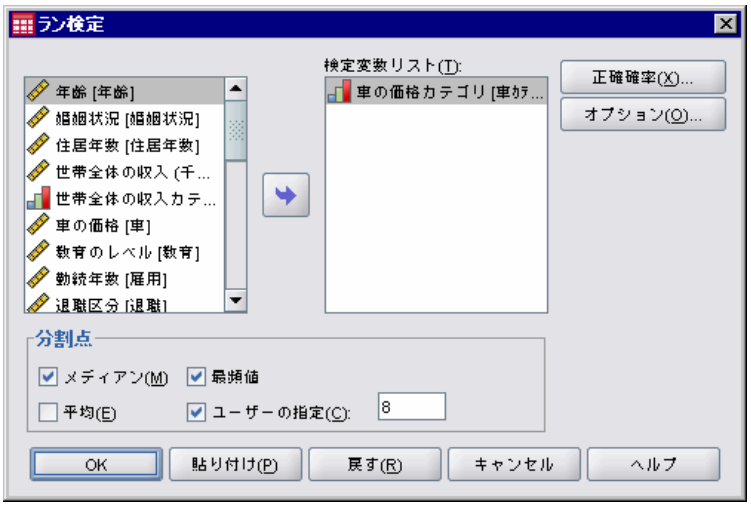
仮定。ノンパラメトリック検定は、基本的な分布形状についての仮定を要求しません。連続確率分布のサンプルを使用します。

ラン検定を行うには

▶ メニューから次の項目を選択します。

分析
ノンパラメトリック検定
ラン...

図 36-5
ユーザー指定の分割点を追加



▶ 1 つ以上の数値検定変数を選択します。

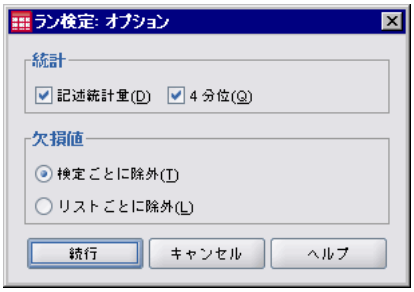
▶ 必要に応じて、[**オプション**] をクリックし、[統計量] で [記述統計量] および [4 分位]、[欠損値] で欠損値データの扱い方を指定できます。

ラン検定の分割点

分割点。選択した変数を 2 分するための分割点を指定します。観測した平均値、中央値、最頻値、またはユーザーが指定した値のどれかを使用できます。分割点未満の値のケースが 1 つのグループに割り当てられ、分割点以上の値のケースがもう 1 つのグループに割り当てられます。指定した分割点ごとに検定が実行されます。

ラン検定のオプション

図 36-6
[ラン検定: オプション] ダイアログ ボックス



統計量。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **4 分位。**25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、すべての分析から除外されます。

NPAR TESTS コマンドの追加機能 (ラン検定)

コマンド シンタックスを使用すると、次の作業も実行できます。

- 変数ごとに分割点を指定 (RUNS サブコマンドを使用)。
- 1 つの変数を異なる分割点で検定 (RUNS サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

1 サンプルによる Kolmogorov-Smirnov 検定

1 サンプルによる Kolmogorov-Smirnov 検定手続きは、1 つの変数で観測された累積分布関数を、正規、一様、ポアソン、または指数のいずれかの、指定した理論分布と比較します。Kolmogorov-Smirnov の *Z* は、観測された累積分布関数と理論的な累積分布関数との間の最大差 (絶対値) から計算されます。この適合度検定は、観測値が指定した分布から取られていると言えるかどうか検定します。

例。パラメトリック検定では、多くの場合、正規分布の変数が必要です。1 サンプルによる Kolmogorov-Smirnov 検定を使用すると、変数 (**収入**) が正規分布しているかを検定することができます。

統計量。平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。

データ。量的変数を使用します (測度の区間または比率レベル)。

仮定。Kolmogorov-Smirnov 検定では、検定分布のパラメータは事前に指定されているものと仮定しています。この手続きは、サンプルからパラメータを推定します。サンプルの平均値と標準偏差は正規分布のパラメータで、サンプルの最小値と最大値は一様な分布の範囲を定義し、サンプルの平均値はポアソン分布のパラメータです。検定の仮設分布からの逸脱の検出力は、大きく逓減する場合があります。推定のパラメータを持つ正規分布の検定については、調整済みの K-S Lilliefors の検定を検討してください([探索的分析] 手続きで使用可)。

1 サンプルによる Kolmogorov-Smirnov 検定を行うには

▶ メニューから次の項目を選択します。

分析
ノンパラメトリック検定
1 サンプルによる K-S 検定...

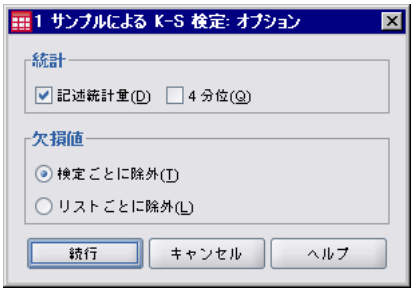
図 36-7
[1 サンプルによる Kolmogorov-Smirnov 検定] ダイアログ ボックス

▶ 1 つ以上の数値検定変数を選択します。各変数は、個別の検定を作成します。

▶ 必要に応じて、[**オプション**] をクリックし、[統計量] で [記述統計量] および [4 分位]、[欠損値] で欠損値データの扱い方を指定できます。

1 サンプルによる Kolmogorov-Smirnov 検定のオプション

図 36-8
[1 サンプルによる K-S 検定のオプション] ダイアログ ボックス



統計量。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **4 分位。**25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、すべての分析から除外されます。

NPAR TESTS コマンド追加機能 (1 サンプルによる Kolmogorov-Smirnov 検定)

コマンド シンタックスを使用すると、検定分布のパラメータを指定することもできます (K-S サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

2 個の独立サンプルの検定

[2 個の独立サンプルの検定] 手続きにより、1 つの変数に関して 2 つのグループのケースを比較できます。

例。より快適な装着感、外観の改善、および矯正期間の短縮などを目的とした歯列矯正器が新たに開発されました。新しい歯列矯正器の装着期間が従来の矯正器よりも短いかどうかを調査するため、古い矯正器および新しい矯正器のそれぞれを、無作為に選択した 10 名の子供たちに装着してもらいました。Mann-Whitney の *U* 検定により、古い矯正器を装着した子供たちに比べ、新しい矯正器を装着した子供たちの方が装着期間が短かったことがわかります。

統計量。平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。検定: Mann-Whitney の *U*、Moses の外れ値反応、Kolmogorov-Smirnov の *Z*、Wald-Wolfowitz のラン。

データ。順位付けできる数値変数を使用します。

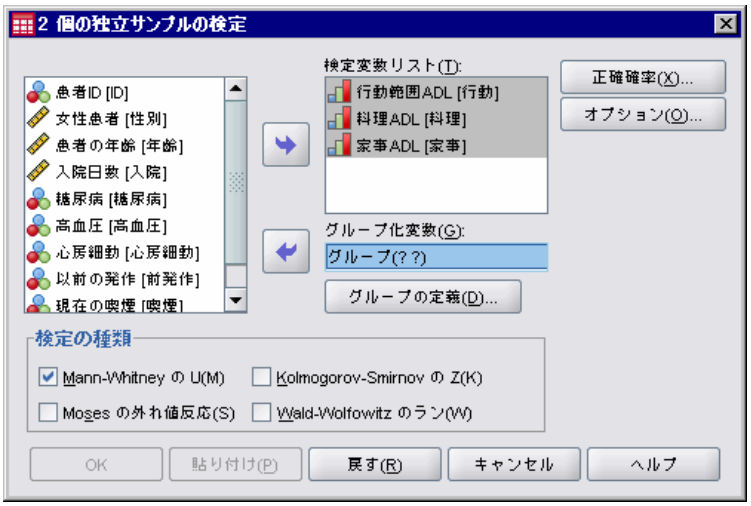
仮定。無作為に抽出された独立サンプルを使用します。[Mann-Whitney の *U*] 検定は、2 個のサンプルのそれぞれの母集団の分布の形が類似しているという仮説を検定します。

2 個の独立サンプルの検定を行うには

▶ メニューから次の項目を選択します。

分析
ノンパラメトリック検定
2 個の独立サンプルの検定...

図 36-9
[2 個の独立サンプルの検定] ダイアログ ボックス



▶ 1 つ以上の変数を選択します。

▶ ファイルを 2 つのグループまたはサンプルに分割するには、任意のグループ化変数を選択し、[**グループの定義**] をクリックします。

2 個の独立サンプルの検定の種類

検定の種類。2 個の独立サンプル (グループ) が同じ分布から取り出されたものであるかを検定するには、4 個の検定が使用できます。

[Mann-Whitney の U] は、[2 個の独立サンプルの検定] ダイアログ ボックスの中で最もよく使用されます。また、2 個のグループについては、Wilcoxon 検定および Kruskal-Wallis の H 検定と同等のものです。[Mann-Whitney の U] は、2

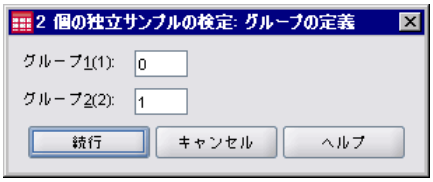
つのグループのサンプルが位置的に同じかどうかを検定します。両方のグループからの観測値を合わせて順位を付けます。観測値が同点の場合は、順位の平均値を割り当てます。同順位の数は、観測の合計数に対して相対的に小さくします。分布が位置的に同じである場合、ランクは 2 個のサンプル間で無作為に混合されます。グループ 1 の得点がグループ 2 の得点より多い場合の回数、またグループ 2 の得点がグループ 1 の得点より多い場合の回数が計算されます。その 2 つのうち小さい方の数字が Mann-Whitney の *U* 検定の統計となります。また、Wilcoxon の順位和検定の *W* 統計も表示されますが、こちらは 2 個の順位合計の小さい方となります。両方のサンプルの観測数が同じ場合、*W* は [2 個の独立サンプルの検定: グループの定義] ダイアログ ボックスに先に名前が載っているグループの順位和になります。

分布の位置と形の両方における差を検出するには、**[Kolmogorov-Smirnov の Z]** および **[Wald-Wolfowitz のラン]** が一般的に使われます。Kolmogorov-Smirnov の正規性の検定は、両方のサンプルの観測累積分布関数間の最大絶対差に基づいています。最大絶対差が有意である場合、2 個の分布は別々のものとして認識されます。Wald-Wolfowitz のラン検定は、両方のグループの観測値を合わせてランク付けします。2 つのサンプルが同じグループから取り出されたものである場合、2 つのグループは、ランク全体を通して無作為に分散されます。

[Moses の外れ値反応] は、実験変数が何人かの被験者をある方向に、また他の被験者を反対方向に影響を及ぼすことを前提としています。そして、対照群と比較した外れ値応答数を検定します。この検定は、対照グループのスパンに注目し、また、実験群の極値が対照グループと結合したときのスパンに及ぼす影響の尺度となります。対照群は、[2 個の独立サンプルの検定: グループの定義] ダイアログ ボックスの中のグループ 1 の値によって定義されます。両方のグループの観測が組み合わされ、ランク付けされます。対照グループのスパンは、そのグループにおけるランクの最大値と最小値の差に 1 を加えたものとなります。見込みの外れ値はスパンの範囲を誤りやすいので、対照グループの両端 5% は自動的にトリムされます。

2 個の独立サンプルの検定: グループの定義

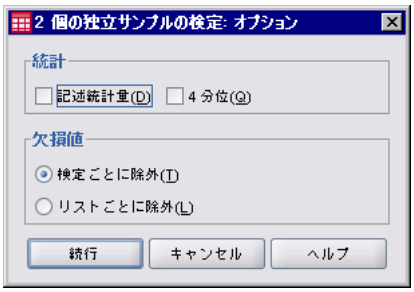
図 36-10
[2 個の独立サンプルの検定: グループの定義] ダイアログ ボックス



ファイルを 2 つのグループまたはサンプルに分割するには、最初の整数値をグループ 1 に、もう 1 つの整数値をグループ 2 に入力します。他の値のあるケースは分析から除外されます。

2 個の独立サンプルの検定: オプション

図 36-11
[2 個の独立サンプルの検定: オプション] ダイアログ ボックス



統計量。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **4 分位。**25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、すべての分析から除外されます。

NPAR TESTS コマンドの追加機能 (2 個の独立サンプルの検定)

コマンド シンタックス言語を使用すると、Moses の外れ値反応の検定に対してトリム化するケースの数を指定する (MOSES サブコマンドを使用) こともできます。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

2 個の対応サンプルの検定

[2 個の対応サンプルの検定] 手続きを使用して、2 つの変数の分布を比較できます。

例。一般に、家族が自分たちの住宅を売りに出す場合、提示価格で売れるでしょうか。Wilcoxon 検定を 10 戸の家のデータに当てはめてみると、7 戸が提示価格よりも安値で、1 戸が高値で、2 戸が提示価格で売れていることがわかります。

統計量。平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。検定: Wilcoxon の符号付き順位検定、符号検定、McNemar 検定。[正確確率検定] オプションが (Windows のオペレーティング システム上でのみ可能) インストールされているなら、周辺等質性検定も可能です。

データ。順位付けできる数値変数を使用します。

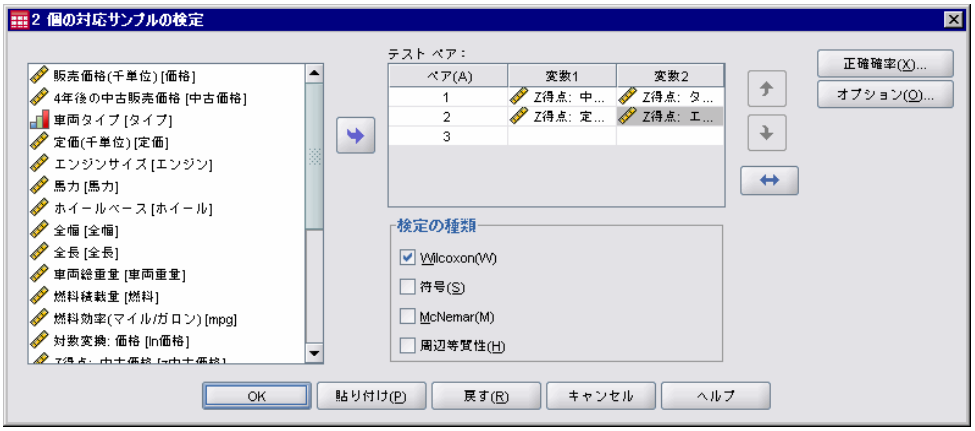
仮定。2 つの変数に対して特定の分布が仮定されていない場合でも、対応のある差の母集団の分布は対称であると仮定されます。

2 個の対応サンプルの検定を行うには

▶ メニューから次の項目を選択します。

分析
ノンパラメトリック検定
2 個の対応サンプルの検定...

図 36-12
[2 個の対応サンプルの検定] ダイアログ ボックス



▶ 1 つ以上の変数ペアを選択します。

2 個の対応サンプルの検定の種類

この項での検定では、2 つの対応した変数の分布が比較されます。データの種類によって、使用する検定が異なります。

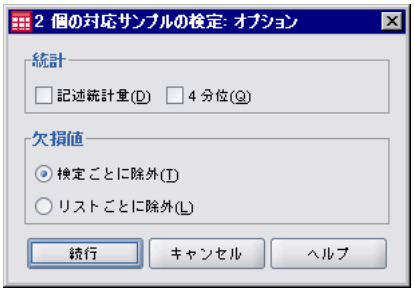
連続的なデータには、符号検定または Wilcoxon の符号付き順位検定を使用します。**符号検定**は、すべてのケースの 2 つの変数間の差を計算し、計算された差が正か負か、または同一かを分類します。2 つの変数の分布が類似している場合は、正と負の数に大きな差は見られません。**Wilcoxon の符号付き順位検定**は、ペア間の差の符号および量の両方に関する情報を考慮します。Wilcoxon の符号付き順位検定は、データに関する情報をより多く含んでいるため、符号検定より強力です。

2 進データには、**McNemar 検定**を使用します。この検定は、各被験者からあるイベントが発生する前と後の 2 回にわたって回答を得る反復測定に使用します。McNemar 検定を使用して、事前の応答比 (イベントが発生する前) と事後の応答比 (イベントが発生した後) が同等かどうかを決定します。この検定は、計画の前後での実験的介入によって発生する応答の変化を検出する場合に役立ちます。

カテゴリのデータには、**周辺等質性検定**を使用します。この検定は、McNemar の検定を 2 値反応から多値反応に拡張したものです。応答の変化を（カイ 2 乗分布を使用して）検定し、計画の前後で実験的介入によって発生する応答の変化を検出する場合に役立ちます。周辺等質性検定は、Exact Tests がインストールされている場合に限り使用できます。

2 個の対応サンプルの検定: オプション

図 36-13
[2 個の対応サンプルの検定: オプション] ダイアログ ボックス



統計量。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **4 分位。**25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、すべての分析から除外されます。

NPAR TESTS コマンドの追加機能（2 個の対応サンプル）

コマンド シンタックス言語を使用すると、指定した変数リストのすべての変数の組み合せについて検定することもできます。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

複数の独立サンプルの検定

[複数の独立サンプルの検定] 手続きでは、1 つの変数で複数のケースのグループを比較します。

例。3 種類の 100 ワット電球が燃え尽きるまでの平均時間に違いはあるのでしょうか。Kruskal-Wallis 一元配置分散分析から、3 種類の電球は平均寿命が違うことを理解できます。

統計量。平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。検定: Kruskal-Wallis の *H*、メディアン検定。

データ。順位付けできる数値変数を使用します。

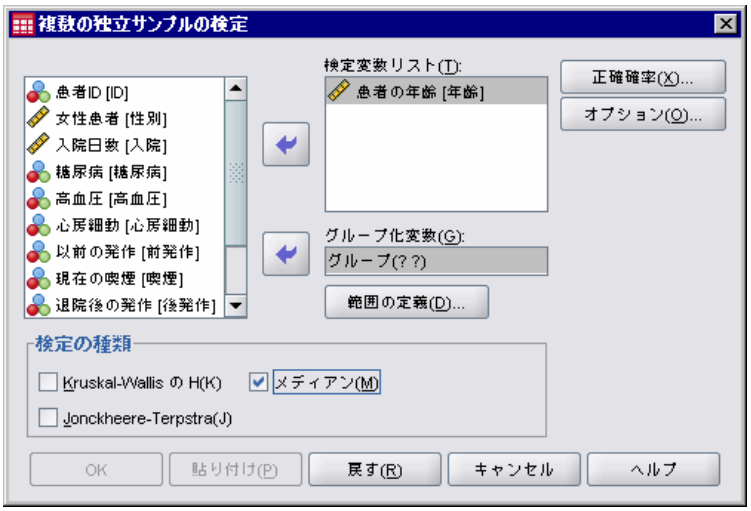
仮定。無作為に抽出された独立サンプルを使用します。Kruskal-Wallis の *H* 検定では、サンプルの分布の形が近似していることが検定されている必要があります。

複数の独立サンプルの検定を行うには

▶ メニューから次の項目を選択します。

分析
ノンパラメトリック検定
K 個の独立サンプルの検定...

図 36-14
メディアン検定の定義



▶ 1 つ以上の変数を選択します。

▶ グループ化変数を選択し、**[範囲の定義]** をクリックしてグループ化変数の最小と最大の整数値を指定します。

複数の独立サンプルの検定の検定の種類

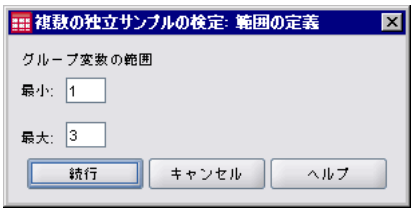
複数の独立サンプルが同じ母集団から発生しているかどうかを判断するには、3 つの検定を利用できます。Kruskal-Wallis の *H* 検定、メディアン検定、および Jonckheere-Terpstra の検定は、すべて複数の独立標本が同じ母集団から発生しているかどうかを検定します。

Mann-Whitney *U* 検定の拡張である **Kruskal-Wallis の H 検定**は、一元配置分散分析のノンパラメトリック版であり、分布の位置の差を検出します。**メディアン検定**はこれよりも一般的な検定ですが、それほど強力ではなく、分布の位置と形の違いを検出します。Kruskal-Wallis の *H* 検定およびメディアン検定は、標本が抽出される *k* 個の母集団の **事前の**順位付けがないものと仮定しています。

k 個の母集団の自然な**事前の**順位付け（昇順または降順）が**あるとき**には、**Jonckheere-Terpstra 検定**の方が効果的です。たとえば、*k* 個の母集団が *k* 段階の上昇温度を表す場合があるとします。異なる温度でも同じ応答分布を示すという仮説は、温度が上昇するにつれて応答の大きさが上昇するという対立仮説に対して検定されます。ここで、対立仮説が順序付けされるので、Jonckheere-Terpstra が最適の検定になります。Jonckheere-Terpstra 検定は、SPSS Exact Tests アドオン モジュールをインストールしている場合にだけ利用できます。

複数の独立サンプルの検定の範囲定義

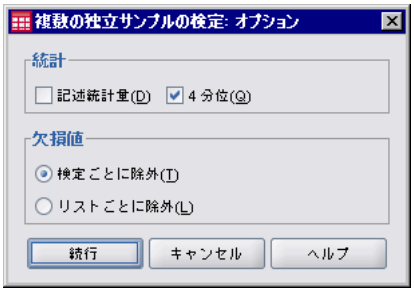
図 36-15
複数の独立サンプルの検定: 範囲の定義ダイアログ ボックス



範囲を定義するには、グループ化変数の最小と最大カテゴリに対応する **最小値**と**最大値**として整数値を入力します。上下限を超える値を含むケースは除外されます。たとえば、1 の最小値と 3 の最大値を指定すると、1 から 3 までの整数値だけが使用されます。最小値は最大値よりも小さくなければなりません。また、必ず両方の値を指定しなければなりません。

複数の独立サンプルの検定のオプション

図 36-16
[複数の独立サンプルの検定: オプション] ダイアログ ボックス



統計量。要約統計量の 1 つまたは両方を選択できます。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **4 分位。**25、50、および 75 パーセンタイルに対応する値を表示します。

欠損値。欠損値の処理を管理します。

- **検定ごとに除外。**いくつかの検定を指定するとき、各検定は欠損値がないか個別に評価されます。
- **リストごとに除外。**変数の欠損値のあるケースは、すべての分析から除外されます。

NPAR TESTS コマンドの追加機能 (K 個の独立サンプルの検定)

コマンド シンタックス言語で、メディアン検定に観測された中央値以外の値を指定することもできます (MEDIAN サブコマンドを使用)。

シンタックスの詳細は、『*Command Syntax Reference*』を参照してください。

複数の対応サンプルの検定

[複数の対応サンプルの検定] 手続きでは、複数の変数の分布を比較します。

例。世間の人が、医者、法律家、警察官、教師という職業に対して抱く権威の大きさに差はあるでしょうか?10 人に依頼して、権威の大きさについてこの 4 種類の職業を順位付けしてもらいました。Friedman の検定では、世間の人がこの 4 つの職業に対して抱く権威の大きさに差があるかどうかを示します。

統計量。平均値、標準偏差、最小値、最大値、欠損値を含まないケース数、および 4 分位。検定: Friedman、Kendall の *W*、および Cochran の *Q*。

データ。順位付けできる数値変数を使用します。

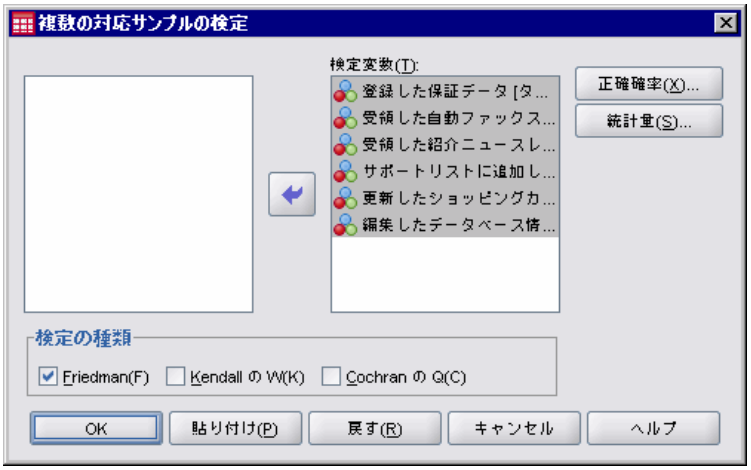
仮定。ノンパラメトリック検定は、基本的な分布形状についての仮定を要求しません。無作為に抽出された対応のあるサンプルを使用します。

複数の対応サンプルの検定を行うには

▶ メニューから次の項目を選択します。

分析
ノンパラメトリック検定
K 個の対応サンプルの検定...

図 36-17
Cochran を検定の種類として選択



▶ 2 つ以上の数値型検定変数を選択します。

複数の対応サンプルの検定の種類

複数の対応する変数の分布を比較するには、3 つの検定が利用できます。

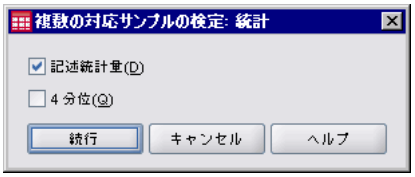
Friedman の検定は、1 サンプルの反復測定の分散分析やセルごとに 1 つの観測を持つ 2 元配置分散分析に対応するノンパラメトリック検定です。Friedman の検定は、*k* 個の対応変数は同じ母集団から発生するという帰無仮説です。各ケースに対して、*k* 個の変数は 1 から *k* に順位付けされます。検定の統計量は、それらの順位に基づきます。

Kendall の W は Friedman の統計の規準です。Kendall の *W* は、評価の一致度の尺度となっている一致係数として解釈することができます。各ケースは評価者で、各変数は評価されている品目または人です。各変数に対して、順位の合計が計算されます。Kendall の *W* は、0 (まったく一致していない) と 1 (完全な一致) の間の値を取ります。

Cochran の Q は、Friedman の検定とまったく同じですが、すべての応答が 2 値データのときに適用します。この検定は、McNemar 検定を *k* 個のサンプルの場合に拡張したものです。Cochran の *Q* は、複数の対応のある 2 分変数には同じ平均値があるという仮説を検定します。変数は同一の個体または対応のある個体上で測定されます。

複数の対応サンプルの検定の統計

図 36-18
[複数の対応サンプルの検定: 統計] ダイアログ ボックス



統計を選択します。

- **記述統計量。**平均値、標準偏差、最小値、最大値、および欠損値を含まないケース数を表示します。
- **4 分位。**25、50、および 75 パーセンタイルに対応する値を表示します。

NPAR TESTS コマンドの追加機能 (K 個の対応サンプルの検定)

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
37

多重回答の分析

多重 2 分および多重カテゴリ グループの分析には、2 つの手続きが利用できます。[多重回答の度数表] 手続きは、度数分布表を表示します。[多重回答のクロス集計表] 手続きは、2 次元および 3 次元のクロス集計表を表示します。どちらの手続きも、使用する前に多重回答グループを定義する必要があります。

例。この例は、市場調査での多重回答項目の使用を示しています。データは架空のものであり、実際のデータとして解釈してはなりません。ある航空会社が、競合する航空会社を評価するために、特定ルートを飛行する旅客を調査するとします。この例では、アメリカン エアラインが、シカゴとニューヨーク間での他の航空会社の利用状況と、航空会社の選択時にスケジュールとサービスが相対的にどの程度重要視されるかを調査します。客室乗務員が、各旅客の搭乗時に簡単なアンケートを手渡します。最初の質問は、「アメリカン、ユナイテッド、トランスワールド、US エア、その他のうち、最近 6 か月間にこの路線で少なくとも 1 回利用したすべての航空会社に丸を付けてください。」というものです。これは、旅客が複数の回答に丸を付けることができるので、多重回答質問です。しかし、変数は各ケースに 1 つの値しか設定できないので、この質問は直接コード化できません。回答をそれぞれの質問にマップするには、複数の変数を使用する必要があります。それには、2 つの方法があります。その 1 つは、それぞれの選択肢 (たとえば、アメリカン、ユナイテッド、トランスワールド、US エア、その他) に対応する変数を定義する方法です。旅客がユナイテッドに丸を付けた場合は、変数「**ユナイテッド**」に 1 のコードが割り当てられ、そうでない場合は 0 になります。これが、**多重 2 分法**による変数のマッピングです。回答をマップするもう 1 つの方法は、**多重カテゴリ法**というもので、利用した航空会社を指定するコードを使用して、質問に対して考えられる最大数の回答を予測し、同じ数の変数を設定する方法です。アンケートのサンプルを詳細に調べると、どの旅客も最近 6 か月間にこの路線で利用した航空会社は、最高 3 社であることがわかります。さらに、航空会社への規制緩和により、その他のカテゴリには 10 社の航空会社名が記されています。多重回答方法を使用し、3 つの変数を定義して、1 = **アメリカン**、2 = **ユナイテッド**、3 = **トランスワールド**、4 = **US エア**、5 = **デルタ**のようにそれぞれをコード化します。特定の旅客がアメリカンとトランスワールドに丸を付けると、最初の変数はコード 1 になり、2 番目の変数はコード 3 となり、3 番目の変数は欠損値コードになります。別の旅客がアメリカンとデルタに丸を付けた場合、。最初の変数はコードが 1 になり、2 番目の変数はコードが 5 になり、3 番目の変数は欠損値コードになります。一方、多重 2 分方法では、14 の別個の変数が使用されます。この調査ではどちらのマッピング方法も利用できますが、回答の分類によって選択する方法が決まります。

多重回答グループを定義

[多重回答グループを定義] 手続きでは、基本変数を多重 2 分グループおよび多重カテゴリ グループに分けて、度数分布表およびクロス集計表を求めることができます。多重回答グループは、20 個まで定義することができます。各グループには一意の名前を付ける必要があります。グループを削除するには、[多重回答グループ] ボックスの一覧でそのグループを選択し、**[除去]** をクリックします。グループを変更するには、同様に選択してグループの定義特性を修正し、**[変更]** をクリックします。

基本変数は、2 分またはカテゴリとしてコード化できます。2 分変数を使用するには、**[2 分]** を選択して多重 2 分グループを作成します。集計値に対して整数値を入力します。集計値が少なくとも 1 回発生する各変数は、多重 2 分割グループのカテゴリになります。**[カテゴリ]** を選択して、成分変数と同じ範囲の値を持つ多重カテゴリ グループを作成します。多重カテゴリ グループのカテゴリの範囲として、最小値と最大値の整数値を入力します。この手続きでは、成分変数全体で範囲内に含まれる各整数値を合計します。空のカテゴリは表にされません。

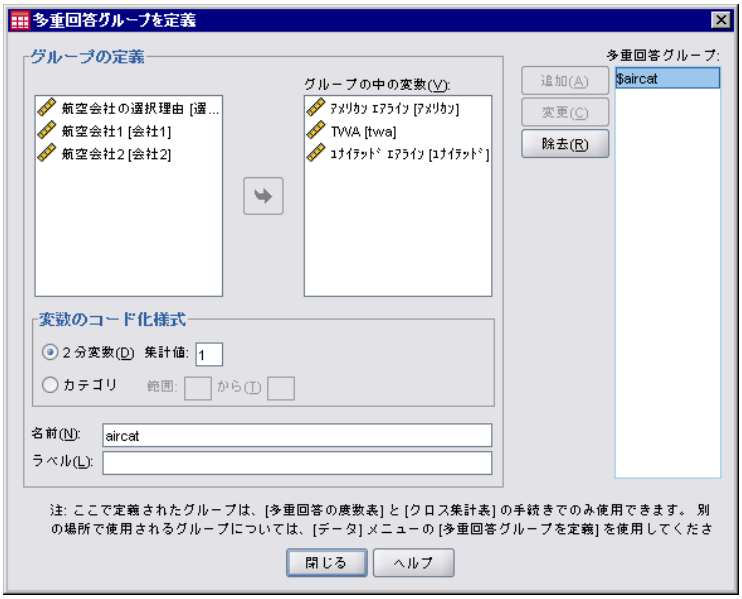
各多重回答グループには、7 文字までの一意の名前を割り当てる必要があります。この手続きでは、ユーザーが割り当てた名前の前にドル記号 ($) を付けます。ただし、予約名 casenum、sysmis、jdate、date、time、length、width は使用できません。多重回答グループの名前は、[多重回答] 手続き専用です。多重回答グループ名を他の手続きで参照することはできません。オプションとして、多重回答グループに記述的な変数ラベルを入力できます。ラベルの最大文字数は、40 文字です。

多重回答グループを定義するには

▶ メニューから次の項目を選択します。

分析
多重回答
グループの定義...

図 37-1
[多重回答グループを定義] ダイアログ ボックス



- 2 つ以上の変数を選択します。
- 変数が 2 分変数としてコード化されている場合には、数えておきたい値を指定します。変数がカテゴリとしてコード化されている場合には、カテゴリの範囲を定義します。
- 各多重回答グループに対して、一意の名前を入力します。
- 定義したグループのリストに多重回答グループを追加するには、[**追加**] をクリックします。

多重回答の度数表

[多重回答の度数表] 手続きは、多重回答グループの度数分布表を作成します。最初に、1 つ以上の多重回答グループを定義する必要があります (「多重回答グループを定義」を参照)。

多重 2 分グループでは、出力に示されるカテゴリ名は、そのグループ内の基本変数に定義した変数ラベルによって決められます。変数名を定義しない場合、変数名はラベルとして使用されます。多重カテゴリ グループでは、カテゴリ ラベルは、グループ内の最初の変数の値ラベルから付けられます。最初の変数に欠損しているカテゴリがグループ内の他の変数に存在する場合は、欠損カテゴリに値ラベルを定義します。

欠損値。欠損値のあるケースは、テーブルごとに除外されます。代わりに、次の中から 1 つまたは両方を選択できます。

- **2 分グループをリストごとに除外。**どの変数に対しても欠損値のあるケースを多重 2 分グループの表から除外します。これは、2 分グループとして定義された多重回答グループだけに適用します。デフォルトでは、ケースの成分変数のどれにも集計値がない場合、そのケースは多重 2 分グループに対して欠損とみなされます。すべての変数にではなく、いくつかの変数に欠損値があるケースは、1 つ以上の変数に集計値があれば、そのグループの集計表に含まれます。
- **カテゴリ グループをリストごとに除外。**どの変数に対しても欠損値のあるケースを多重カテゴリ グループの表から除外します。これは、カテゴリ グループとして定義されている多重回答グループにのみ適用します。デフォルトでは、ケースの成分変数のどれにも定義された範囲内に有効な値がない場合にだけ、多重カテゴリ グループに対して欠損とみなされます。

例。調査の質問から作成したそれぞれの変数は、基本変数です。多重回答項目を分析するには、変数を 2 種類の多重回答グループ、つまり、多重 2 分グループまたは多重カテゴリ グループのどちらか 1 つに結合する必要があります。たとえば、航空会社の調査で、3 つの航空会社 (アメリカン、ユナイテッド、トランスワールド) のうち最近 6 か月間に利用したのはどの会社かという質問をした場合、2 分変数を使用して **多重 2 分グループ**を定義すると、グループ内の 3 つの変数はそれぞれグループ化変数のカテゴリになります。3 つの航空会社の度数とパーセントは 1 つの度数分布表に表示されます。回答者が 3 つ以上の航空会社を言及しないとわかった場合は、各航空会社に 1 つずつ、計 3 つのコードを持った 2 つの変数を作成できます。**多重カテゴリ グループ**を定義する場合、値は基本変数で同じコードを一緒に追加して集計されます。結果としての値のグループは、基本変数のものと同じになります。たとえば、ユナイテッドとして答えた 30 の回答は、航空会社 1 として ユナイテッドを答えた 5 の回答と航空会社 2 として ユナイテッド を答えた 25 の合計です。3 つの航空会社の度数とパーセントは 1 つの度数分布表に表示されます。

統計量。度数を表示する度数分布表、回答のパーセント、ケースのパーセント、有効なケース数、および欠損ケースの数。

データ。多重回答グループを使用します。

仮定。度数とパーセントは、分布から取り込んだデータについて有効な記述をします。

関連手続き。[多重回答グループを定義] 手続きでは、多重回答グループを定義できます。

多重回答の度数分布表を作成するには

▶ メニューから次の項目を選択します。

分析
多重回答
度数分布表...

図 37-2
[多重回答の度数表] ダイアログ ボックス

- 1 つ以上の多重回答グループを選択します。

多重回答のクロス集計表

[多重回答のクロス集計表] 手続きは、多重回答グループ、基本変数、または結合をクロス集計します。ケースまたは回答に基づいたセル パーセントの表示、欠損値の処理の修正、または対応するクロス集計表の作成などを行うこともできます。最初に、1 つ以上の多重回答グループを定義する必要があります (「多重回答グループを定義するには」を参照)。

多重 2 分グループでは、出力に示されるカテゴリ名は、そのグループ内の基本変数に定義した変数ラベルによって決められます。変数名を定義しない場合、変数名はラベルとして使用されます。多重カテゴリ グループでは、カテゴリ ラベルは、グループ内の最初の変数の値ラベルから付けられます。最初の変数に欠損しているカテゴリがグループ内の他の変数に存在する場合は、欠損カテゴリに値ラベルを定義します。この手続きでは、1 行あたり最大 8 文字で、3 つの行にある列にカテゴリ ラベルを表示します。単語が分割しないようにするには、行と列の項目を反転するか、またはラベルを再定義します。

例。多重 2 分変数グループと多重カテゴリ グループの両方は、この手続きで他の変数とクロス集計できます。ある航空会社の調査で、旅客に次のように尋ねました。「次の航空会社のうち、最近 6 か月間に少なくとも 1 回利用したすべての航空会社に丸を付けてください (アメリカン、ユナイテッド、トランスワールド)。航空会社の選択時に、スケジュールとサービスのどちらがより重要ですか。1 つだけ選択してください。」。2 分変数または多重カテゴリとしてデータを入力してから、そのデータを 1 セットに結合すると、サービスまたはスケジュールに関連する質問の航空会社の選択肢をクロス集計できます。

統計量。セル、行、列および総数付き集計表、およびセル、行、列、および合計パーセント。セル パーセントは、ケースまたは回答に基づくことができます。

データ。多重回答グループまたは数値カテゴリ変数を使用します。

仮定。度数とパーセントは、分布から取り込んだデータについて有効な記述をします。

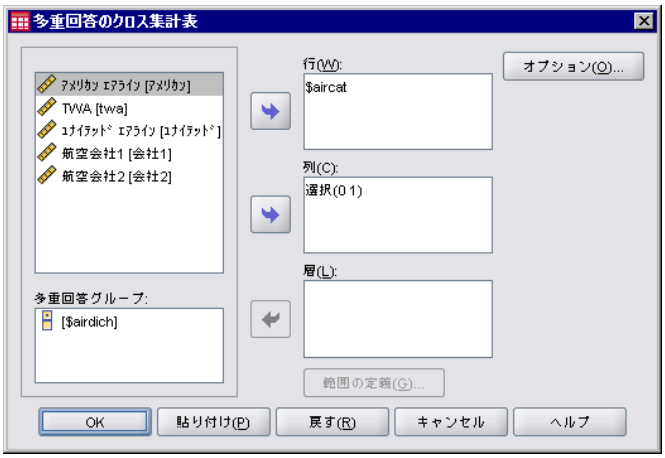
関連手続き。[多重回答グループを定義] 手続きでは、多重回答グループを定義できます。

多重回答のクロス集計表を作成するには

▶ メニューから次の項目を選択します。

分析
多重回答
クロス集計表...

図 37-3
[多重回答のクロス集計表] ダイアログ ボックス



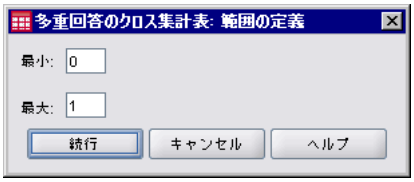
▶ 1 つ以上の数値変数を選択するか、クロス集計表の各次元に多重回答グループを選択します。

▶ 各基本変数の範囲を定義します。

オプションとして、管理変数の各カテゴリに対し、2 元配置のクロス集計表または多重回答グループを作成できます。[層] リストに対して 1 つ以上の項目を選択します。

多重回答のクロス集計表の範囲の定義

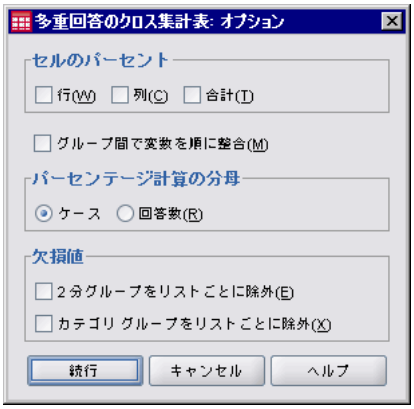
図 37-4
[多重回答のクロス集計表: 範囲の定義] ダイアログ ボックス



値の範囲は、クロス集計表のどの基本変数にも定義する必要があります。集計したい整数の最小カテゴリ値と最大カテゴリ値を入力します。範囲外のカテゴリは、分析から除外されます。範囲内の値は整数であるとみなされます (整数以外は切り捨てられます)。

多重回答のクロス集計表のオプション

図 37-5
[多重回答のクロス集計表: オプション] ダイアログ ボックス



セルのパーセント。セル度数は常に表示されます。選択すると、行パーセント、列パーセント、および二元配置 (全体) パーセントを表示できます。

パーセンテージ計算の分母。セルのパーセントをケース (または回答) に基づいて設定できます。これは、多重カテゴリ グループ全体で変数の結合を選択すると利用できません。セル パーセントは、回答に基づくこともできます。多重 2 分割グループでは、回答数はケース全体でカウントされる値の数に等しくなります。多重カテゴリ グループでは、回答数は定義された範囲内の値数です。

欠損値。次の中から 1 つまたは両方を選択することができます。

- **2 分グループをリストごとに除外。**どの変数に対しても欠損値のあるケースを多重 2 分グループの表から除外します。これは、2 分グループとして定義された多重回答グループだけに適用します。デフォルトでは、ケースの成分変数のどれにも集計値がない場合、そのケースは多重 2 分グループに対し

て欠損とみなされます。すべての変数にではなく、いくつかの変数に欠損値があるケースは、1 つ以上の変数に集計値があれば、そのグループの集計表に含まれます。

- **カテゴリ グループをリストごとに除外。**どの変数に対しても欠損値のあるケースを多重カテゴリ グループの表から除外します。これは、カテゴリ グループとして定義されている多重回答グループにのみ適用します。デフォルトでは、ケースの成分変数のどれにも定義された範囲内に有効な値がない場合にだけ、多重カテゴリ グループに対して欠損とみなされます。

デフォルトでは、2 つの多重カテゴリ グループをクロス集計すると、最初のグループ内にある各変数を 2 番目のグループ内にある各変数と集計して各セルの度数を合計するので、いくつかの回答が 1 つのテーブルで複数回表示されることもあります。次のオプションを選択できます。

グループ間で変数を順に整合。1 番目のグループにある最初の変数を 2 番目のグループにある最初の変数とペアにし、以後同じようにします。このオプションを選択すると、セル パーセントは、回答者ではなく回答数に基づきます。対応は、多重 2 分グループまたは基本変数には利用できません。

MULT RESPONSE コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 5 次元までのクロス集計表の作成 (BY サブコマンドを使用)。
- 値ラベルの抑制を含む出力書式オプションの変更 (FORMAT サブコマンドを使用)。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
38

結果の報告

ケースのリストおよび記述統計量は、データを調査、提示するための基本ツールです。データ エディタまたは [ケースの要約] 手続きを使ってケースのリストを、[度数分布表] 手続きを使って度数と記述統計量を、および [グループの平均] 手続きを使ってグループ別の母集団の統計量を得ることができます。それぞれの手続きでは、情報を明確にするためにレイアウトされた書式を使用します。情報を別の書式で表示したい場合は、[報告書: 行の集計] および [報告書: 列の集計] ダイアログ ボックスで、データ表示を必要に応じて指定できます。

報告書の行の集計

[報告書: 行の集計] 手続きでは、各種の要約統計量が行に割り当てられる報告書を作成します。要約統計量を伴うまたは伴わないでケースのリストを表示することもできます。

例。ある小売店のチェーンを持っている会社が、給与、勤務年数、および各従業員が働く店舗と部門を含む従業員情報を記録しているとします。各店舗、部門、および各店舗内の部門の要約統計量 (たとえば、平均給与) を、店舗と部門別 (ブレーク変数) に分類した個別の従業員情報 (リスト) を示した報告書を作成できます。

データ列。必要とするケースのリストまたは要約統計量のための変数を表示して、データ列の表示書式を指定します。

ブレーク列。報告書をグループに分割する任意のブレーク変数を表示して、要約統計量およびブレーク列の表示書式を指定します。ブレーク変数が複数ある場合は、リスト内の先行するブレーク変数のカテゴリ内にある各ブレーク変数の各カテゴリに対して個別のグループができます。ブレーク変数は、ケースを一定数の有意なカテゴリに分割するカテゴリ変数でなければなりません。各ブレーク変数の個々の値は、データ列の左に、個別の列でソートされて表示されます。

報告書。全体の要約統計量、欠損値の表示、ページ番号、および表題を含む報告書全体の特徴を指定します。

ケースの表示。それぞれのケースに対して、データ列変数の実際の値 (または値ラベル) を表示します。これにより、集計報告書よりもかなり長くなるリスト報告書が作成されます。

プレビュー。報告書の最初のページのみを表示します。このオプションは、報告書全体を処理しないで報告書の書式を下見するときに有効です。

データは並べ替え済み。ブレーク変数のある報告書では、報告書を作成する前に、データ ファイルをブレーク変数値でソートする必要があります。データ ファイルがブレーク変数で既に並び替えられている場合、このオプションを選択すると処理時間を節約できます。このオプションは、プレビュー報告書を実行した後には特に有効です。

集計報告書の行の集計を行うには

- メニューから次の項目を選択します。

  分析
  報告書
  報告書: 行の集計...

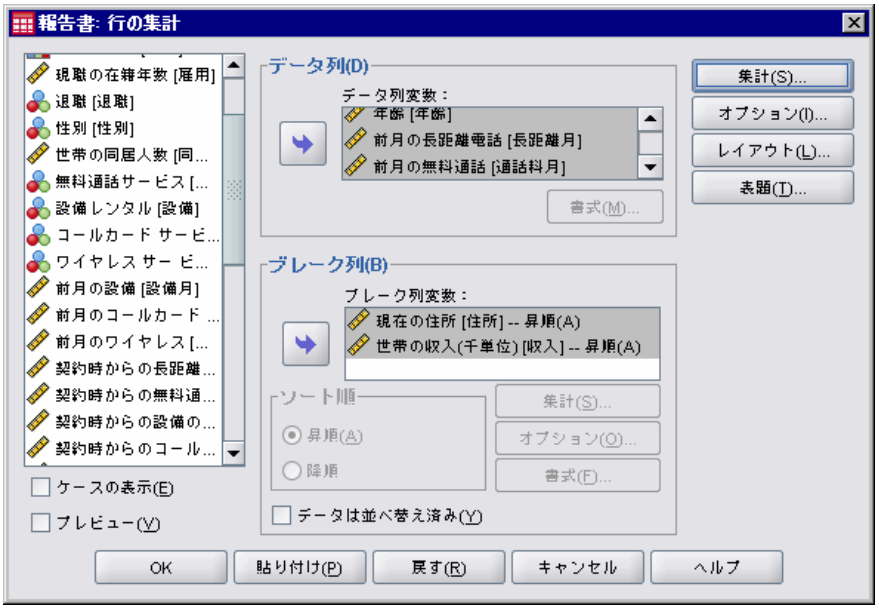
- データ列に 1 つ以上の変数を選択します。選択された各変数に対して報告書内に 1 つの列が生成されます。

- サブグループで並べ替えられ、表示された報告書に対しては、ブレーク列に 1 つ以上の変数を選択します。

- ブレーク変数で定義されたサブグループの要約統計量を表示する報告書に対しては、ブレーク列変数リスト内のブレーク変数を選択して、ブレーク列グループ内の [**集計**] をクリックして集計項目を指定します。

- 全体の要約統計量を表示する報告書では、[**集計**] をクリックして集計項目を指定します。

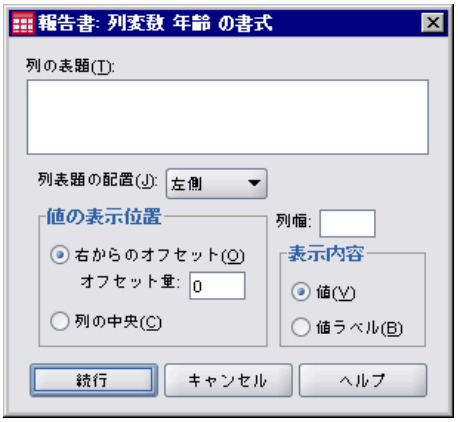
図 38-1
[報告書: 行の集計] ダイアログ ボックス



報告書のデータ列/ブレーク列の書式

[書式] ダイアログ ボックスでは、列の表題、列幅、テキストの位置合わせ、およびデータ値または値ラベルの表示を指定します。データ列の [書式] では、報告書ページの右側のデータ列の書式を指定します。ブレーク列の [書式] では、左側のブレーク列の書式を指定します。

図 38-2
[報告書: 列の集計] ダイアログ ボックス



列の表題。選択した変数に対して、列の表題を指定します。長い表題は、列内で自動的に折り返されます。表題を折り返したい場所に手動で改行を挿入するには、Enter キーを押します。

値の表示位置。選択した変数に対して、列内のデータ値または値ラベルの位置合わせを指定します。値またはラベルの位置合わせは、列見出しの位置合わせに影響しません。指定した文字数だけ列の内容をインデントすることもできますが、内容を中央に設定することもできます。

表示内容。選択した変数に対して、データ値または定義済みの値ラベルのどちらを表示するかを指定します。データ値は、常に定義済みの値ラベルを持っていない値に対して表示されます。(列集計報告書のデータ列では使用できません)。

報告書の集計行/最終集計行

2 つの集計行ダイアログ ボックスでは、ブレーク グループおよび全体の報告書に対して要約統計量の表示を指定します。[集計行] ダイアログ ボックスは、ブレーク変数で定義された各カテゴリのサブグループ統計量を制御します。[最終集計行] ダイアログ ボックスでは、報告書の終わりに表示される全体の統計量を指定します。

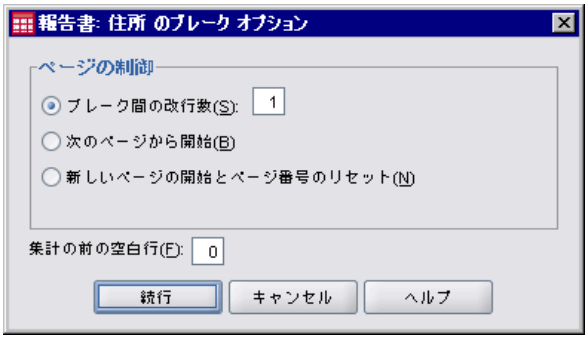
図 38-3
[報告書: 最終集計行] ダイアログ ボックス

使用できる要約統計量としては、合計、平均値、最小値、最大値、ケースの数、指定した値の上下でのケースのパーセント、指定した範囲値内のケースのパーセント、標準偏差、尖度、分散、および歪度があります。

報告書のブレーク オプション

ブレーク列のオプションでは、ブレーク カテゴリ情報の改行とページ番号を指定します。

図 38-4
[報告書: ブレークのオプション] ダイアログ ボックス



ページの制御。選択したブレーク変数のカテゴリに対して改行およびページ番号を指定します。ブレーク カテゴリ間の空白行数を指定することもできますが、新しいページから各ブレーク カテゴリを開始することもできます。

集計の前の空白行。ブレーク カテゴリ ラベルまたはデータと要約統計量間の空白行の数を指定します。これは、ブレーク カテゴリに個々のケース リストと要約統計量の両方を含む報告書には特に有効です。この報告書では、ケース リストと要約統計量間にスペースを挿入できます。

報告書のオプション

[報告書: オプション] ダイアログ ボックスでは、欠損値の扱いと表示および報告書のページ番号を指定します。

図 38-5
[報告書: オプション] ダイアログ ボックス

欠損値のあるケースをリストごとに除外。報告書の変数のどれかに対して欠損値のあるケースを (報告書から) 除去します。

欠損値の表示。データ ファイル内の欠損値を表す記号を指定できます。記号は 1 文字だけ使用でき、この記号を使用して、**システム欠損値**と **ユーザー欠損値**の両方を表すことができます。

開始ページ番号。報告書の最初のページにページ番号を指定できます。

報告書のレイアウト

[報告書: レイアウト] ダイアログ ボックスでは、各報告書ページの幅と長さ、ページ上での報告書の配置、および空白行とラベルの挿入を指定します。

図 38-6
[報告書: レイアウト] ダイアログ ボックス

ページの設計。行単位（上端と下端）と文字単位（左と右）で表されるページ余白および余白内での報告書の位置合わせを指定します。

ページ表題と脚注。ページ表題と脚注を報告書の本文から分離する行数を指定します。

ブレーク列。ブレーク列の表示を指定します。複数のブレーク変数を指定すると、変数を個別の列または最初の列に入れることができます。最初の列にすべてのブレーク変数を置くと、幅の狭い報告書が作成されます。

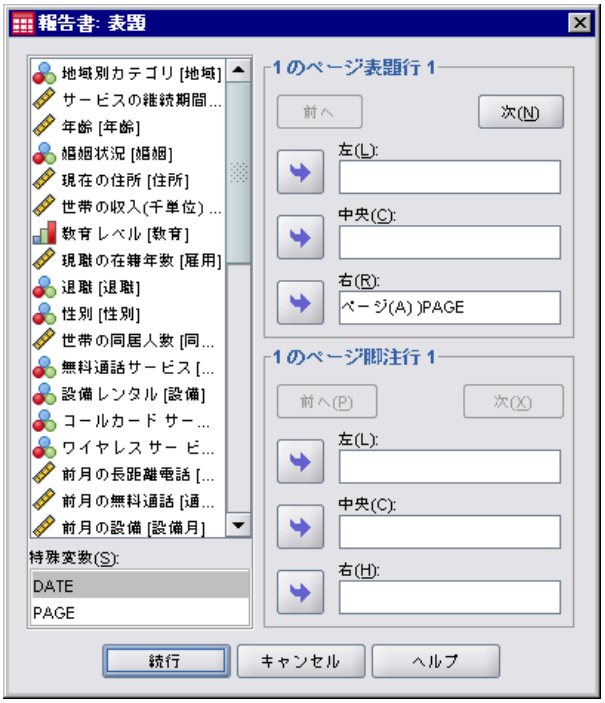
列の表題。表題に付ける下線、表題と報告書の本文との間の空白行、および列表題の縦の位置合わせを指定します。

データ列の行とブレーク列のラベル。各ブレーク カテゴリの最初のブレーク ラベルに関してデータ列情報（データ値または要約統計量）の配置を指定します。データ列情報の最初の行は、ブレーク カテゴリ ラベルと同じ行、またはブレーク カテゴリ ラベル行から指定した行数後のどちらかから開始できます。（列集計報告書では使用できません）。

報告書の表題

[報告書: 表題] ダイアログ ボックスでは、報告書の表題と脚注の内容および配置を指定します。各行の内容を左詰め、中央、または右詰めにして、10 行までのページ表題、および 10 行までのページ脚注を指定できます。

図 38-7
[報告書: 表題] ダイアログ ボックス



変数を表題または脚注に挿入した場合は、現在の値ラベルまたは変数の値が表題または脚注に表示されます。表題では、ページの始めに変数の値に対応する値ラベルが表示されます。脚注では、ページの終わりに変数の値に対応する値ラベルが表示されます。値ラベルがない場合は、実際の値が表示されます。

特殊変数。特殊変数 DATE と PAGE を使用して、現在の日付またはページ番号を報告書の表題または脚注の任意の行に挿入できます。データ ファイルに DATE または PAGE という変数がある場合は、この変数を報告書の表題または脚注で使用できません。

報告書の列の集計

[報告書: 列の集計] では、個々の要約統計量が個別の列に表示される要約統計量を作成します。

例。ある小売店のチェーンを持っている会社が、給与、勤務年数、および各従業員が働く部門を含む従業員情報を記録しているとします。各部門に給与の要約統計量 (たとえば、平均値、最小値、最大値) を示す報告書を作成できます。

データ列。要約統計量を必要とする変数を表示し、各変数に表示される表示書式と要約統計量を指定します。

ブレーク列。報告書をグループに分割する任意のブレーク変数を表示し、ブレーク列の表示書式を指定します。ブレーク変数が複数ある場合は、リスト内の先行するブレーク変数のカテゴリ内にある各ブレーク変数の各カテゴリに対して個別のグループができます。ブレーク変数は、ケースを一定数の有意なカテゴリに分割するカテゴリ変数でなければなりません。

報告書。欠損値、ページ番号、および表題を含む報告書全体の特徴を指定します。

プレビュー。報告書の最初のページのみを表示します。このオプションは、報告書全体を処理しないで報告書の書式を下見するときに有効です。

データは並べ替え済み。ブレーク変数のある報告書では、報告書を作成する前に、データ ファイルをブレーク変数値でソートする必要があります。データ ファイルがブレーク変数で既に並び替えられている場合、このオプションを選択すると処理時間を節約できます。このオプションは、プレビュー報告書を実行した後には特に有効です。

集計報告書の列の集計を行うには

- メニューから次の項目を選択します。

  分析
  報告書
  報告書: 列の集計...

- データ列に 1 つ以上の変数を選択します。選択された各変数に対して報告書内に 1 つの列が生成されます。

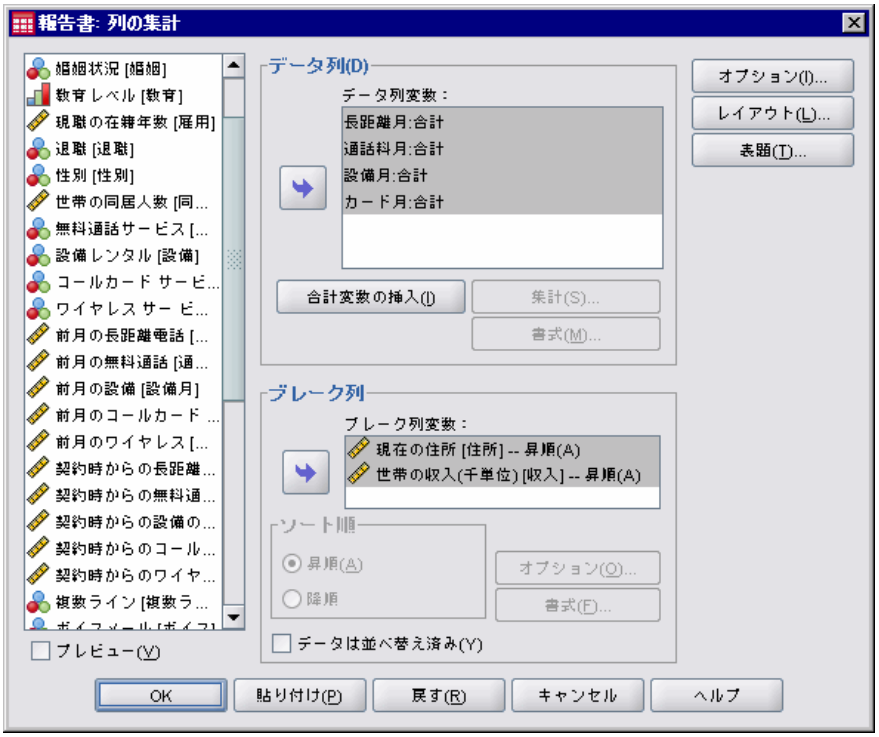
- 変数の集計項目を変更するには、データ列変数の変数を選択して [**集計**] をクリックします。

- 1 つの変数に複数の集計項目を使用するには、右側の変数リストで変数を選択して [データ列変数] ボックスに移動し、それに希望する集計項目の 1 つを割り当てます。この操作を希望する集計項目の数だけ繰り返します。

- 既存の列の合計値、平均値、比率、またはその他の関数を含む列を表示するには、[**列合計の挿入**] をクリックします。これで、**合計**という変数がデータ列リストに挿入されます。

- サブグループで並べ替えられ、表示された報告書に対しては、ブレーク列に 1 つ以上の変数を選択します。

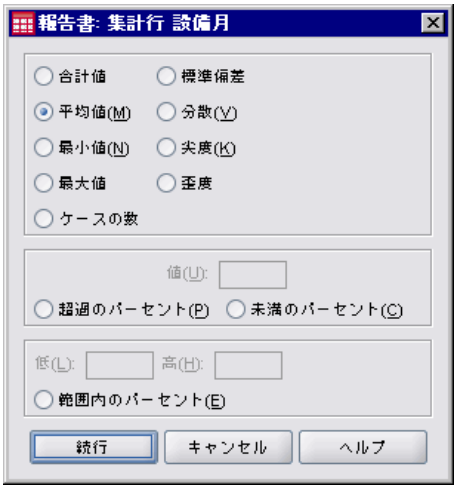
図 38-8
[報告書: 列の集計] ダイアログ ボックス



データ列の集計関数

集計行ダイアログ ボックスでは、選択されたデータ列変数に表示される要約統計量を指定します。

図 38-9
[報告書: 最終集計行] ダイアログ ボックス



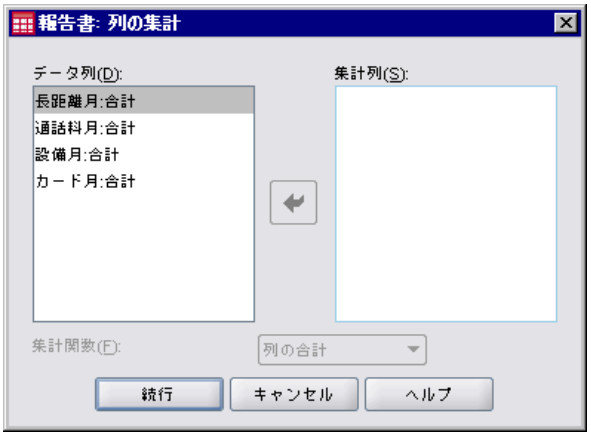
使用できる要約統計量としては、合計、平均値、最小値、最大値、ケースの数、指定した値の上下でのケースのパーセント、指定した範囲値内のケースのパーセント、標準偏差、尖度、分散、および歪度があります。

合計列のデータ列集計

列の集計では、2 つ以上のデータ列を集計する合計の要約統計量を指定します。

使用できる合計の要約統計量としては、列の合計、列の平均値、最小値、最大値、2 つの列間の値の差、1 つの列の値を別の列の値で割った商、および列の値を掛けた積があります。

図 38-10
[報告書: 列の集計] ダイアログ ボックス



列の合計。合計の列は、[集計列] リストにある列の合計です。

列の平均。合計の列は、[集計列] リストにある列の平均です。

列の最小。合計の列は、[集計列] リストにある列の最小値です。

列の最大。合計の列は、[集計列] リストにある列の最大値です。

1 列目 –2 列目。合計の列は、[集計列] リストにある列の差です。[集計列] リストには 2 つの列が含まれている必要があります。

1 列目/2 列目。合計の列は、[集計列] リストにある列の商です。[集計列] リストには 2 つの列が含まれている必要があります。

1 列目/2 列目の %。合計の列は、[集計列] リストにある第 2 列に対する最初の列のパーセントです。[集計列] リストには 2 つの列が含まれている必要があります。

列の積。合計の列は、[集計列] リストにある列の積です。

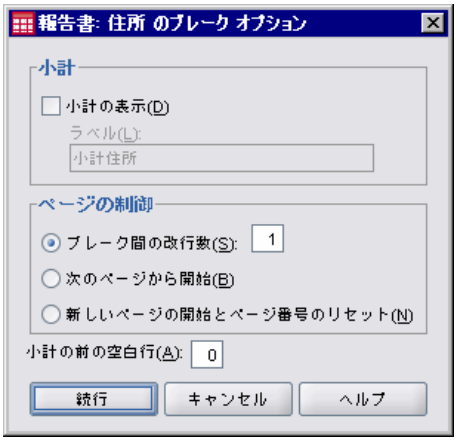
報告書の列の書式

[報告書: 列の集計] ダイアログ ボックスのデータ列とブレーク列の書式オプションは、[報告書: 行の集計] ダイアログ ボックスで説明したものと同じです。

報告書の列の集計でのブレーク列のオプション

ブレーク列のオプションでは、ブレーク カテゴリの小計表示、改行、およびページ番号を指定します。

図 38-11
[報告書: ブレークのオプション] ダイアログ ボックス



小計。ブレーク カテゴリの小計表示を指定します。

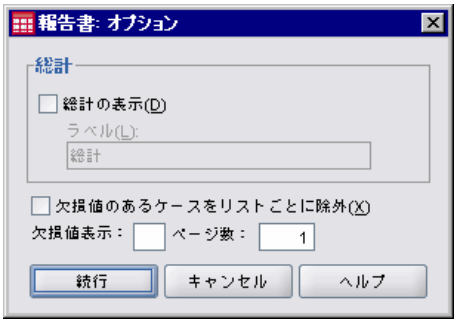
ページの制御。選択したブレーク変数のカテゴリに対して改行およびページ番号を指定します。ブレーク カテゴリ間の空白行数を指定することもできますが、新しいページから各ブレーク カテゴリを開始することもできます。

小計の前の空白行。ブレーク カテゴリ データと小計間の空白行の数を指定します。

報告書の列の集計のオプション

オプションで、列集計報告書内の総計の表示、欠損値の表示、およびページ番号を指定します。

図 38-12
[報告書: オプション] ダイアログ ボックス



総計。列の下端に表示される各列の総計の表示とラベル付けを行います。

欠損値。報告書から欠損値を除外できますが、報告書内の欠損値を示すために 1 つの文字を選択することもできます。

報告書の列の集計のレイアウト

[報告書: 列の集計] ダイアログ ボックスの [報告書: レイアウト] オプションは、[報告書: 行の集計] ダイアログ ボックスで説明したものと同じです。

REPORT コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 1 つの集計行の列に別の集計関数を表示。
- データ列変数以外の変数に対して、または集計関数の各種結合 (複合関数) に対して、集計行をデータ列に挿入。
- 中央値、最頻値、度数、およびパーセントを集計関数として使用。
- 要約統計量の表示書式をさらに正確に指定。
- 報告書内のさまざまな点に空白行を挿入。
- 報告書のリストで *n* 番目のケースごとに空白行を挿入。

REPORT シンタックスは複雑なため、ダイアログ ボックスから作成した報告書に近づけるためにシンタックスで新しい報告書を構築するときには、対応するシンタックスをコピーして貼り付け、そのシンタックスを修正して希望する正確な報告書を作成するようにしてください。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
39

信頼性分析

信頼性分析では、測定尺度とスケールを構成する項目の特性を調査できます。[信頼性分析] 手続きは、一般に使用される多数の尺度の信頼性の測度を計算し、尺度の各項目間の関係についての情報を提供します。級内相関係数は、評価者間の信頼性推定値の計算に役立ちます。

例。質問表は消費者の満足度の測定に役立つでしょうか。信頼性分析を使用して、質問表の項目間の関連の度合いを決定し、反復性の全体的な指標や尺度全体の内部的一貫性を求めることができます。また、問題のある項目を識別して尺度から除去できます。

統計量。各変数と尺度の記述統計、項目での要約統計、項目間の相関と共分散、信頼性推定値、分散分析表、級内相関係数、Hotelling の T^2 乗、Tukey の加法性の検定。

モデル。次の信頼性のモデルを使用できます。

- **アルファ (Cronbach)。**これは、項目間の平均相関に基づく内部一貫性のモデルです。
- **折半法。**このモデルは尺度を 2 つに分割し、分割した部分間の相関を調べます。
- **Guttman。**このモデルは真の信頼性の Guttman の下限を計算します。
- **平行モデル。**このモデルでは、繰り返しを通じてすべての項目の分散と誤差分散が等しいと仮定します。
- **厳密平行モデル。**このモデルは平行モデルの仮定に加えて、すべての項目で平均が等しいと仮定します。

データ。データは 2 分データ、順序データ、または間隔データでなければなりません。また、数値でコード化されている必要があります。

仮定。観測値は独立している必要があります。また、誤差は項目間で相関してはなりません。項目の各組は 2 変量正規分布でなければなりません。スケールは加法的でなければならず、その結果、各項目は合計得点に対して線型になります。

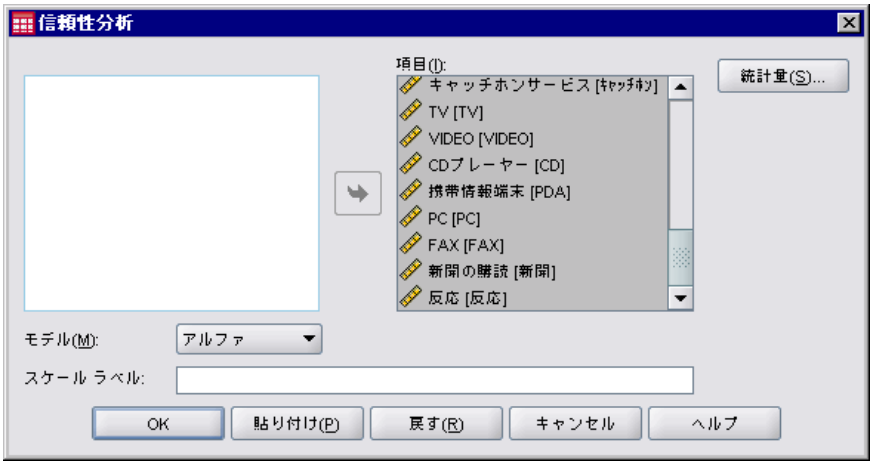
関連手続き。尺度項目の各次元を詳細に調べる (項目得点のパターンの説明に複数の構成が必要かどうかを調べるため) 場合には、因子分析または多次元尺度法を使用します。等質な変数グループを識別するには、変数を寄せ集める階層的クラスタ分析を使用できます。

信頼性分析を行うには

▶ メニューから次の項目を選択します。

分析
尺度
信頼性分析...

図 39-1
[信頼性分析] ダイアログ ボックス

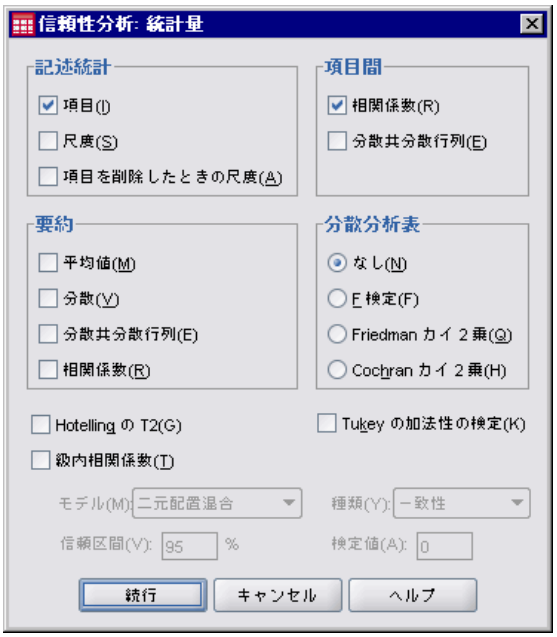


▶ 加法的尺度の成分の候補として、複数の変数を選択します。

▶ [モデル] ボックスの一覧からモデルを選択します。

信頼性分析の統計

図 39-2
[信頼性分析: 統計量] ダイアログ ボックス



スケールと項目を表すさまざまな統計を選択できます。デフォルトで求める統計は、ケース数、項目数、および次の信頼性推定値です。

- **アルファ モデル。**アルファ係数。2 分変数の場合には、Kuder-Richradson の公式 20 (KR20) 係数と同じです。
- **折半法のモデル。**形式間の相関、Guttman の折半法の信頼性、Spearman-Brown の信頼性 (等長と不等長) および 2 つの部分ごとのアルファ係数。
- **Guttman のモデル。**信頼係数ラムダ 1 からラムダ 6 まで。
- **平行モデルと厳密平行モデル。**モデルの適合度の検定、誤差分散の推定値、共通分散の推定値、および真の分散の推定値、共通項目間相関の推定値、信頼性の推定値、信頼性の不偏推定値。

記述統計。すべてのケースで、尺度または項目の記述統計量を求めます。

- **項目。**すべてのケースで、項目の記述統計量を求めます。
- **尺度。**尺度の記述統計量を求めます。
- **項目を削除したときの尺度。**各項目を他の項目から構成される尺度に比較した要約統計量を表示します。この統計量には、項目が尺度から削除された場合の尺度の平均と分散、項目と他の項目から構成される尺度との間の相関、および項目が尺度から削除された場合の Cronbach のアルファが含まれます。

要約。尺度のすべての項目で項目分布の記述統計量を求めます。

- **平均値 (信頼性分析)。**項目平均値の要約統計量。最小項目平均値、最大項目平均値、平均項目平均値、項目平均値の範囲と分散、および最大項目平均値と最小項目平均値の比率が表示されます。
- **分散。**項目分散の要約統計量。項目分散の最小、最大、平均、項目分散の範囲、分散、および項目分散の最大と最小の比率が表示されます。
- **共分散 (ピボットテーブル 回帰)。**項目間共分散の要約統計量。最小項目間共分散、最大項目間共分散、平均項目間共分散、項目間共分散の範囲と分散、および最大項目間共分散と最小項目間共分散の比率が表示されます。
- **相関係数。**項目間相関の要約統計量。最小項目間相関、最大項目間相関、平均項目間相関、項目間相関の範囲と分散、および最大項目間相関と最小項目間相関の比率が表示されます。

項目間。項目間の相関行列または分散共分散行列を作成できます。

分散分析表。平均が等しい検定を行います。

- **F 検定。**反復測定 (経時的変化) による分散分析表を表示します。
- **Friedman カイ 2 乗。**Friedman のカイ 2 乗および Kendall の一致係数を表示します。このオプションは、ランクの形式になっているデータに適しています。カイ 2 乗検定は、分散分析表の通常の F 検定に代わるものです。
- **Cochran カイ 2 乗。**Cochran の Q を表示します。このオプションは、2 分データに適しています。Q 統計量は、分散分析表の通常の F 統計量に代わるものです。

Hotelling の T2。スケール上の全項目が同じ平均を持つという帰無仮説の多変量の検定を行うことができます。

Tukey の加法性の検定。項目間に (乗法的) 交互作用がないという仮説の検定を行うことができます。

級内相関係数。ケース内の一致度または一致値を測定します。

- **モデル。**級内相関係数を計算するモデルを選択します。使用できるモデルは二元配置混合、二元配置変量、および一元配置変量です。人的効果が変量で項目効果が固定である場合は **[二元配置混合]** を、人的効果と項目効果の両方が変量である場合は **[二元配置変量]** を選択するか、人的効果が変量である場合は **[一元配置変量]** を選択します。
- **種類。**指標の種類を選択します。使用できる種類は一致性と絶対一致です。
- **信頼区間。**信頼区間の係数を指定します。デフォルト値は 95% です。
- **検定値。**仮説検定の係数の仮説値を指定します。この値が観測値と比較される値です。デフォルト値は 0 です。

RELIABILITY コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 相関行列の読み込み、および分析。

- その後の分析で使用する相関行列の書き込み。
- 折半法で等分割以外の指定。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
40

多次元尺度法

多次元尺度法は、対象またはケース間の一群の距離測度の中に構造を見つける手法です。この手法では、観測値を概念的な空間（通常は 2 次元または 3 次元）の中の特定の場所に割り当てます。この場合、空間内の点の間の距離が、指定した非類似度にできるだけ近接するようにします。多くの場合、データを深く理解するために、この概念的な空間の次元を解釈して使用します。

変数が客観的に測定されている場合には、多次元尺度法をデータ削減のための手法として使用できます（必要であれば、[多次元尺度法] 手続きは多変量データの距離を計算します）。また、多次元尺度法は、対象または概念の非類似度を主観的に評価する場合にも適用できます。また、[多次元尺度法] 手続きでは、複数の評価者や複数のアンケートの回答者からのデータがある場合に、それらの複数のデータから非類似度を処理することもできます。

例。人はどのようにして異なる車の関連に気づくでしょうか。車の異なった作りとモデルの類似度を評価した回答者からのデータがある場合は、多次元尺度法を使用して、消費者の認識を表す次元を識別できます。たとえば、価格と車の大きさが 2 次元空間を定義する場合があります。この空間は、回答者によって示される類似度を説明しています。

統計量。各モデル: データ行列、最適尺度化データ行列、S ストレス（Young）、ストレス（Kruskal）、RSQ、刺激座標、平均ストレス、各刺激の RSQ（RMDS モデル）。個人差（INDSCAL）モデル: 被験者の重み付けと各被験者の異常指標。複合多次元尺度法モデルの各行列: ストレスと各刺激の RSQ。作図: 刺激座標（2 次元または 3 次元）、差異と距離の散布図。

データ。データが非類似度データの場合、すべての非類似度は数量でなければならず、さらに同じ測定基準で測定される必要があります。データが多変量データの場合、変数は数量データ、2 値データ、または度数データとなります。尺度の違いは解に影響することもあるので、変数の尺度は重要な問題です。変数の尺度に大きな違いがある（たとえば、ある変数はドルで測定され、別の変数は年で測定されている）場合には、変数の標準化を検討する必要があります（[多次元尺度法] 手続きでは、標準化は自動的に実行できます）。

仮定。[多次元尺度法] 手続きは比較的、分布の仮定を必要としません。結果を正しく計算するには、[多次元尺度法: オプション] ダイアログ ボックスで適切な尺度（[順序データ]、[区間データ]、または [比データ]）を選択する必要があります。

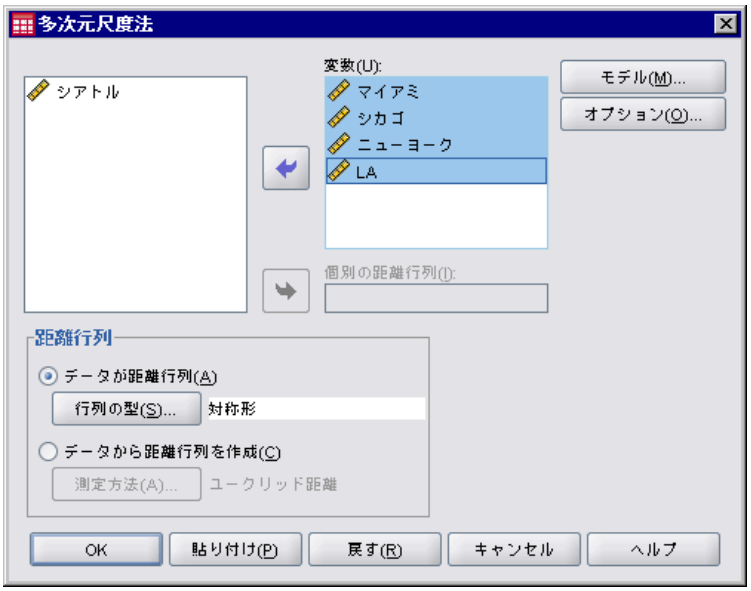
関連手続き。データの削減を目的としている場合で、特に変数が数量のときには、代わりに因子分析を使用できます。類似ケースのグループを識別する場合には、多次元尺度分析を階層クラスタ分析または大規模ファイルのクラスタ分析で補強することを検討します。

多次元尺度法分析を行うには

▶ メニューから次の項目を選択します。

分析
尺度
多次元尺度法...

図 40-1
[多次元尺度法] ダイアログ ボックス



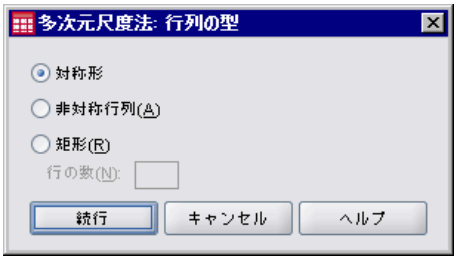
▶ [距離行列] で、**[データが距離行列]** または **[データから距離行列を作成]** をクリックします。

▶ データが距離行列の場合には、分析では少なくとも 4 個の数値型変数を選択する必要があります。(また、**[行列の型]** をクリックして、距離行列の形式を指定することもできます。)

▶ まず、データから距離行列を作成し、その後に分析を行う場合には、最低 1 つの数値型変数を選択します。(また、**[測定方法]** をクリックして、希望の距離測度の種類を指定することもできます。)グループ化変数 (数値型または文字型のどちらも可能) の各カテゴリの行列を個別に作成するには、[個別の距離行列] ボックスにその変数を移動します。

多次元尺度法のデータの形式

図 40-2
[多次元尺度法: 型] ダイアログ ボックス

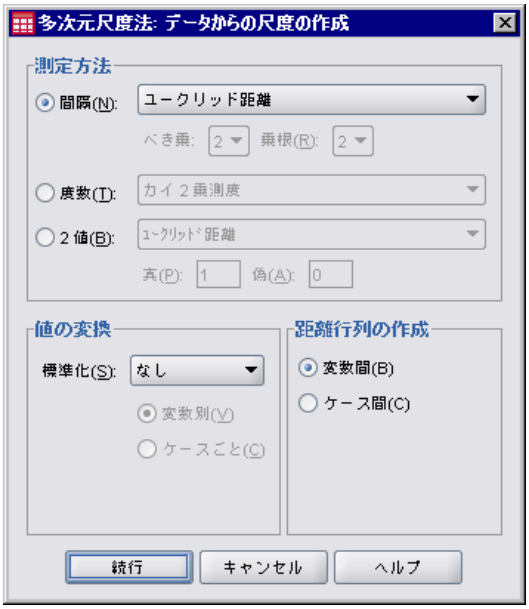


作業データセットが対象グループ内の距離、または 2 組の対象グループ間の距離を表している場合に、正しい結果を得るには、データ行列の形式を指定する必要があります。

注: [モデル] ダイアログ ボックスの [条件付け] で [行] を指定している場合、[**対称行列**] は選択できません。

多次元尺度法の尺度の作成

図 40-3
[多次元尺度法: データからの尺度の作成] ダイアログ ボックス



多次元尺度法では、非類似度データを使用して尺度解を作成します。データが多変量データ（測定された複数の変数の値）の場合は、非類似度データを作成し、多次元尺度法の解を計算します。データから非類似度の測度を作成する方法を詳細に指定できます。

測定方法。分析で使用する非類似度の測定方法を指定できます。データの種類に応じて [測定方法] のオプションを 1 つクリックします。選択したオプションのドロップダウン リストから測定方法を 1 つ選択します。使用可能な選択肢は次のとおりです。

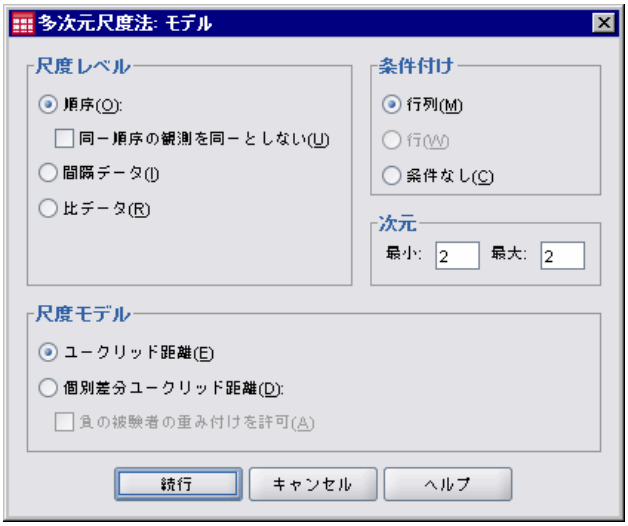
- **間隔。**[ユークリッド距離]、[平方ユークリッド距離]、[Chebychev]、[都市ブロック]、[Minkowski]、または [カスタマイズ]。
- **度数。**[カイ 2 乗測度] または [ファイ 2 乗測度]。
- **2 値。**[ユークリッド距離]、[平方ユークリッド距離]、[サイズの差異]、[パターンの差異]、[分散]、または [Lance と Williams]。

距離行列の作成。分析の単位を選択します。[変数間] または [ケース間] をクリックします。

値の変換。たとえば、変数がまったく異なるスケールで測定されている場合などで、値を標準化してから近接を計算します (2 値データには適用できません)。[標準化] ドロップダウン リストの一覧から標準化の方法を選択します。標準化を行う必要がない場合は、[なし] を選択します。

多次元尺度法のモデル

図 40-4
[多次元尺度法: モデル] ダイアログ ボックス



データとモデル自体の状態を指定することで、多次元尺度法のモデルを正確に推定します。

尺度レベル。データのレベルを指定できます。[順序データ]、[区間データ]、または [比データ] のいずれか 1 つをクリックします。変数が順序データの場合には、**[同一順序の観測を同一としない]** チェック ボックスをオンにすると、変数が連続変数として扱われるため、同順位のデータ (ケースが違っても同じ値を持つデータ) が最適に分析されます。

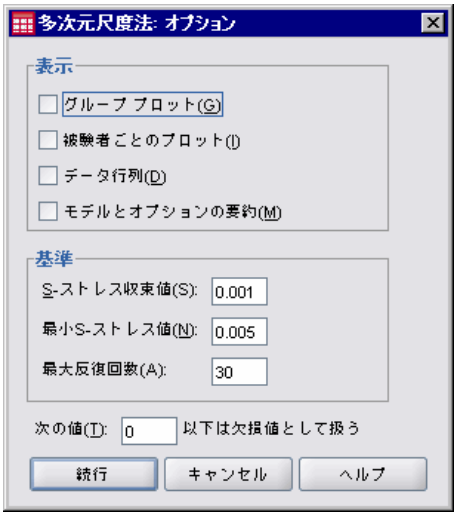
条件付け。意味のある比較を指定します。[行列]、[行]、[条件なし] のいずれか 1 つをクリックします。

次元。尺度法で求める解の次元数を指定します。範囲内の各数値に対して、解が 1 つ計算されます。ここでは、1 から 6 までの整数を指定できます。最小値の 1 を指定できるのは、尺度モデルに **[ユークリッド距離]** を指定した場合だけです。解を 1 つだけ求めるには、最小値と最大値に同じ数を指定します。

尺度モデル。尺度法を実行するための仮定を指定します。[ユークリッド距離] または [個人差ユークリッド距離] (INDSCAL とも言います) のどちらかをクリックします。個人差ユークリッド距離モデルでは、**[負の被験者の重み付けを許可]** を選択できます (ただし、そのデータに適当な場合)。

多次元尺度法のオプション

図 40-5
[多次元尺度法: オプション] ダイアログ ボックス



多次元尺度法でオプションを指定できます。

表示。各種の出力形式を指定できます。[グループ プロット]、[被験者ごとのプロット]、[データ行列]、[モデルとオプションの要約] が使用できます。

基準。反復を停止させる基準を指定します。デフォルトを変更するには、**[S-ストレス収束値]**、**[最小 S-ストレス値]**、および **[最大反復回数]** に値を入力します。

距離が n 以下は欠損値として扱う。指定した値以下の距離が分析から除外されます。

ALSCAL コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 多次元尺度法の文献で、ASCAL、AINDS、および GEMSCAL とよばれる追加モデルの使用。
- 区間データと比データの多項式変換の実行。
- (距離ではなく) 類似度を順序データで分析。
- 名義データの分析。
- さまざまな座標や重み付き行列をファイルに保存し、分析で再使用。
- 多次元展開の制約。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

章
41

比率統計量

比率統計量手続きでは、2 つのスケール変数間の比率を示す要約統計量の総合的なリストを作成できます。

出力は、グループ化変数値により昇順または降順に並べ替えることができます。比率統計量報告書は出力内で抑制することができ、結果は外部ファイルとして保存することができます。

例。家屋の査定額と販売額間の比率は、5 か国間でほぼ同じなのでしょうか。出力結果から、比率の分布は国によって大きく異なることがわかります。

統計量。ユーザー指定の範囲または中央値比率に含まれたパーセンテージで計算された中央値、平均値、重み付き平均値、信頼区間、散らばり係数 (COD)、中央値中心化の変動係数、平均値中心化の変動係数、価格関連格差 (PRD)、標準偏差、平均絶対偏差 (AAD)、範囲、最小値と最大値、および濃度インデックス。

データ。数値コードまたは文字列を使用してグループ化変数をコード化します (名義または順序尺度)。

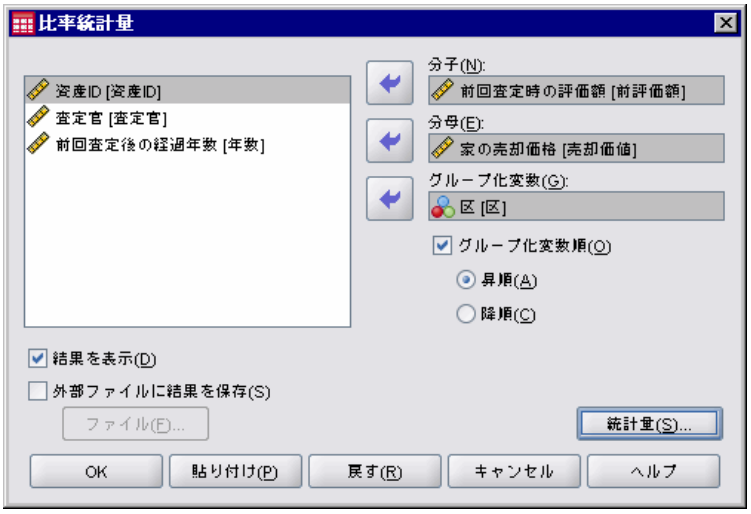
仮定。分子と分母の比率を定義する変数は、正の値を取るスケール変数でなければなりません。

比率統計を行うには

▶ メニューから次の項目を選択します。

分析
記述統計
比率...

図 41-1
[比率統計量] ダイアログ ボックス



- ▶ 分子変数を選択します。
- ▶ 分母変数を選択します。

  次のオプションが選択できます。

  - ■ グループ化変数を選択し、結果のグループ順序を指定します。
  - ■ ビューアに結果を表示するかどうかを選択します。
  - ■ 後で使用できるように結果を外部ファイルに保存するかどうかを選択し、結果を保存するファイル名を指定します。

比率統計量

図 41-2
[比率統計量] ダイアログ ボックス

中心傾向。中心傾向の測度は、比率の分布を示す統計量です。

- **中央値。**この中央値より小さい比率の数と、大きい比率の数は同じになります。
- **平均値。**比率の合計の結果を比率の総数で割った値です。
- **重み付き平均値。**分子の平均値を分母の平均値で割った値です。重み付き平均値は、分母によって重みを付けられた比率の平均値にもなります。
- **信頼区間。**平均値、中央値、および重み付き平均値の信頼区間を表示します(要求がある場合)。信頼水準として、0 以上 100 未満の値を指定します。

散らばり。観測値での変動量、または広がり量を測定する統計量です。

- AAD。平均絶対偏差は、中央値の比率の絶対偏差合計を比率の総数で割った値です。
- COD。散らばり係数は、平均絶対偏差を中央値のパーセントで表した値です。
- PRD。価格関連格差 (回帰指数として知られている) は、平均値を加重平均で割ったものです。
- **中央値 中心化の** COV。中央値を基準にした変動係数は、中央値からの偏差のルート平均平方を中央値のパーセントで表した値です。
- **平均値 中心化の** COV。平均値中心化の変動係数は、標準偏差を平均値のパーセントで表した値です。

- **標準偏差。**標準偏差とは、平均値の比率の平方偏差合計を比率の総数から 1 引いた数で割り、正の平方根を取った値です。
- **範囲。**範囲とは、最大比から最小比を引いた値のことです。
- **最小値。**最少値とは、最少比のことです。
- **最大値。**最大値とは、最大比のことです。

集中指数です。集中係数は、区間に入る比のパーセントを表します。濃度係数は、次の 2 種類の方法で計算できます。

- **範囲間の比率。**この区間は、区間の最低値および最高値を指定することにより明示的に定義されます。最低値と最高値の比率を入力し、**[追加]** をクリックして区間を指定します。
- **範囲内の比率。**この区間は、中央値のパーセントを指定することにより黙示的に定義されます。0 ～ 100 までの値を入力し、**[追加]** をクリックします。区間の下限は (1 – 0.01 × 値) × 中央値に等しく、上限は (1 + 0.01 × 値) × 中央値に等しくなります。

章
42

ROC 曲線

被験者を分類する 2 つのカテゴリに 1 つの変数がある場合に、この手続きを使用して分類方法のパフォーマンスを評価できます。

例。銀行は、ローン返済の履行、不履行について顧客を正しく分類することに関心があります。したがって、そのための意思決定を行う特別な方法が開発されています。ROC 曲線を使用して、こうした方法のパフォーマンスを評価できます。

統計量。信頼区間を持つ ROC 曲線の下の領域と ROC 曲線の座標点。作図: ROC 曲線。

方法。ROC 曲線の下の面積の推定値は、ノンパラメトリックまたは 2 負指数モデルを使用するパラメトリックのいずれかで計算できます。

データ。検定変数は量的変数です。ほとんどの検定変数は、判別分析またはロジスティック回帰または任意のスケール上の得点 (これは一方のカテゴリあるいは他方のカテゴリに被験者を分類する評価者の「確信の強さ」を示しています) からの確率です。状態変数はどのような種類のものでもよく、被験者が属する真のカテゴリを示しています。状態変数の値は、どのカテゴリを **正**とみなすべきかを示します。

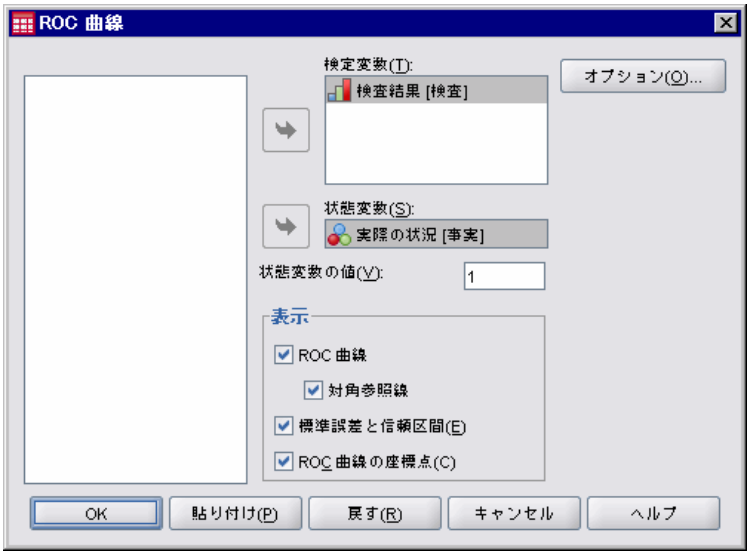
仮定。評価者のスケール上の数値が大きければ大きいほど、被験者が一方のカテゴリに属すという確信が強まり、このスケール上の数値が小さければ小さいほど、被験者が他方のカテゴリに属すという確信が強まると仮定します。ユーザーは、どの方向が **正**であるかを指定する必要があります。さらに、各被験者が属している **真**のカテゴリがわかっていると仮定します。

ROC 曲線を作成するには

▶ メニューから次の項目を選択します。

分析
ROC 曲線...

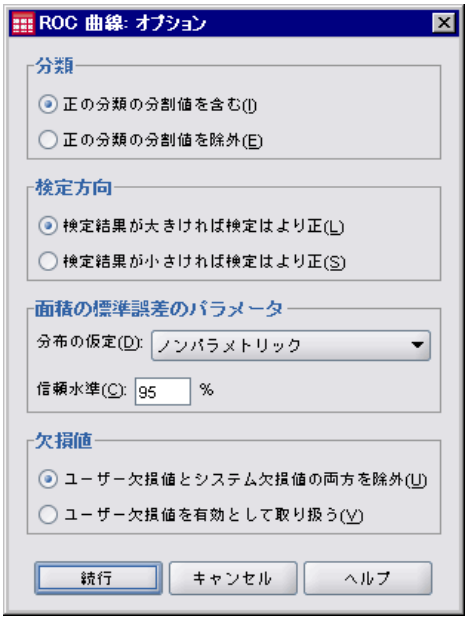
図 42-1
[ROC 曲線] ダイアログ ボックス



- ▶ 1 つ以上の検定変数を選択します。
- ▶ 1 つの状態変数を選択します。
- ▶ 状態変数の **正**の値を識別します。

ROC 曲線のオプション

図 42-2
[ROC 曲線: オプション] ダイアログ ボックス



ROC 分析に次のオプションを指定できます。

分類。正の分類を実行するときに、分割点の値を含めるか除外するかを指定できます。この設定は、現在、出力に影響しません。

検定方向。正のカテゴリに対するスケールの方向を指定できます。

面積の標準誤差のパラメータ。曲線の下の面の標準誤差を推定する方法を指定できます。使用できる方法は、ノンパラメトリックと 2 負指数です。さらに、信頼区間の水準を設定することもできます。有効な範囲は 50.1～99.9% です。

欠損値。欠損値の扱い方を指定できます。

章
43

図表作成機能の概要

高解像度図表およびプロットを作成するには、[グラフ] メニューに用意されている手続きと、[分析] メニューに用意されている各種の手続きを使用します。この章では、図表作成機能の概要を述べます。

図表の作成および編集

図表を作成する前に、データ エディタにデータを表示する必要があります。データは、データ エディタに直接入力したり、以前に保存されたデータ ファイルを開いたり、スプレッドシート、タブ区切りのデータ ファイル、データベース ファイルなどを読み込んだりしてエディタに取り込むことができます。[ヘルプ] メニューの [チュートリアル] をクリックすると、図表を作成および修正する例が表示されます。また、オンライン ヘルプには、あらゆる種類の図表の作成および修正方法についての説明も用意されています。

図表の作成

図表ビルダーを使用すると、定義済みのギャラリ図表から、または図表を構成する個々のパーツ (軸や棒など) から、図表を作成することができます。ギャラリ図表や基本要素を **キャンバス** ([図表ビルダー] ダイアログ ボックスの [変数] リストの右にある広い領域) にドラッグ アンド ドロップして図表を作成します。

図表を作成していくと、図表のプレビューがキャンバスに表示されてきます。プレビューでは定義済みの変数ラベルおよび尺度を使用しますが、実際のデータが表示されるわけではありません。代わりに、無作為に生成されたデータを使用して図表の大まかな姿を見せるものです。

図表の作成に慣れていないユーザーには、ギャラリを使用する方法が適しています。ギャラリの使用方法は、482 ページのギャラリからの図表の作成を参照してください。

図表ビルダーを起動する方法

▶ メニューから次の項目を選択します。

グラフ
チャート ビルダ

[図表ビルダー] ダイアログ ボックスが開きます。

図 43-1
[図表ビルダー] ダイアログ ボックス

ギャラリからの図表の作成

図表を作成する最も簡単な方法は、ギャラリを使用することです。ここでは、ギャラリから図表を作成する際の一般的な手順について説明します。

- [ギャラリ] タブが表示されていない場合は、タブをクリックします。
- [以下から選択] リストから図表のカテゴリを選択します。各カテゴリに複数の種類が用意されています。
- 希望の図表の絵をキャンバスにドラッグします。絵をダブルクリックする方法もあります。すでにキャンバスに図表が表示されている場合は、図表中の軸グループとグラフ要素がギャラリ図表に置き換えられます。
- [変数] リストから変数をドラッグし、軸ドロップ ゾーンおよび (使用可能な場合は) グループ ドロップ ゾーンにドロップします。既に軸ドロップ ゾーンに統計量が

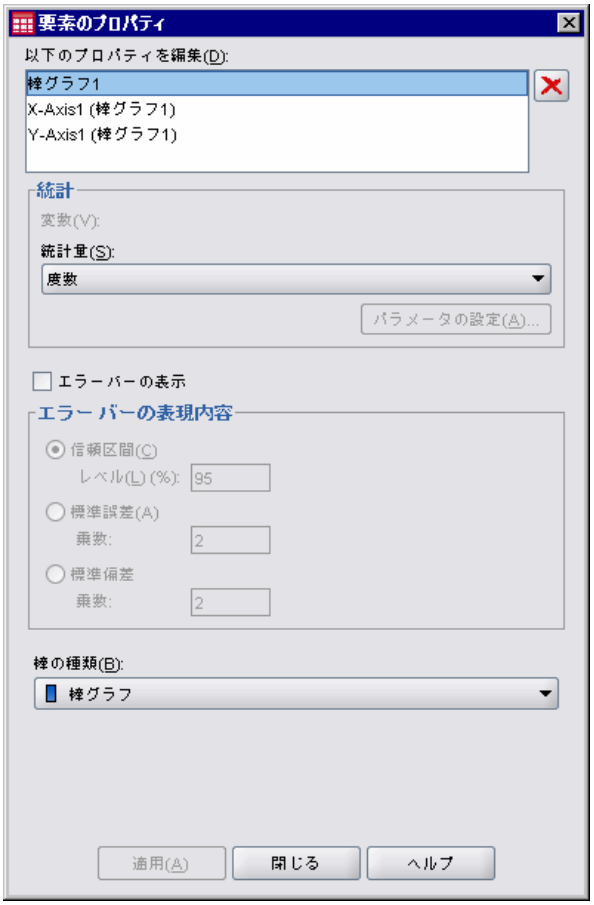
表示されていてその統計量を使用したい場合は、軸ドロップ ゾーンに変数をドラッグする必要はありません。ゾーンに変数を追加する必要があるのは、ゾーン内のテキストの色が青のときだけです。テキスト色が黒の場合、ゾーンには変数または統計量が既に入っています。

注: 変数の測定レベルは重要です。図表ビルダーでは、図表の構築の際、尺度を基にデフォルトを設定します。さらに、尺度が異なると、作成される図表の外観も異なってくることがあります。変数を右クリックし、オプションを選択することで、変数の尺度を一時的に変更できます。

図 43-2
ドロップ領域がすべて設定された [図表ビルダー] ダイアログ ボックス

▶ 統計量を変更したり、軸や凡例の属性 (スケールの範囲など) を修正する場合は、[**要素のプロパティ**] をクリックします。

図 43-3
[要素プロパティ] ダイアログ ボックス



- [以下のプロパティを編集] リストから、変更したい項目を選択します。特定のプロパティの詳細は、[**ヘルプ**] をクリックして参照してください。

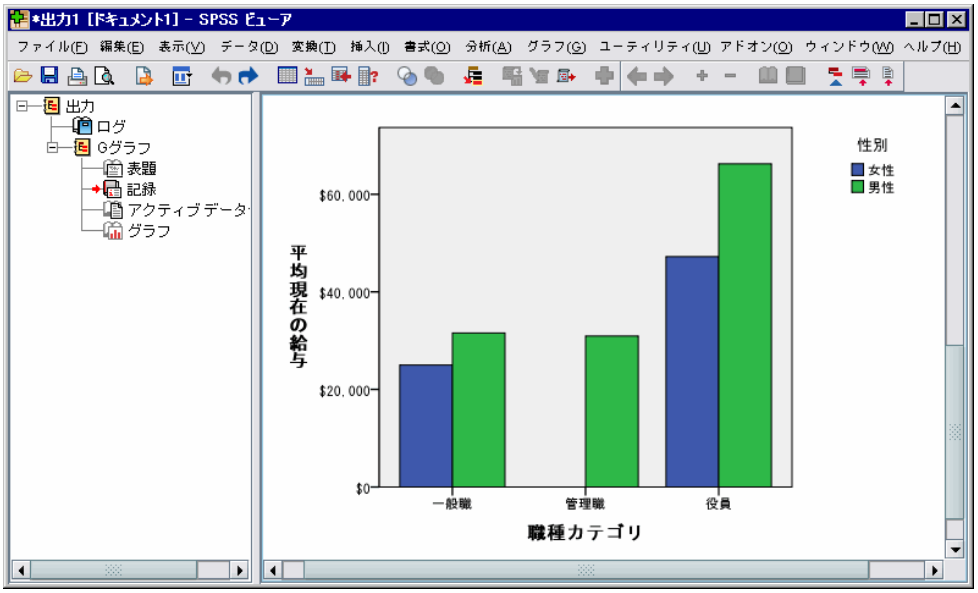
- 変更を行ったら、[**適用**] をクリックします。

- クラスタ化やパネル化などのために変数をさらに図表に追加する必要があるときは、[図表ビルダー] ダイアログ ボックスの [**グループ/ポイント ID**] タブをクリックし、1 つ以上のオプションをクリックします。次に、キャンバスに表示される新しいドロップ ゾーンにカテゴリ変数をドラッグします。

- 図表の縦横を入れ換える (棒を水平にするなど) には、[**基本要素**] タブをクリックしてから、[**行と列の入れ換え**] をクリックします。

- [OK] をクリックして、図表を作成します。ビューアに図表が表示されます。

図 43-4
[ビューア] ウィンドウに表示される棒グラフ



図表の編集

図表エディタは、図表をカスタマイズしたり、データを分析するための、強力で使いやすい環境を提供します。図表エディタは次のような機能があります。

- **簡単に操作しやすいユーザー インターフェイス。**メニュー、コンテキスト メニュー、およびツールバーを使用して図表の各部分を選択して素早く編集できます。また、図表に直接テキストを入力することもできます。
- **各種書式設定オプションおよび統計オプション。**幅広い範囲のスタイルおよび統計オプションが用意されています。
- **強力な探索ツール。**ラベル付け、並べ替え、回転などの操作を行って、データをいろいろな方法で分析できます。図表の種類および図表内の変数の役割を変更できます。さらに、分布曲線、当てはめ線、線補間、参照線を追加することもできます。
- **一貫した外観と動作を提供する柔軟なテンプレート。**カスタマイズしたテンプレートを作成し、それを元に目的の外観やオプションを持つ図表を簡単に作成できます。軸ラベルを特定の方向に位置付けたい場合は、テンプレートに方向を指定し、テンプレートを他の図表に適用します。

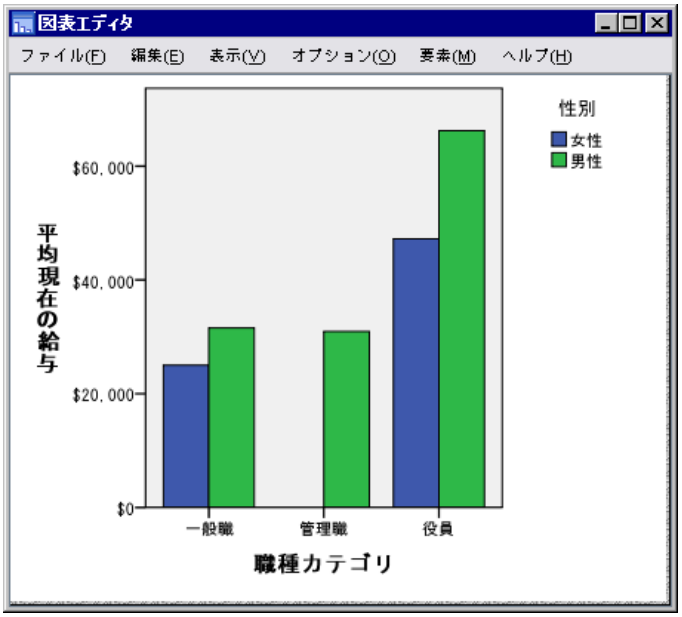
図表エディタを表示する方法

▶ SPSS で図表を作成するか、図表を含んだビューア ファイルを開きます。

▶ ビューアで図表をダブルクリックします。

図表エディタに図表が表示されます。

図 43-5
図表エディタに図表が表示されます。



図表エディタの基礎

図表エディタでは、いろいろな方法で図表を操作できます。

メニュー

図表エディタで実行できる多くの処理は、メニューを通じて行います。特に、項目を図表に追加する場合がこれに該当します。たとえば、メニューを使用して線の当てはめを散布図に追加します。通常、項目を図表に追加した後は、[プロパティ] ダイアログ ボックスを使用して、追加した項目のオプションを指定します。

[プロパティ] ダイアログ ボックス

[プロパティ] ダイアログ ボックスには、図表および図表要素のオプションが用意されています。

[プロパティ] ダイアログ ボックスを表示するには、次の操作を行います。

▶ 図表要素をダブルクリックします。

または

▶ 図表要素を選択し、メニューから次の項目を選択します。

編集
プロパティ

また、[プロパティ] ダイアログ ボックスは、項目を図表に追加すると自動的に表示されます。

図 43-6
[テーブル プロパティ] ダイアログ ボックスの [罫線] タブ

[プロパティ] ダイアログ ボックスには、オプションを設定したり、図表にその他の変更を加えるためのタブが用意されています。[プロパティ] ダイアログ ボックスに表示されるタブは、選択項目に基づいて変化します。

一部のタブには、変更を適用した場合に選択項目 (図表要素) にどのような結果が得られるかを示すプレビュー機能があります。ただし、**[適用]** をクリックしない限り、変更は図表に反映されません。**[適用]** をクリックするのは、複数のタブで変更を加えた後でかまいません。選択項目を変更して図表内の異なる要素を修正する必要がある場合、**[適用]** をクリックしてから、選択項目を変更してください。**[適用]** をクリックせずに選択項目 (図表要素) を変更した場合、[プロパティ] ウィン

ドウでの変更内容は反映されません。[適用] のクリックによって変更が反映されるのは、現在選択されている図表要素のみです。

選択した項目によっては、ある設定のみが使用できます。個々のタブのヘルプを参照すると、タブを表示するには、どの項目を選択する必要があるかがわかります。複数の項目を選択した場合は、それらの要素すべてに共通の設定だけを変更できます。

ツールバー

[プロパティ] ダイアログ ボックスのいくつかの機能に対するショートカットを提供します。たとえば、[プロパティ] ダイアログ ボックスの [テキスト] タブを使用する代わりに [編集] ツールバーを使用して、テキストのフォントやスタイルを変更できます。

変更の保存

図表変更は図表エディタを閉じると保存されます。[ビューア] に変更された図表が続いて表示されます。

図表定義のオプション

図表ビルダーの図表を定義しているときに、図表作成のために表題を追加したりオプションを変更することができます。

表題および脚注の追加と編集

図表をわかりやすくするために、表題や脚注を追加することができます。脚注にはエラー バーの情報が自動的に表示されます。

表題と脚注を追加する方法

- [表題/脚注] タブをクリックします。
- 1 つ以上の表題と脚注を選択します。図表に表題と脚注が追加されたことを示すテキストが、キャンバスに表示されます。
- [要素のプロパティ] ダイアログ ボックスで、表題および脚注のテキストを編集します。

表題または脚注を削除する方法

- [表題/脚注] タブをクリックします。
- 削除したい表題または脚注の選択を解除します。

表題または脚注のテキストを編集する方法

表題や脚注を追加した場合、それぞれに関連付けられているテキストを図表上で直接編集することはできません。テキストの編集は、図表ビルダー内の他の項目と同様に、[要素のプロパティ] ダイアログ ボックスで行います。

- [要素のプロパティ] ダイアログ ボックスが表示されていなければ、**[要素のプロパティ]** をクリックします。
- [以下のプロパティを編集] リストから、表題、副題、または脚注 (**[表題 1] など**) を選択します。
- [Content] ボックスに、表題、副題、または脚注に関連付けるテキストを入力します。
- **[適用]** をクリックします。

[全般] オプションの設定

図表ビルダーでは、図表全般に関するオプションが用意されています。図表上の特定の項目ではなく、図表全体に適用されるオプションです。全般オプションでは、欠損値の処理、テンプレート、図表サイズ、パネルの折り返しなどを指定できます。

- **[オプション]** をクリックします。
- 全般オプションを変更します。以下に詳細を示します。
- **[適用]** をクリックします。

ユーザー欠損値

ブレーク変数。カテゴリやサブグループを定義するために使用する変数に欠損値がある場合、**[含む]** を選択すると、ユーザー欠損値 (ユーザーによって欠損値として識別された値) のカテゴリが図表に含まれるようになります。これらの「欠損した」カテゴリは、統計量を計算する際のブレーク変数としても機能します。「欠損」カテゴリは、カテゴリ軸上または凡例内に表示され、たとえば、余分な棒を追加したり、分割を円グラフに追加します。欠損値がない場合、「欠損」カテゴリは表示されません。

このオプションがオンのときに、図表が描かれた後にカテゴリが表示されないようにするには、図表エディタで図表を開き、[編集] メニューから **[プロパティ]** を選択します。[カテゴリ] タブで、非表示にするカテゴリを [除外] リストに移動します。しかし、「欠損した」カテゴリを隠すと、統計量は再計算されません。このため、パーセントの統計量などは、「欠損した」カテゴリを考慮する必要があります。

注: このコントロールはシステム欠損値に影響しません。システム欠損値は図表から除外されます。

要約統計量とケース値。欠損値を持つケースを除外するに関しては、次のオプションのどちらかを選択できます。

- **図表に対する一致したケース基準を得るためにリストごとに除外する。**与えられたケースについて図表内の変数のどれかに欠損値がある場合、そのケース全体が図表から除外されます。
- **データを最大限に利用するために変数ごとに削除する。**選択された変数に欠損値がある場合、変数の分析時にそれらの欠損値を含むケースが除外されます。

欠損値のリストごとの除外と変数ごとの除外の違いについては、次の図を参照してください。2 つのオプションそれぞれについて棒グラフが示されています。

図 43-7
欠損値のリストごとの除外

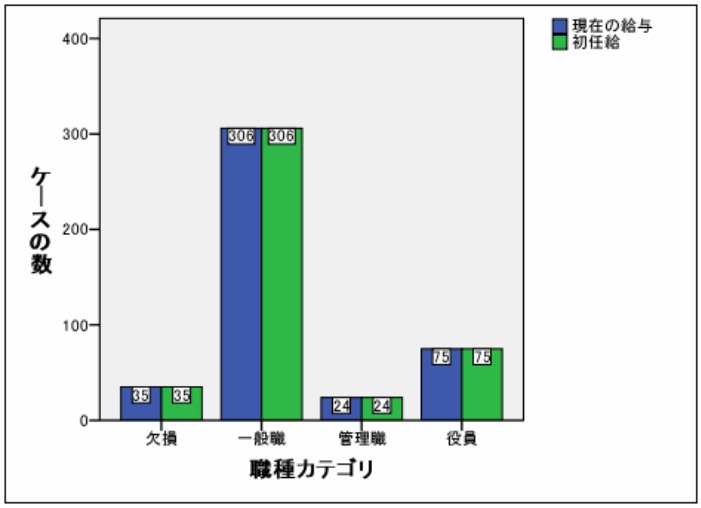
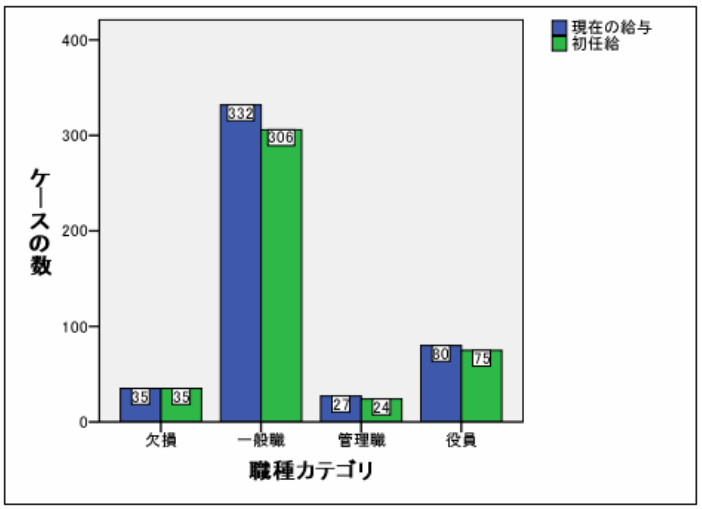


図 43-8
欠損値の変数ごとの除外



これらの図表は、あるバージョンの Employee data.sav ファイルを基に作成したものです。ただしこのファイルは、現在の給与と職種のカテゴリの変数にシステム欠損 (空白) 値がいくつか含まれるよう編集されています。それ以外に、値 0 を入れて欠損として定義した場合もあります。どちらの図表においても、**[欠損値グループの表示]** オプションがオンになっており、表示されている他の職種カテゴリに **「欠損」** カテゴリが追加されています。またそれぞれの図表では、**「ケースの数」**という集計関数の値が棒のラベル内に表示されています。

両方の図表内で、26 個のケースに職種カテゴリのシステム欠損値があり、13 個のケースにユーザー欠損値 (0) があります。リストごとの図表では、ケースの数は各棒クラスタ内の両方の変数について同じです。値が欠損値である限り、すべての変数に関してそのケースが除外されているからです。変数ごとの図表では、カテゴリ内の各変数の非欠損ケースの個数は他の変数内の欠損値と無関係にプロットされます。

テンプレート (テンプレート)

図表テンプレートを使用すると、図表の属性を別の図表に適用することができます。図表エディタで図表を開いたときは、それをテンプレートとして保存することができます。こうして保存したテンプレートは、作成時にここで指定して適用したり、後で図表エディタで適用することができます。

デフォルト テンプレート。これは、[オプション] で指定されるテンプレートです。このオプションにアクセスするには、データ エディタの [編集] メニューから **[オプション]** を選択し、**[図表]** タブをクリックします。デフォルト テンプレートは最初に適用されるテンプレートです。これは、他のテンプレートに置き換えることができます。

テンプレート ファイル。[追加] をクリックし、標準のファイル選択ダイアログ ボックスで 1 つ以上のテンプレートを指定します。各テンプレートは、その表示順に適用されます。したがって、リスト末尾のテンプレートはリスト先頭のテンプレートを上書きすることができます。

図表サイズとパネル

図表サイズ。パーセントの値として 100 より大きな値を指定すると図表は拡大され、100 より小さな値を指定すると縮小されます。パーセントの値は、デフォルトの図表サイズに対する割合を表します。

パネル。パネルの列が多くなる場合は、**[パネルを折り返す]** を選択することにより、パネルを 1 行で表示するのではなく、複数の行に折り返して表示することができます。このオプションが選択されていないと、パネルは 1 行の中に表示できるよう縮小されます。

章
44

ユーティリティ

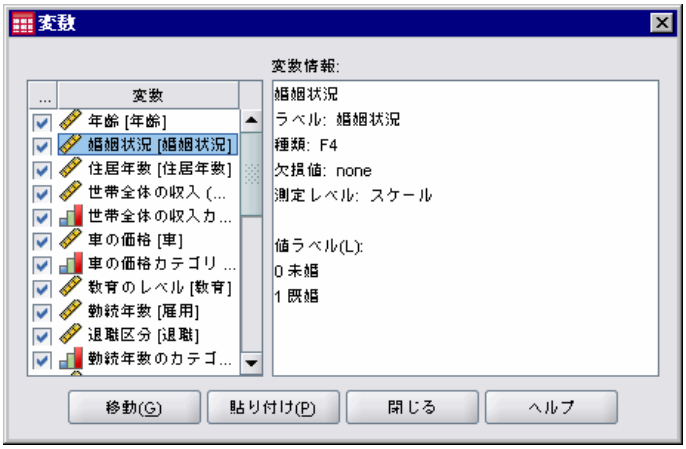
この章では、[ユーティリティ] メニューにある機能を使用して目標変数リストを並べ替える機能について説明します。

変数情報（変数）

[変数] ダイアログ ボックスには、現在選択されている変数の変数定義情報が表示されます。この情報には、次のものが含まれています。

- 変数ラベル
- データ形式
- ユーザー欠損値
- 値ラベル
- 尺度

図 44-1
[変数] ダイアログ ボックス



表示されます。変数一覧の表示列は、[データ エディタ] および ダイアログ ボックス変数リストに変数が見えているかどうかを示します。表示は変数セットによりコントロールされています。 詳細は、493 ページの 変数セットを参照してください。

移動。データ エディタ ウィンドウ内の選択した変数へ移動します。

貼り付け。選択した変数を指定したシンタックス ウィンドウのカーソルの置かれている位置に貼り付けます。

変数の定義を修正するには、データ エディタの変数ビューを使用します。

変数情報を表示するには

▶ メニューから次の項目を選択します。

ユーティリティ
変数...

▶ 定義情報を表示したい変数を選択します。

データ ファイルのコメント

データ ファイルに分かりやすいコメントを付けられます。SPSS 形式のデータ ファイルでは、コメントはデータ ファイルといっしょに保存されます。

データ ファイルのコメントを追加、変更、削除、または表示するには

▶ メニューから次の項目を選択します。

ユーティリティ
データ ファイルのコメント...

▶ ビューアでコメントを表示するには、**[出力時にコメントを表示する]** を選択します。

コメント全体の長さに制限はありませんが、1 行あたり 80 バイト (1 バイト言語では 80 文字) に制限されます。行は、80 文字で自動的に折り返されます。コメントは、テキスト出力と同じフォントで表示され、ビューアに表示されたときの外観を正確に反映します。

コメントを追加または変更すると、日付スタンプ (カッコ内に示される現在の日付) が自動的にコメントのリストの最後に追加されます。既存のコメントを変更したり既存のコメントの間に新しいコメントを追加した場合、コメントに関連付けられている日付があいまいになることがあります。

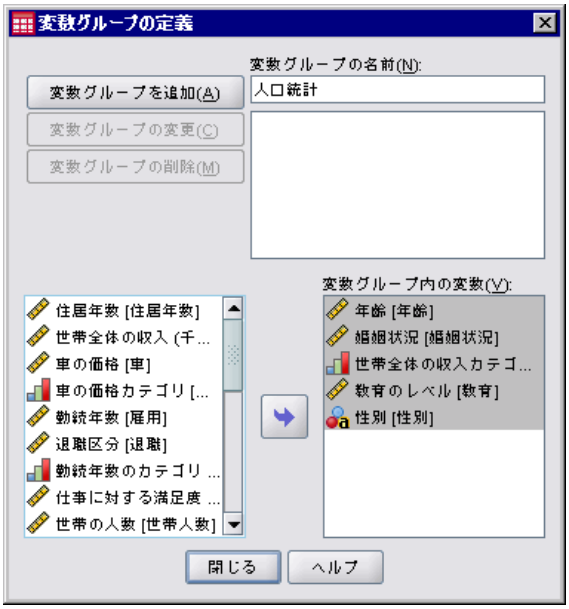
変数セット

データ エディタおよびダイアログ ボックスの変数リストに表示される変数は、変数グループを定義、使用することにより制限できます。これは、変数が大量に含まれているデータ ファイルには特に有効です。小規模な変数グループにより、分析に必要な変数を簡単に検出および選択することが可能になります。

変数グループの定義

[変数グループの定義] ダイアログ ボックスでは、データ エディタやダイアログ ボックスの変数リストに表示される変数のサブグループを作成します。定義した変数グループは、SPSS 形式のデータ ファイルとともに保存されます。

図 44-2
[変数グループの定義] ダイアログ ボックス



変数グループの名前。設定名の長さは、最高で 64 バイトです。空白を含むすべての文字を使うことができます。

グループ内の変数。数値および文字列の変数の組み合せならどれでも、グループに含めることができます。グループの中の変数の順序が、データ エディタまたはダイアログ ボックスの変数リストの変数表示順に影響することはありません。1 つの変数を複数のグループに所属させることができます。

変数グループを定義するには

▶ メニューから次の項目を選択します。

ユーティリティ
変数グループの定義...

▶ グループに含めたい変数を選択します。

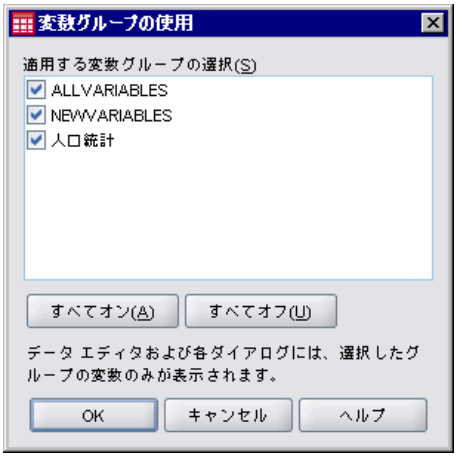
▶ グループ名を入力します (最大 64 バイト)。

▶ [**変数グループを追加**] をクリックします。

変数グループの使用

[変数グループの使用] ダイアログ ボックスで、データ エディタやダイアログ ボックスの変数リストに表示される変数を、選択した (チェックマークを付けた) グループの変数に制限します。

図 44-3
[変数グループの使用] ダイアログ ボックス



- データ エディタやダイアログ ボックスの変数リストに表示される変数は、選択したグループすべてを結合したものです。
- 1 つの変数が複数の選択したグループに含まれていてもかまいません。
- 選択したグループに含まれる変数の順序や選択したグループの順序が、データ エディタまたはダイアログ ボックスの変数リストの変数表示順に影響することはありません。
- 定義した変数グループは SPSS 形式のデータ ファイルとともに保存されますが、現在選択しているグループのリストは、データ ファイルを開くたびにデフォルトの組み込みグループにリセットされます。

使用可能な変数グループのリストには、アクティブなデータセットに対して定義されている変数グループの他、次の 2 つの組み込みグループも含まれます。

- **全変数。**このグループには、セッション中に作成された新変数を含め、データ ファイルに含まれているすべての変数が含まれます。
- **新変数。**このグループには、そのセッション中に作成された新しい変数だけが含まれます。

  注: 新しい変数を作成した後でデータ ファイルを保存しても、データ ファイルを閉じて再度開くまでは、それらの変数は NEWVARIABLES グループに 含まれたままとなります。

変数グループは、少なくとも 1 つ選択しなければなりません。ALLVARIABLES にはすべての変数が含まれているので、このグループを他のグループとともに選択しても、表示される変数は制限されません。

表示する変数グループを選択するには

▶ メニューから次の項目を選択します。

ユーティリティ
変数グループの使用...

▶ データ エディタやダイアログ ボックスの変数リストに表示したい変数を含む定義済み変数グループを選択します。

すべての変数を表示するには

▶ メニューから次の項目を選択します。

ユーティリティ
すべての変数を表示

目標変数リストの並べ替え

変数は、ダイアログ ボックスの目標リストに、ソース変数リストから選択された順序で表示されます。目標リストにおける変数の順序を変更したいのに、すべての変数の選択を解除して新たな順序で選択し直したくない場合は、Ctrl キー (Macintosh では Command キー) を使用して目標リストの変数を上下に移動することができます。変数は連続して (一緒にグループ化されて) いれば、複数の変数を同時に移動できます。連続していない変数グループを同時に移動することはできません。

章
45

オプション

オプションで、次を含む広範囲の SPSS の設定をコントロールします。

- あらゆるセッションでのコマンドすべての記録を保管する SPSS セッション ジャーナル
- ダイアログ ボックスの選択元リストでの変数の表示順
- 新しい出力結果で項目の表示/非表示
- 新規ピボット テーブルのテーブルルック
- ユーザーの指定による通貨書式

SPSS オプション設定を変更するには

▶ メニューから次の項目を選択します。
編集
オプション...

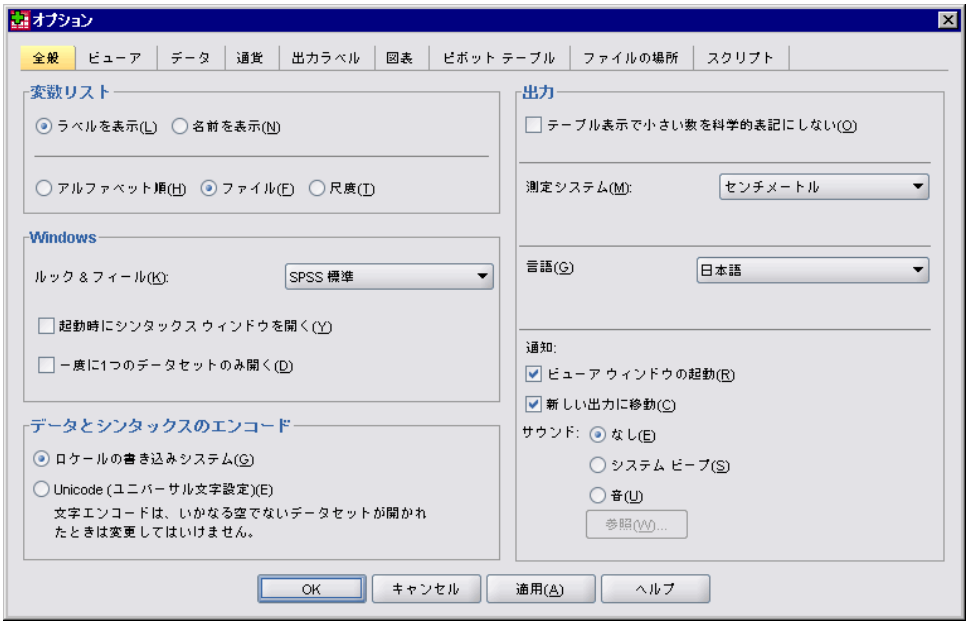
▶ 変更したい設定のタブをクリックします。

▶ 設定を変更します。

▶ [OK] または [**適用**] をクリックします。

[全般] オプション

図 45-1
[オプション] ダイアログ ボックス: [全般] タブ



変数リスト

これらの設定はダイアログ ボックスのリストでの変数の表示の仕方をコントロールします。変数名または変数ラベルを表示できます。名前またはラベルは、アルファベット順またはファイル順に表示されるか、または尺度によりグループ化されます。表示順は選択元変数リストにのみ影響します。目標変数リストは、常に変数が選択された順序を反映します。

Windows

外観と使い勝手。ウィンドウおよびダイアログ ボックスの基本外見をコントロールします。

起動時にシンタックス ウィンドウを開く。シンタックス ウィンドウは、SPSS コマンドの入力、編集および実行に使用するテキスト ファイル ウィンドウです。コマンド シンタックスで頻繁に作業をする場合、このオプションを選択すると SPSS の起動時にシンタックス ウィンドウが自動的に開きます。これは、ダイアログ ボックスの代わりにコマンド シンタックスで作業する SPSS ユーザーには特に有効です。

一度に 1 個のデータセットを開く。メニューおよびダイアログ ボックスを使用して別のデータ ソースを開くときはいつも現在開けているデータ ソースを閉じます。デフォルトでは、新規のデータ ソースを開けるためにメニューおよびダイアログ ボッ

クスを使用するときはいつも、新規 [データ エディタ] ウィンドウ内で開いているデータ ソース、および他の [データ エディタ] ウィンドウ内で開いている他のデータ ソースは、明示的に閉じるまでセッションの間は開いたままで利用可能です。

このオプションを選択するとすぐに効果がありますが、設定が変更されたときに開いていたデータセットはすべて閉じられることはありません。この設定は、複合データセットをコントロールする DATASET コマンドに基づくコマンド シンタックスを使用して開けたデータ ソースには影響を及ぼしません。詳細は、108 ページの 6 章の多重データ ソースの使用方法を参照してください。

データ ファイルおよびシンタックス ファイルの文字エンコード

これは、データ ファイルおよびシンタックス ファイルの読み込みおよび書き込みのエンコードを決定するデフォルトの動作をコントロールします。この設定は開いているデータ ソースがないときにだけ変更でき、明示的に変更されるまで次のセッションにも効果があります。

- **ロケールの書き込みシステム。**ファイルの読み込みおよび書き込みのエンコードを決定するために現在のロケール設定を使用します。この方法は **コード ページ モード**と呼ばれます。
- **Unicode (ユニバーサル文字セット)。**ファイルの読み込みおよび書き込みに Unicode エンコード (UTF-8) を使用します。この方法は **Unicode モード**と呼ばれます。

Unicode モードおよび Unicode ファイルに関しては、次のような重要な意味があります。

- Unicode エンコードで SPSS 形式のデータ ファイルおよびシンタックス ファイルを保存すると、16.0 以前の SPSS のリリースでは使用できません。シンタックス ファイルでは、ファイルを保存するときにエンコードを指定できます。データ ファイルは、以前のバージョンでファイルを読み込みたいときは必ずコード ページ モードでデータ ファイルを 開き、再度保存してください。
- コード ページ データ ファイルが Unicode モードで読み込まれるときは、すべての文字列変数の定義幅は 3 倍になります。変数の最長観測値に各文字変数の幅を自動的に設定するには、[データを開く] ダイアログ ボックスの [**観測値に基づいて文字幅を最小化**] を使用します。

出力 (X11 ARIMA)

テーブル表示で小さい数を科学的表記にしない。出力で小さな小数値が科学的表記で表示されないようにします。非常に小さな小数値は 0 (または 0.000) として表示されます。

測定システム。ピボット テーブルのセルの余白、セル幅、および印刷するためのテーブル間の空白のような属性を指定するときに使用する測定システム (ポイント、インチ、またはセンチメートル)。

言語。出力で使用する言語を指定します。単純なテキスト出力に適用しないでください。利用できる言語の一覧は、現在インストールしてある言語ファイルで変わります。

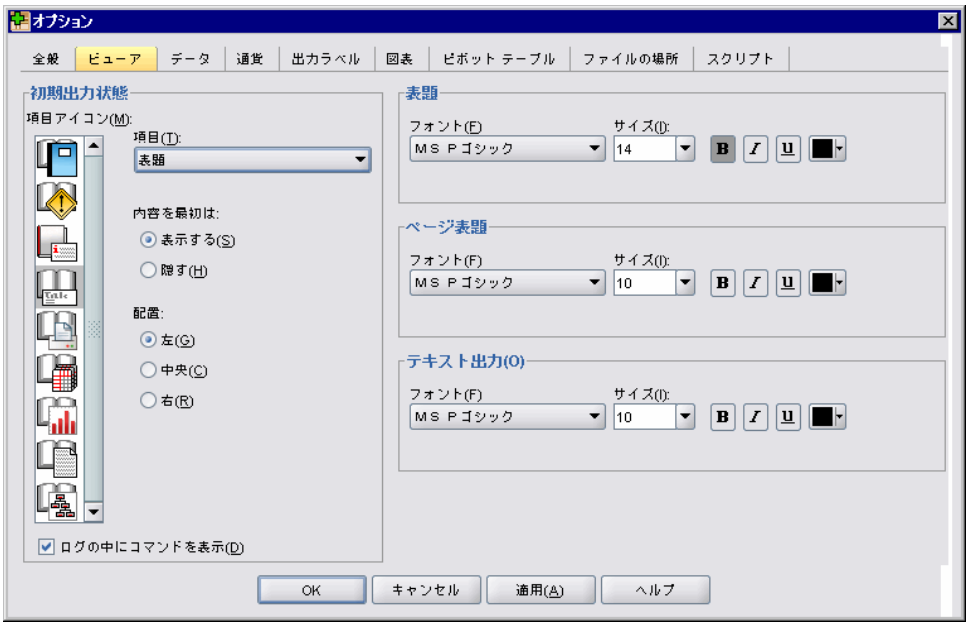
注: 出力言語を変更した場合、出力にある言語固有のテキスト文字列に依存しているユーザー指定のスクリプトが、正常に実行されない可能性もあります。 詳細は、512 ページの[スクリプト] オプションを参照してください。

通知。手続きの実行を終了したことおよび結果がビューアで利用できることを、SPSS がユーザーに通知する方法をコントロールします。

[ビューア] オプション

ビューアの出力表示オプションは、設定を変更した後で作成した新しい出力にのみ影響します。ビューアにすでに表示されている出力は、設定を変更しても変更されません。

図 45-2
[オプション] ダイアログ ボックス: [ビューア] タブ



初期出力状態。手続きを実行するたびに自動的に項目の表示/非表示をコントロールし、さらに項目の初期割り当て方法もコントロールします。表示をコントロールできるのは、ログ、警告、記録、表題、ピボット テーブル、図表、樹形図およびテキスト出力です。また、ログ中のコマンドの表示のオンまたはオフを切り替えることができます。ログからコマンド シンタックスをコピーし、シンタックス ファイルに保存できます。

注: 出力項目のすべては、ビューアでは左位置合わせで表示されます。印刷した出力の位置合わせだけが位置合わせ設定で影響を受けます。中央位置合わせと右位置合わせは、項目の上左に付けた小さな記号で識別されます。

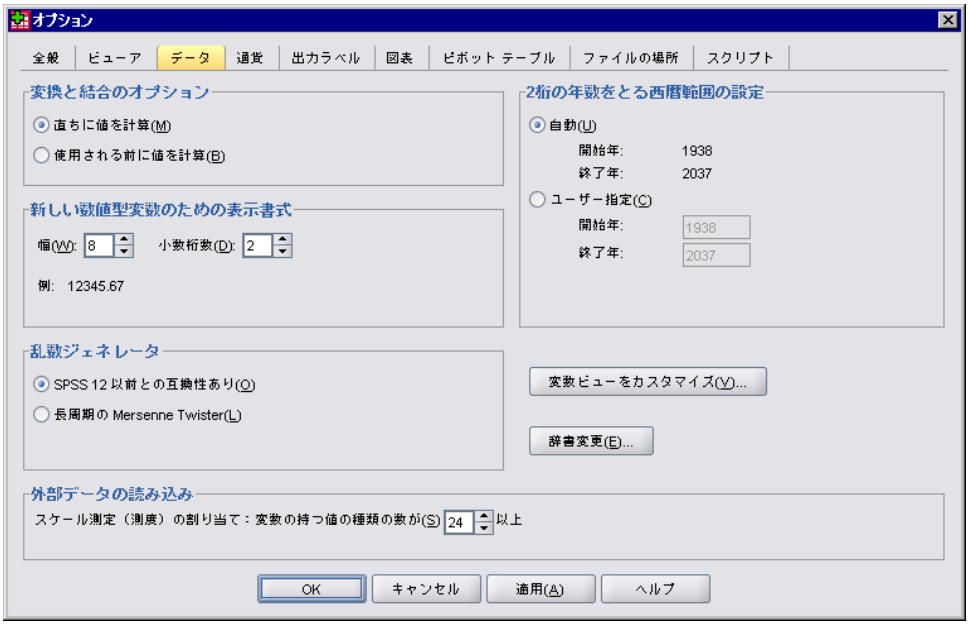
表題。新しい出力表題のフォントの種類、大きさ、および色をコントロールします。

ページ表題。TITLE コマンド シンタックスと SUBTITLE コマンド シンタックスを使って生成したか [**挿入**] メニューの [**新しいページ表題**] で作成した新しいページ表題のフォントの種類、大きさ、および色を制御します。

テキスト出力。出力に使用するフォント。SPSS のテキストの出力は、等幅 (固定ピッチ) フォントで使用するようになっています。プロポーショナル フォントを選択すると、集計表出力は正しく位置合わせされません。

[データ] オプション

図 45-3
[**オプション**] ダイアログ ボックス: [**データ**] タブ



変換と結合のオプション。SPSS は、コマンドを実行するたびにデータ ファイルを必要とします。いくつかのデータ変換 (たとえば、計算と値の再割り当て) およびファイル変換 (たとえば、変数の追加およびケースの追加) はデータの個別パスを必要としないので、統計手続きやグラフ作成手続きのような別のコマンドを実行するために、SPSS がデータを読み込むまでコマンドの実行を遅らせることができます。

- 大きなデータ ファイルではデータの読み込みに時間がかかる場合があるので、[**使用される前に値を計算**] を選択すると実行時間を遅らせて処理時間を節約することができます。このオプションをオンにすると、たとえば [変数の計算] ダイアログ ボックスで行った変換の結果は、データ エディタに直ちに表示さ

れません。変換によって作成された新しい変数はデータ値を持たない状態で表示され、保留されている変換がある間はデータ エディタ内のデータ値を変更できません。統計手続きやグラフ作成手続きのようにデータを読み込むコマンドによって保留中の変換が実行され、データ エディタに表示されているデータが更新されます。または、[変換] メニューの **[保留された変換の実行]** を使用することもできます。

- デフォルトの設定は **[直ちに値を計算]** です。この場合に、ダイアログ ボックスからコマンド シンタックスを貼り付けると、各変換コマンドの後に EXECUTE コマンドが貼り付けられます。 詳細は、281 ページの 12 章の複数の Execute コマンドを参照してください。

新しい数値型変数のための表示書式。新しい数値型変数のデフォルトの表示幅および小数桁数を指定します。新規文字型変数にはデフォルトの表示書式はありません。値が指定した表示書式に大きすぎる場合、SPSS は最初に小数桁数を概算してから値を科学表記に変換します。表示書式は内部のデータ値には影響しません。たとえば、123456.78 という値は 123457 と丸められて表示されますが、どの計算でも丸められていない元の値が使用されます。

2 桁の年数をとる西暦範囲の設定。2 桁の年数 (たとえば、10/28/86、29-OCT-87 など) で入力したり表示したりする日付書式変数の年数の範囲を定義します。自動範囲設定は現在の年に基づいており、現在の年から数えて 69 年前に始まり 30 年後に終わります (現在の年を加えると合計 100 年になります)。ユーザーの指定による範囲では、終了年は開始年として入力された値に基づいて自動的に決定されます。

乱数ジェネレータ。2 種類の異なる乱数ジェネレータを使用できます。

- **バージョン 12 互換。**バージョン 12 と以前のリリースで使用された乱数ジェネレータです。指定されたシード値に基づいて以前のリリースで作成されたランダム化された結果を再生する必要がある場合は、この乱数ジェネレータを使用してください。
- **Mersenne Twister。**シミュレーションを行う場合に利用できる、より信頼性の高い新しい乱数ジェネレータです。バージョン 12 またはそれ以前にランダム化された結果の再生ができる場合は、この乱数ジェネレータを使用します。

外部データ ファイルの読み込み。外部ファイル形式および前の SPSS 形式の データ ファイル (リリース 8.0 以前) では、数値型変数に対して、その変数がスケール変数か名義変数かを分類するためのデータ値の最小数を指定できます。その一意の値の数が指定された数より少ない変数は名義変数として分類されます。

変数ビューをカスタマイズ。[変数ビュー] の属性のデフォルト表示および順序を制御します。 詳細は、503 ページのデフォルト変数の変更を参照してください。

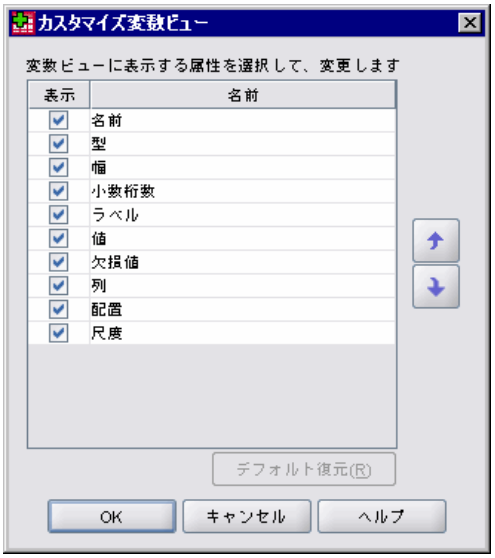
辞書を変更。[変数ビュー] の項目のスペルをチェックするために使用された辞書の言語バージョンを制御します。 詳細は、98 ページの 5 章のスペル チェック変数と値ラベルを参照してください。

デフォルト変数の変更

[変数ビュー] (例えば、名前、型、ラベル) にデフォルトで表示された属性および表示された順序を制御するには、[変数ビューをカスタマイズ] を使用できます。

[**変数ビューをカスタマイズ**] をクリックします。

図 45-4
変数ビューをカスタマイズ (デフォルト)



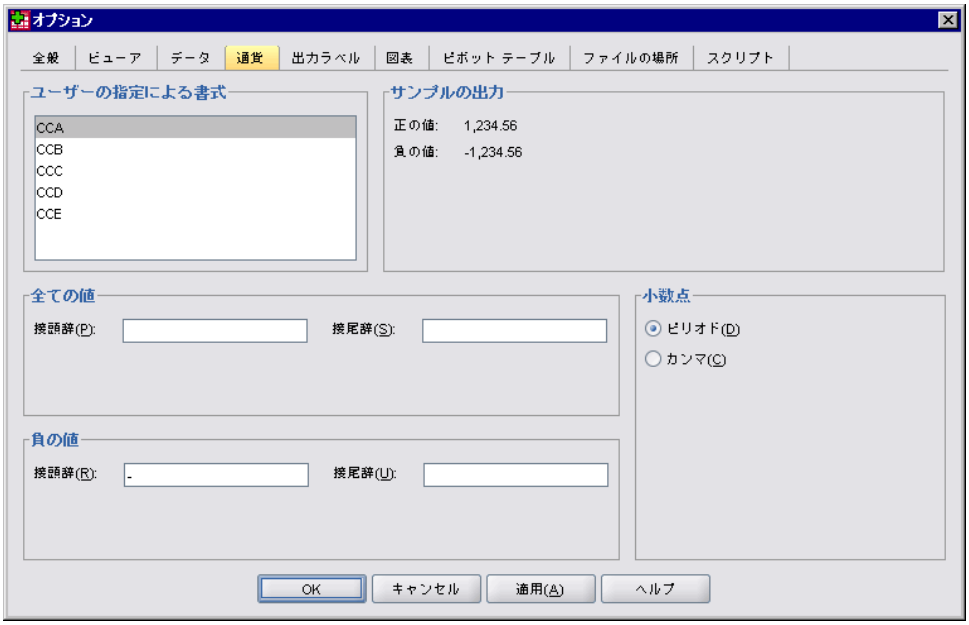
- 表示する変数属性を選択 (チェック) します。
- 属性の表示順を変更するには上および下矢印を使用します。

[通貨] オプション

特別な接頭辞や接尾辞、および負の値に対して特別な処理を使用できる、ユーザーの指定の通貨表示書式を、5 つまで作成することができます。

5 つのユーザーの指定による通貨書式の名前は、CCA、CCB、CCC、CCD、および CCE になります。書式名は変更できません。また、新しい書式名を追加することもできません。ユーザーの指定による通貨書式を修正するには、修正元リストから書式名を選択して希望する変更を行います。

図 45-5
[オプション] ダイアログ ボックス: [通貨] タブ



ユーザーの指定の通貨書式で定義した接頭辞、接尾辞、および小数点は、表示目的専用です。ユーザーの指定の通貨書式文字を使用してデータ エディタに値を投入することはできません。

通貨書式を作成するには

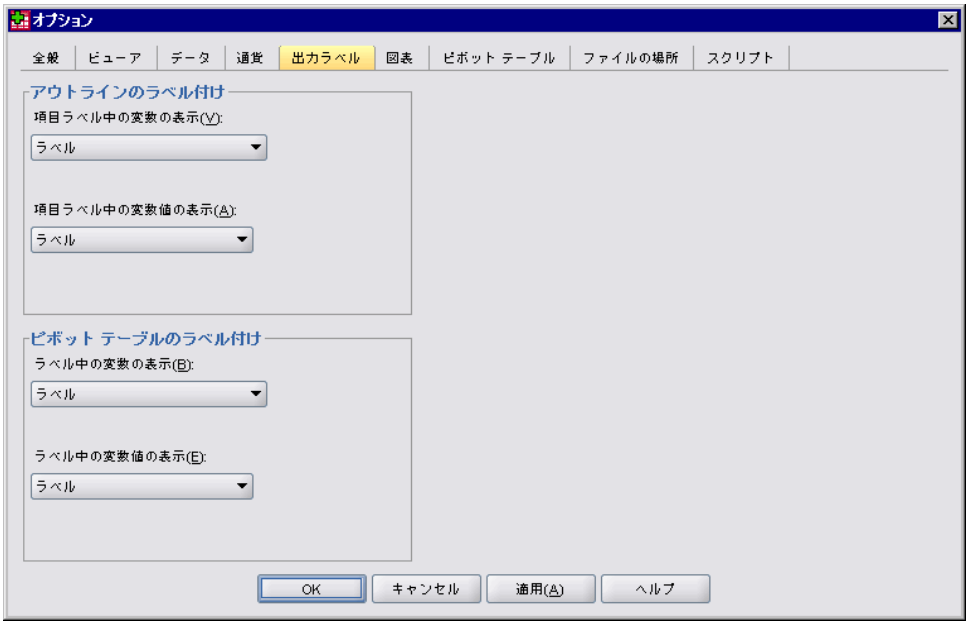
- [通貨] タブをクリックします。
- [ユーザーの指定による書式] ボックスの一覧 (CCA、CCB、CCC、CCD、CCE) から、通貨書式を 1 つ選択します。
- 接頭辞、接尾辞、および小数点を入力します。
- [OK] または [適用] をクリックします。

[出力ラベル] オプション

出力ラベル オプションは、アウトラインおよびピボット テーブル内の変数とデータ値情報の表示をコントロールします。変数名、定義済みの変数ラベルと実際のデータ値、定義済みの値ラベル、またはこれらの組み合せを表示することができます。

記述変数と値ラベル (データエディタ、[ラベル] 列、および [値] 列の [変数ビュー]) を使用すると多くの場合、結果を容易に解釈することができます。ただし、テーブルによっては、長いラベルは不便な場合もあります。

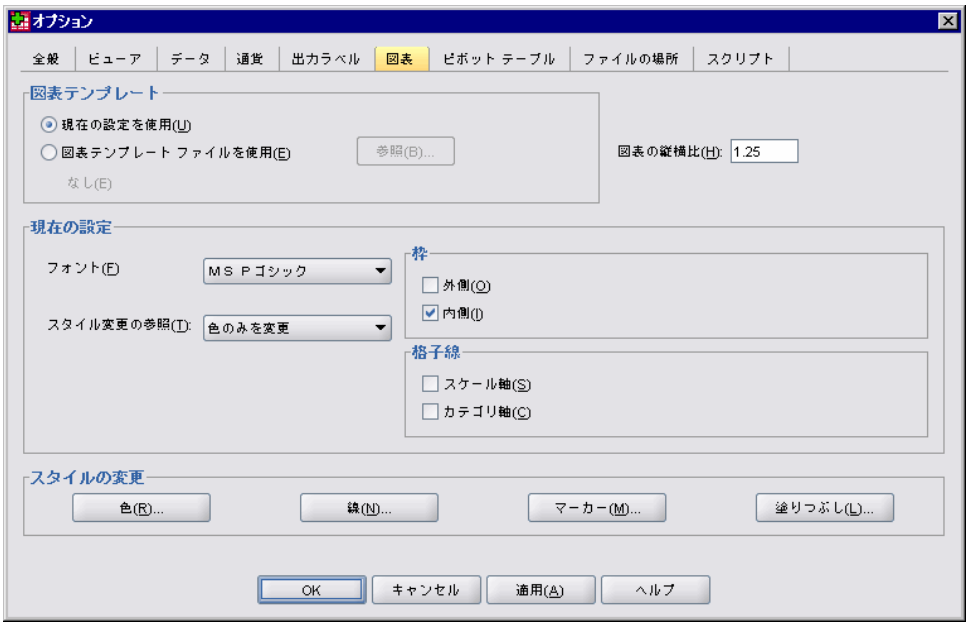
図 45-6
[オプション] ダイアログ ボックス: [出力ラベル] タブ



出力ラベル オプションは、設定の変更後に作成した新しい出力にのみ影響します。ビューアにすでに表示されている出力は、設定を変更しても変更されません。設定はピボット テーブルの出力にのみ影響します。テキストの出力はこの設定に影響されません。

[グラフ] オプション

図 45-7
[オプション] ダイアログ ボックス: [グラフ] タブ



図表テンプレート。新しい図表は、ここで選択した設定または図表テンプレート ファイルから取り入れた設定のどれかを使用できます。**[参照]** をクリックして図表テンプレート ファイルを選択します。図表テンプレート ファイルを作成するには、希望する属性で図表を作成してからそれをテンプレートとして保存します ([ファイル] メニューの **[図表テンプレートの保存]** を選択)。

図表の縦横比。新しい図表の外枠の幅対高さの比です。幅対高さの比は、0.1 から 10.0 まで指定することができます。値を 1 よりも小さくすると幅よりも高い図表になります。値を 1 よりも大きいと、幅が高さよりも大きくなります。値が 1 では正方形の図表になります。図表をいったん作成すると、その縦横比は変更することができません。

現在の設定。使用できる設定は次のとおりです。

- **フォント。**新しい図表ですべてのテキストに使用するフォント。
- **スタイル変更の参照。**新しい図表の色および塗りつぶしの初期割り当てです。**[色のみを変更]** は、色だけを使用して図表要素を区別し、塗りつぶしは使用しません。**[模様のみを変更]** は、線の種類、マーカー、または塗りつぶしパターンを使用して図表要素を区別し、色は使用しません。
- **枠。**新しい図表の内枠と外枠の表示をコントロールします。
- **格子線。**新しい図表でスケール軸およびカテゴリ軸のグリッド線の表示をコントロールします。

スタイルの変更。新しい図表に適用する色、線の種類、マーカー、および塗りつぶしパターンをカスタマイズします。新しい図表が作成されるときに使用される色やパターンの順序を変更できます。

データ要素 色

図表内のデータ要素に使用する色の順序を指定します (棒やマーカー)。色は、メイン [オプション] ダイアログ ボックスの [グラフ] タブにある [スタイル変更の参照] グループで名前に **色**を含むオプションを選択すると使用されます。

たとえば、2 つのグループを使用したクラスタ棒グラフを作成し、メイン [オプション] ダイアログ ボックスの [グラフ] タブで **[色のみを変更]** または **[模様のみを変更]** を選択すると、[グループ化されたグラフ] リストの最初の 2 つの色が新しい図表の棒の色として使用されます。

色を使用する順序を変更するには

▶ **[単純なグラフ]** を選択し、カテゴリを持たない図表に使用する色を選択します。

▶ **[グループ化されたグラフ]** を選択し、カテゴリを持つ図表に使用する色の順序を変更します。カテゴリの色を変更するには、カテゴリを選択した後、そのカテゴリの色をパレットから選択します。

オプションとして、次の選択が可能です。

- 選択したカテゴリの上に新しいカテゴリを挿入する。
- 選択したカテゴリを移動する。
- 選択したカテゴリを削除する。
- 順序を元の順序に戻す。
- 色を編集する (色のボックスを選択し、**[編集]** をクリックします)。

データ要素 線

新しい図表内の線データ要素に使用するスタイルの順序を指定します。線の種類は、図表に線データ要素が含まれるときに、メイン [オプション] ダイアログ ボックスの [グラフ] タブにある [スタイル変更の参照] グループで名前に **模様**を含むオプションを選択すると使用されます。

たとえば、2 つのグループを使用した折れ線グラフを作成し、メイン [オプション] ダイアログ ボックスの [グラフ] タブで **[模様のみを変更]** を選択すると、[グループ化されたグラフ] リストの最初の 2 つのスタイルが新しい図表の線の模様として使用されます。

線の種類を使用する順序を変更するには

- [**単純なグラフ**] を選択し、カテゴリを持たない折れ線グラフに使用する線の種類を選択します。

- [**グループ化されたグラフ**] を選択し、カテゴリを持つ折れ線グラフに使用する模様の順序を変更します。カテゴリの線の種類を変更するには、カテゴリを選択した後、そのカテゴリの線の種類をパレットから選択します。

  オプションとして、次の選択が可能です。
  - 選択したカテゴリの上に新しいカテゴリを挿入する。
  - 選択したカテゴリを移動する。
  - 選択したカテゴリを削除する。
  - 順序を元の順序に戻す。

データ要素 マーカー

新しい図表内のマーカー データ要素に使用するシンボルの順序を指定します。マーカーの種類は、図表にマーカー データ要素が含まれるときに、メイン [オプション] ダイアログ ボックスの [グラフ] タブにある [スタイル変更の参照] グループで名前に **模様**を含むオプションを選択すると使用されます。

たとえば、2 つのグループを使用した散布図を作成し、メイン [オプション] ダイアログ ボックスの [グラフ] タブで [**模様のみを変更**] を選択すると、[グループ化されたグラフ] リストの最初の 2 つのシンボルが新しい図表のマーカーとして使用されます。

マーカーの種類を使用する順序を変更するには

- [**単純なグラフ**] を選択し、カテゴリを持たない図表に使用するマーカー シンボルを選択します。

- [**グループ化されたグラフ**] を選択し、カテゴリを持つ図表に使用する模様の順序を変更します。カテゴリのマーカー シンボルを変更するには、カテゴリを選択した後、そのカテゴリのシンボルをパレットから選択します。

  オプションとして、次の選択が可能です。
  - 選択したカテゴリの上に新しいカテゴリを挿入する。
  - 選択したカテゴリを移動する。
  - 選択したカテゴリを削除する。
  - 順序を元の順序に戻す。

データ要素 塗りつぶし

新しい図表内の棒および面データ要素に使用する塗りつぶしスタイルの順序を指定します。塗りつぶしスタイルは、図表に棒または面データ要素が含まれるときに、メインの [図表オプション] ダイアログ ボックスの [グラフ] タブにある [スタイル変更の参照] グループで名前に **模様**を含むオプションを選択すると使用されます。

たとえば、2 つのグループを使用したクラスタ棒グラフを作成し、メイン [オプション] ダイアログ ボックスの [グラフ] タブで **[模様のみを変更]** を選択すると、[グループ化されたグラフ] リストの最初の 2 つのスタイルが新しい図表の棒の塗りつぶし模様として使用されます。

塗りつぶしスタイルを使用する順序を変更するには

- **[単純なグラフ]** を選択し、カテゴリを持たない図表に使用する塗りつぶし模様を選択します。
- **[グループ化されたグラフ]** を選択し、カテゴリを持つ図表に使用する模様の順序を変更します。カテゴリの塗りつぶし模様を変更するには、カテゴリを選択した後、そのカテゴリの塗りつぶし模様をパレットから選択します。

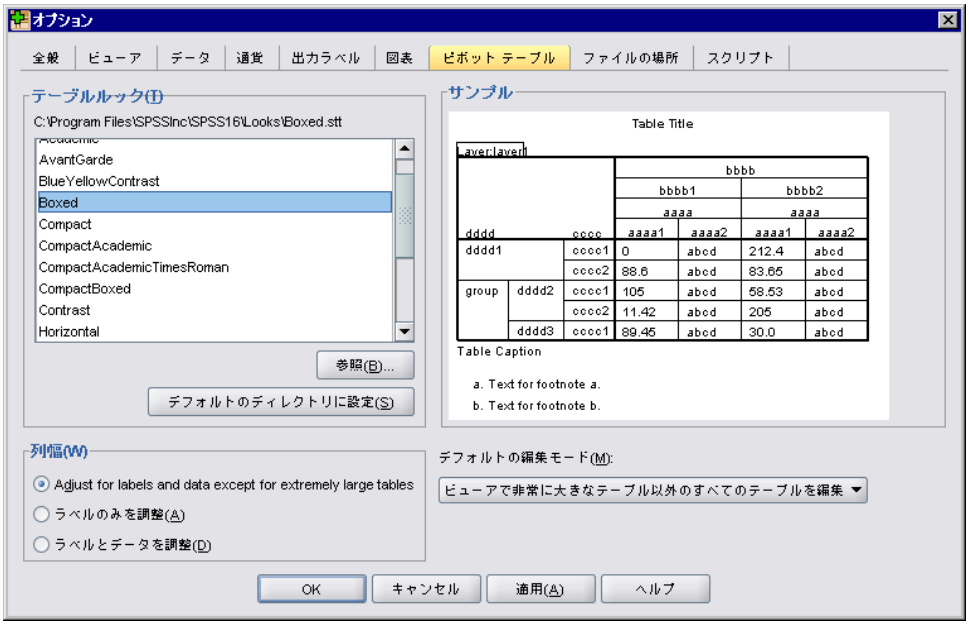
オプションとして、次の選択が可能です。

- 選択したカテゴリの上に新しいカテゴリを挿入する。
- 選択したカテゴリを移動する。
- 選択したカテゴリを削除する。
- 順序を元の順序に戻す。

[ピボット テーブル] オプション

ピボット テーブル オプションは、新規ピボット テーブル出力に使用するデフォルトのテーブルルックを設定します。テーブルルックは、グリッド線の表示と幅、フォント スタイル、大きさと色、および背景色を含む、各種のピボット テーブルの属性をコントロールできます。

図 45-8
[オプション] ダイアログ ボックス: [ピボット テーブル] タブ



テーブルルック。ファイルのリストからテーブルルックを選択し、[OK] または [**適用**] をクリックします。SPSS と共に提供されるテーブルルックの 1 つを使用することもできますが、ピボット テーブル エディタ ([**書式**] メニューの [**テーブルルック**] を選択します) で独自のものを作成することもできます。

- **参照。**別のディレクトリからテーブルルックを選択できます。
- **デフォルトのディレクトリに設定。**デフォルトのテーブルルックのディレクトリを変更できます。

注:SPSS の以前のバージョンで作成されたテーブルルックは、バージョン 16.0 以降では使用できません。

列幅の調整。ピボット テーブルでの列幅の自動調整をコントロールします。

- **ラベルのみ。**列幅を列ラベルの幅に調整します。こうすると、コンパクトなテーブルが作成されますが、ラベルよりも広いデータ値は表示されません (アスタリスクは、値の幅が広すぎて表示できないことを示しています)。
- **ラベルとデータ。**列幅を、列ラベルまたは最大データ値のどちらか大きい方に調整します。こうすると、幅の広い表が作成されますが、すべての値が表示されるようになります。

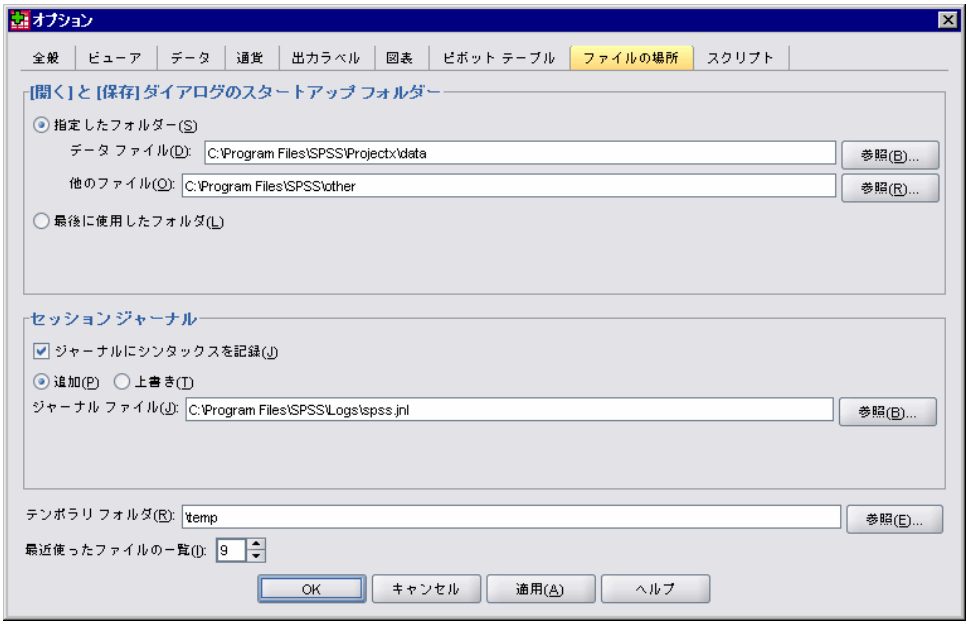
デフォルトの編集モード。ピボットテーブルをビューア ウィンドウで開くか、別のウィンドウで開くかを制御します。デフォルトでは、ピボットテーブルをダブルクリックすると、ビューア ウィンドウに非常に大きなテーブルが開かれます。ピボットテーブルを別のウィンドウで開くことを選択したり、小さなピボットテーブルは

ビューア ウィンドウで、大きなピボットテーブルは別のウィンドウで開くようにサイズ設定を選択することもできます。

ファイル位置オプション

[ファイル位置] タブのオプションは、アプリケーションが各セッションの開始時にファイルを開けたり、保存したりするために使用するデフォルト位置、ジャーナル ファイルの位置、テンポラリ フォルダの位置、最近使用したファイル リストに表示されるファイル数を制御します。

図 45-9
[オプション] ダイアログ ボックス: [ファイル位置] タブ



ダイアログ ボックスを開いたり保存したりするための [スタートアップ] フォルダ

- **指定したフォルダ。**指定したフォルダは各セッションの開始時のデフォルト位置として使用されます。データおよび他の (データ以外) ファイルの別のデフォルト位置を指定できます。
- **最近使用されたフォルダ。**前のセッションでファイルを開いたり、保存するために最近使用されたフォルダは、次のセッションの開始時にデフォルトとして使用されます。これはデータおよび他の (データ以外) ファイルの両方に適用されます。

これらの設定はファイルを開いたり、保存したりするダイアログ ボックスのみに適用され、「最近使用したフォルダ」はファイルを開いたり、保存したりする最近のダイアログ ボックスにより決定されます。コマンド シンタックスを通して開いた

り、保存したファイルは、これらの設定に影響を及ぼすことはありませんし、これらの設定により影響を受けません。

セッション ジャーナル

セッションで実行されるコマンドを自動的にレコードするセッション ジャーナルを使用できます。実行されるコマンドとしては、シンタックス ウィンドウへ入力、実行するコマンド、およびダイアログ ボックスの選択で生成されるコマンドがあります。ジャーナル ファイルを編集したり、他のセッションでそれらのコマンドを再び使用したりできます。ジャーナルはオン、オフすることもでき、追加または上書きもでき、さらにジャーナル ファイル名と位置を選択することもできます。ジャーナル ファイルからコマンド シンタックスをコピーし、シンタックス ファイルに保存できます。

テンポラリ フォルダ

セッション中に作成されるテンポラリ ファイルの場所を指定します。ディストリビュートモード (サーバー バージョンで使用できます) では、この指定はテンポラリ データ ファイルの格納場所には影響しません。ディストリビュート モードでは、テンポラリ データ ファイルの格納場所は SPSSTMPDIR 環境変数で指定され、SPSS サーバー バージョンが稼働しているコンピュータでのみ設定できます。テンポラリ ディレクトリの場所を変更する必要がある場合には、システム管理者にお尋ねください。

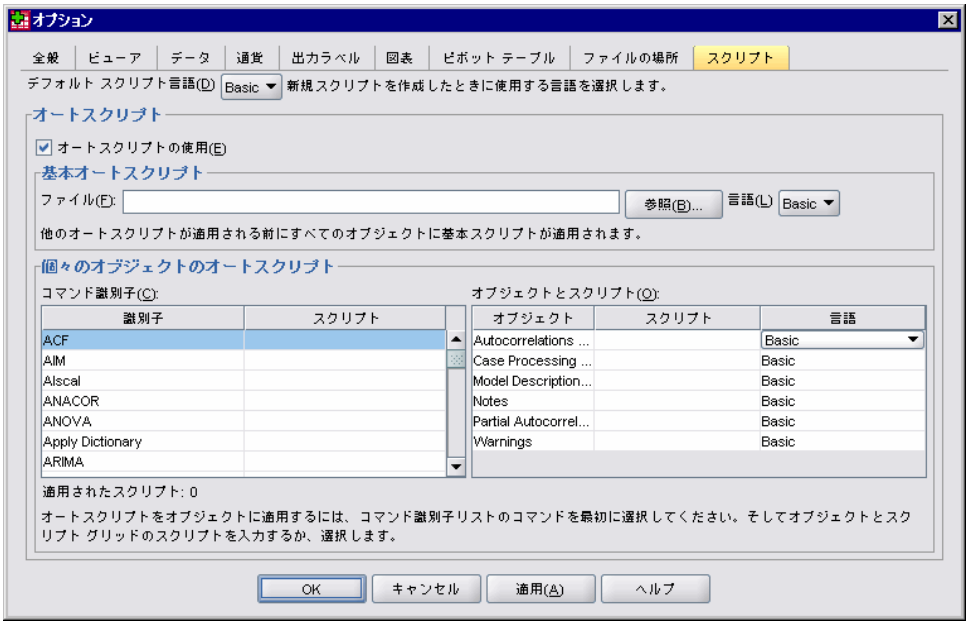
最近使ったファイルの一覧

[ファイル] メニュー内に表示される最近使用したファイルの数をコントロールします。

[スクリプト] オプション

使用したいデフォルト スクリプト言語およびオートスクリプトを指定する [スクリプト] タブを使用します。スクリプトを使用すると、ピボット テーブルのカスタマイズを含め、SPSS の多くの機能を自動化することができます。

図 45-10
[オプション] ダイアログ ボックス: [スクリプト] タブ



注: 従来の Sax Basic のユーザーは、ユーザー指定のオートスクリプトを手動で変換する必要があります。16.0 以前のバージョンと共にインストールされたオートスクリプトは、SPSS がインストールされているディレクトリの「**サンプル**」サブディレクトリに位置する別のスクリプト ファイルのセットとして利用できます。デフォルトでは、出力項目はオートスクリプトとは関係ありません。次に説明するように、すべてのオートスクリプトは望む出力項目を必ず手動で関連付けてください。

デフォルト スクリプト言語。デフォルト スクリプト言語は、新規スクリプトが作成されたときに起動するスクリプト エディタを決定します。実行形式ファイルがオートスクリプトを実行するために使用されるデフォルト言語を指定します。利用可能なスクリプト言語はプラットフォームに応じて異なります。Windows では、使用可能なスクリプトは、Base システムち共にインストールされた Basic および Python プログラム言語です。他のプラットフォームでは、スクリプトは Python プログラム言語です。

注: Python プログラミング言語でスクリプトを作成できるようにするには、別途 Python と SPSS-Python Integration Plug-in をインストールする必要があります。Python プログラミング言語は Base システムと一緒にインストールされませんが、Windows 用のインストール CD に含まれています。SPSS-Python Integration Plug-in は、すべてのプラットフォーム用のインストール CD に含まれています。

オオートスクリプトトの使用。チェック ボックスでオートスクリプトを有効にするか無効にできます。デフォルトでは、オートスクリプトは有効です。

オートスクリプト。他のオートスクリプトが適用される前に、オプションのスクリプトがすべての新規の [ビューア] オブジェクトに適用されます。実行形式ファイルがスクリプトを実行するために使用される言語と同じく基本オートスクリプトとして使用されるスクリプトを指定します。

オートスクリプトを出力項目に適用するには

- コマンド識別子グリッド内で、オートスクリプトが適用される出力項目を生成するコマンドを選択します。

  オブジェクトおよびスクリプト グリッドでは、**オブジェクト** 列は、選択したコマンドと関連するオブジェクトの一覧を表示します。**スクリプト** 列は選択したコマンドの既存のスクリプトを表示します。

- **オブジェクト** 列に表示された項目のスクリプトを指定します。対応する [**スクリプト**] セルをクリックします。スクリプトを参照するには、スクリプトへのパスを入力するか、省略符号 (...) をクリックしてください。

- 実行形式ファイルがスクリプトを実行するために使用される言語を指定します。**注:** 選択された言語は、デフォルト スクリプト言語を変更することにより影響を受けます。

- [OK] または [**適用**] をクリックします。

オートスクリプト連関を削除するには

- オブジェクトおよびスクリプト グリッドで、関連付けを解くには、スクリプトと対応するスクリプト列のセルをクリックします。

- スクリプトへのパスを削除して、オブジェクトとスクリプト グリッドの他のセルをクリックします。

- [OK] または [**適用**] をクリックします。

スクリプト オプションを設定するには

メニューから次の項目を選択します。

編集
オプション

- [スクリプト] タブをクリックします。

- 必要な設定を選択します。

- [OK] または [**適用**] をクリックします。

章
46

メニューとツールバーのカスタマイズ

メニュー エディタ

メニュー エディタを使用して、メニューをカスタマイズできます。メニュー エディタでは、次の操作ができます。

- カスタマイズされたスクリプトを実行するメニュー項目を追加する。
- コマンド シンタックス ファイルを実行するメニュー項目を追加する。
- 他のアプリケーションを起動して自動的にデータを他のアプリケーションに送るメニュー項目を追加する。

他のアプリケーションにデータを送信するときに使用できるのは、SPSS、Excel、Lotus 1-2-3、タブ区切り、dBASE IV の各書式です。

メニューに項目を追加するには

▶ メニューから次の項目を選択します。
分析
メニュー エディタ...

▶ [メニュー エディタ] ダイアログ ボックスで、新規項目を追加するメニューをダブルクリック (またはアイコンの + サインをクリック) します。

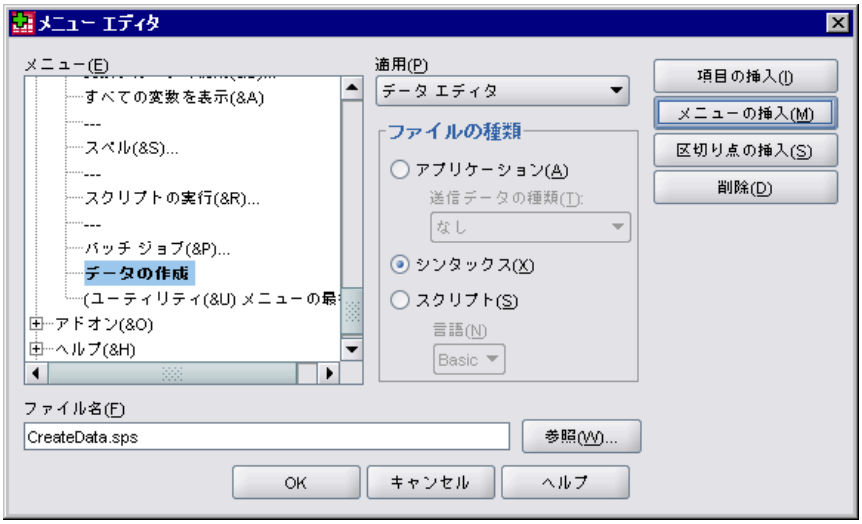
▶ 新しいメニュー項目を挿入する位置のメニュー項目を選択します。

▶ **[項目の挿入]** をクリックして、新しいメニュー項目を挿入します。

▶ この新しい項目に対するファイルの種類 (スクリプト ファイル、コマンド シンタックス ファイル、または外部アプリケーション) を選択します。

▶ **[参照]** をクリックして、メニュー項目に添付するファイルを選択します。

図 46-1
[メニュー エディタ] ダイアログ ボックス



メニュー項目の間に、まったく新しいメニューとセパレータを追加することもできます。オプションとして、メニュー上で別のアプリケーションを選択すると、データ エディタの内容を自動的にそのアプリケーションに送ることもできます。

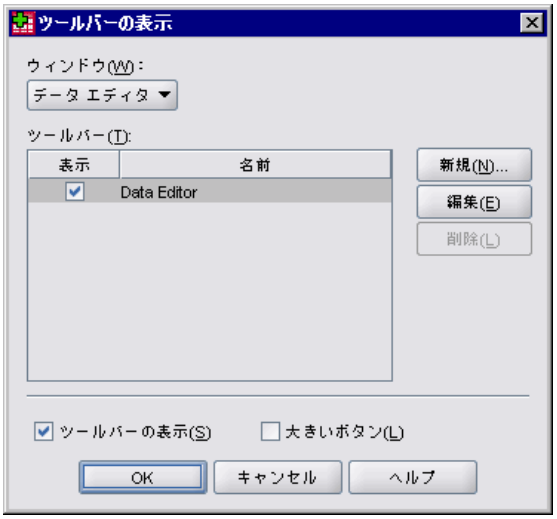
ツールバーのカスタマイズ

ツールバーをカスタマイズして、新しいツールバーを作成できます。メニューの全アクション用のツールをはじめ、使用できるツールならどれでもツールバーに含めることができます。さらに、他のアプリケーションを起動させるツール、コマンド シンタックス ファイルを実行するツール、およびスクリプト ファイルを実行するツールなどのカスタム ツールを含めることもできます。

ツールバーの表示

[ツールバーの表示] は、ツールバーの表示/非表示の切り替え、ツールバーのカスタマイズ、新しいツールバーの作成に使用します。メニューの全アクション用のツールをはじめ、使用できるツールならどれでもツールバーに含めることができます。さらに、他のアプリケーションを起動させるツール、コマンド シンタックス ファイルを実行するツール、およびスクリプト ファイルを実行するツールなどのカスタム ツールを含めることもできます。

図 46-2
[ツールバーの表示] ダイアログ ボックス



ツールバーをカスタマイズするには

- メニューから次の項目を選択します。

  表示
  ツールバー
  カスタマイズ

- カスタマイズするツールバーを選択し [**編集**] をクリックするか、[**新しいツールバー...**] を選択して新しいツールバーを作成します。
- 新しいツールバーについて、ツールバーの名前を入力し、このツールバーを表示するウィンドウを選択した後、[**編集**] をクリックします。
- カテゴリリストで項目を選択し、そのカテゴリで使用できるツールを表示します。
- ダイアログ ボックスに表示されたツールバーに目的のツールをドラッグ アンド ドロップします。
- ツールをツールバーから削除するには、ダイアログ ボックス内に表示されたツールバーから目的のツールを外にドラッグします。

  ファイルを開いたり、コマンド シンタックス ファイルまたはスクリプトを実行するためのカスタム ツールを作成するには、次の操作を行います。

- [ツールバーの編集] ダイアログ ボックスの [**新しいツール**] をクリックします。
- ツールを説明するラベルを入力します。
- このツールで行うアクション (ファイルを開く、コマンド シンタックス ファイルまたはスクリプトを実行する) を選択します。

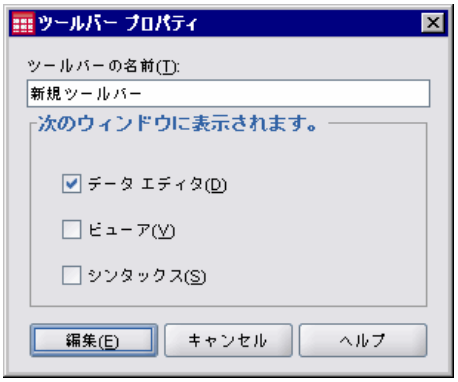
- [参照] をクリックして、このツールに関連付けるアプリケーションまたはファイルを選択します。

  新しいツールは、ユーザー定義のメニュー項目も含む [ユーザー定義] カテゴリに表示されます。

ツールバーのプロパティ

[ツールバー プロパティ] は、選択したツールバーを表示するウィンドウの種類を選択するときに使用します。このダイアログ ボックスは、新しいツールバーに名前を付けるときにも使用します。

図 46-3
[ツールバー プロパティ] ダイアログ ボックス



ツールバーのプロパティを設定するには

- メニューから次の項目を選択します。

  表示
  ツールバー
  カスタマイズ

- 既存のツールバーの場合は、[編集] をクリックし、その後 [ツールバーの編集] ダイアログ ボックスの [プロパティ] をクリックします。
- 新しいツールバーの場合は、[新しいツールバー] をクリックします。
- ツールバーを表示するウィンドウの種類を選択します。新しいツールバーの場合は、ツールバー名も入力します。

[編集] ツールバー

[ツールバーの編集] ダイアログ ボックスは、既存のツールバーのカスタマイズと新しいツールバーの作成に使用します。メニューの全アクション用のツールをはじめ、使用できるツールならどれでもツールバーに含めることができます。

さらに、他のアプリケーションを起動させるツール、コマンド シンタックス ファイルを実行するツール、およびスクリプト ファイルを実行するツールなどのカスタム ツールを含めることもできます。

図 46-4
[ツールバーのカスタマイズ] ダイアログ ボックス

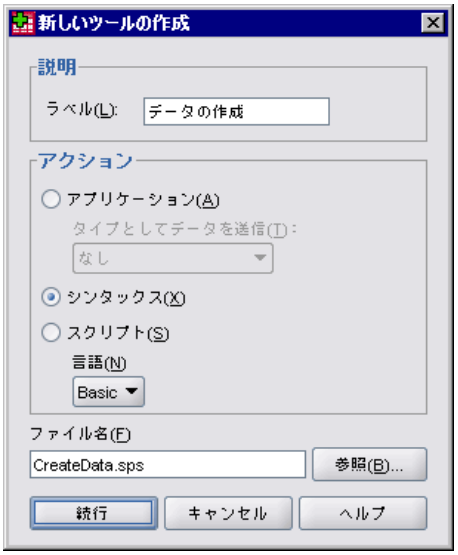
ツールバーのイメージを変更するには

- ツールバーの表示のイメージうぃ変更したいときはツールを選択します。
- [**イメージ変更**] をクリックします。
- ツールに使用したいイメージ ファイルを選択します。利用可能なイメージ形式は、BMP、PNG、GIF および JPG です。
  - イメージは正方形でなければなりません。正方形でないイメージは正方形領域にクリップされます。
  - イメージに合せて自動的に表示。オプションの表示では、小さいツールバーでは 16x16 ピクセルまたは大きいツールバーでは 32x32 ピクセルです。

新しいツールの作成

[新しいツールの作成] ダイアログ ボックスは、他のアプリケーションの起動、コマンド シンタックス ファイルの実行、スクリプト ファイルの実行のためのカスタム ツールを作成するときに使用します。

図 46-5
[新しいツールの作成] ダイアログ ボックス



章
47

バッチ ジョブ

バッチ ジョブでは、SPSSを自動的に実行することができます。プログラムは自動的に実行され、最後のコマンドの実行後に終了するため、実行中に他の作業を行うことができるほか、バッチ ジョブを特定の時間に自動的に実行するようスケジュールすることもできます。バッチ ジョブは、週ごとのレポートのような、時間のかかる一連の分析を繰り返し実行することが多い場合に役立ちます。

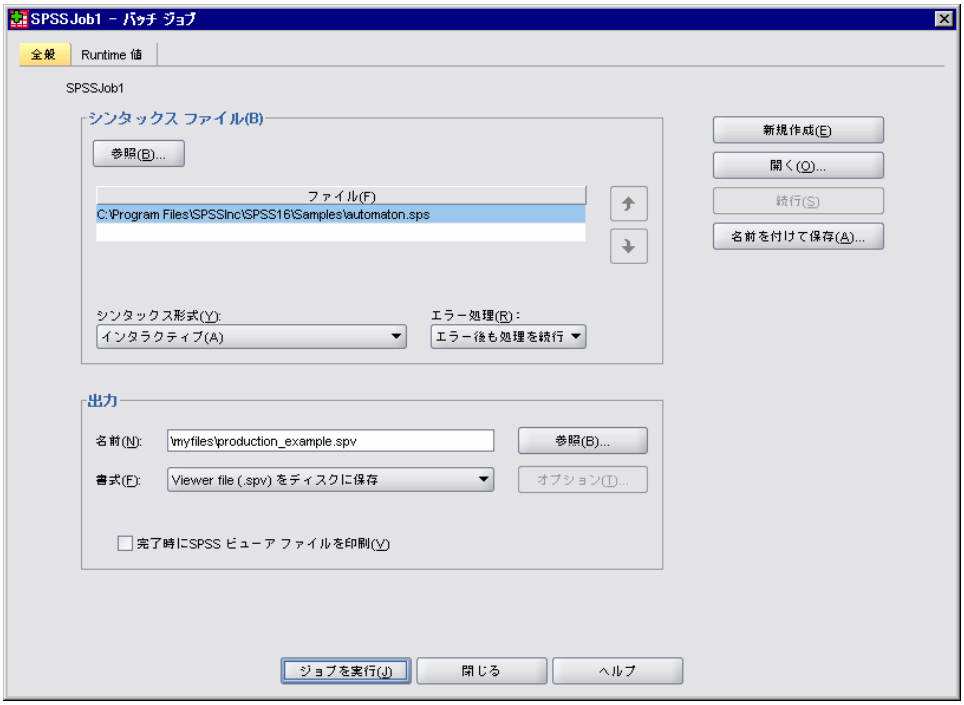
バッチ機能では、コマンド シンタックス ファイルを使って、何をすべきかを SPSS に指示します。コマンド シンタックス ファイルはコマンド シンタックスを記述した単純なテキスト ファイルです。コマンド シンタックス ファイルは、適当なテキスト エディタを使用して作成します。また、コマンド シンタックスの作成は、ダイアログ ボックスで選択したものをシンタックス ウィンドウに貼り付けるか、ジャーナル ファイルを編集しても行えます。 詳細は、275 ページの 12 章のコマンド シンタックスを使用する作業を参照してください。

バッチ ジョブを作成するには、次の手順を実行します。

▶ 開いているウィンドウのメニューから、次の項目を選択します。

ユーティリティ
バッチ ジョブ

図 47-1
[バッチ ジョブ] ダイアログ ボックス



▶ [新規] をクリックして新しいバッチ ジョブを作成するか、既存のバッチ ジョブを開くには、[開く] をクリックします。

注: 16.0 より前のバージョンで作成されたバッチ機能ジョブ ファイル (.spp) は 16.0 以降では使用できません。Windows と Macintosh バッチ機能ジョブ ファイルをバッチ ジョブ (.spj) に変換するには変換ユーティリティを使用してください。詳細は、529 ページのバッチ ジョブ ファイル変換を参照してください。

▶ 1 つ以上のコマンド シンタックス ファイルを選択します。

注: ディストリビュート アナリシス モードを使用する場合、シンタックス ファイルの場所は手動で入力されなければいけなく、相対パスは リモート サーバーに関係します。詳細は、81 ページの 4 章の絶対パスと相対パス指定を参照してください。

▶ 出力ファイル名、場所、形式を選択します。

▶ バッチ ジョブを保存するには、[保存] または [名前を付けて保存] をクリックします。

シンタックス形式。ジョブに使用されるシンタックス ルールの形式を制御します。

- **インタラクティブ。**各コマンドはピリオドで終わる必要があります。ピリオドはコマンド内のどの位置でも使用でき、コマンドは複数の行にわたって継続できます。ただし、ピリオドが行の最後の空白以外の文字として使用されている場合、コマンドの終点として解釈されます。継続行と新しいコマンドは、新しい行

の任意の位置で開始できます。これらは、シンタックス ウィンドウでコマンドを選択および実行する際に有効になる、「インタラクティブな」ルールです。

- **バッチ学習。**各コマンドは、新しい行の始点で開始する必要があります (コマンドの始点の前に空白スペースを入れません)。継続行は、スペースを 1 つ以上入れてインデントする必要があります。新しいコマンドをインデントする場合は、行の始点で最初の文字としてプラス記号、ダッシュ、またはピリオドを使用して、実際のコマンドをインデントできます。コマンドの最後のピリオドは、付けても付けなくても構いません。この設定は、INCLUDE コマンドによって含まれるコマンド ファイルのシンタックス ルールに相当します。

  注: シンタックス ファイルが、GPL 構文を含む GGRAPH コマンド シンタックスを含んでいる場合はバッチ オプションを使用しては**いけません**。GPL 構文はインタラクティブ ルールでは実行できません。

エラー処理。ジョブのエラー条件の処理を制御します。

- **エラーのあと処理を続行。**ジョブのエラーにより、自動的にコマンドの処理が停止されることはありません。バッチ ジョブ ファイルのコマンドは通常のコマンド ストリームの一部として処理され、コマンドの処理は通常どおりに継続されます。
- **処理を即座に停止。**バッチ ジョブ ファイルでの最初のエラーの発生時に、コマンドの処理が停止されます。これは、INCLUDE コマンドによって含まれるコマンド ファイルの動作に相当します。

出力。バッチ ジョブの結果の名前、場所、および形式を制御。次の形式オプションを使用できます。

- **ディスクへのビューア ファイル (.spv)。**結果は、指定したファイル場所の SPSS ビューア形式に保存されます。
- **PES レポジトリへのビューア ファイル (.spv)。**これをするには、SPSS Adaptor for Enterprise Services オプションが必要です。
- **PES レポジトリへの Web 報告書 (.spv)。**これをするには、SPSS Adaptor for Enterprise Services オプションが必要です。
- **Word/RTF (*.doc)。**ピボット テーブルは、セルの罫線、フォント スタイル、背景色など、あらゆる書式属性とともに、Word テーブルとしてエクスポートされます。テキスト出力は書式付きの RTF としてエクスポートされます。Windows オペレーティング システム上では、図表は EMF (拡張メタファイル) 形式でドキュメントに挿入されます。他のオペレーティング システムでは、図表は PNG 形式で挿入されます。テキスト出力は、常に固定ピッチ (等幅) フォントで表示され、同じフォント属性のままでエクスポートされます。スペース区切りテキスト出力を正しく位置合わせするには、固定ピッチ フォントが必要です。

  注: Microsoft Word では、横幅が極端に長い表は表示されないことがあります。
- **Excel (*.xls)。**ピボット テーブルの行、列、セルは、Excel の行、列、セルとしてエクスポートされます。この場合、セルの罫線、フォント スタイル、背景色など、すべての書式属性も反映されます。テキスト出力は、すべてのフォント属性とともにエクスポートされます。テキスト出力の各行は、Excel ファイル

の行になります。この場合、1 行の内容はすべて 1 つのセルに入れられます。図表は含まれません。

- **HTML (*.htm)。**ピボット テーブルは HTML テーブルとしてエクスポートされます。テキスト出力は、事前にフォーマットされた HTML としてエクスポートされます。図表は参照により埋め込まれます。HTML 文書を含む適当な形式で図表をエクスポートする必要があります (たとえば、PNG および JPEG)。
- **Portable Document Format (*.pdf)。**出力はすべて、[印刷プレビュー] に表示されるとおりにエクスポートされます。またその際、すべての書式属性は維持されます。
- **PowerPoint ファイル (*.ppt)。**ピボット テーブルは Word テーブルとしてエクスポートされ、PowerPoint ファイルへはそれぞれのピボット テーブルが個別のスライドへ 1 つずつ埋め込まれます。ピボット テーブルは、セルの罫線、フォント スタイル、背景色など、あらゆる書式属性が保持されます。図表は TIFF 形式でエクスポートされます。テキスト出力は含まれません。

  注: PowerPoint へのエクスポートは、Windows オペレーティング システム上のみで利用でき、Student 版では利用できません。
- **テキスト (*.txt)。**出力形式は、テキスト形式、UTF-8、および UTF-16 を含みます。ピボット テーブルは、タブ区切り書式またはスペース区切り書式でエクスポートできます。すべてのテキスト出力はスペース区切り書式でエクスポートされます。図表では、各図表に対してエクスポートする図表のファイル名を示す行がテキスト ファイルに挿入されます。

SPSS ビューア ファイル印刷完了。バッチ ジョブの終了時に、最終ビューア出力ファイルをプリンタに自動的に送ります。

ジョブを実行。異なるセッションのバッチ ジョブを実行します。展開する前に新規バッチ ジョブの検定には非常に有効です。

OUTPUT コマンドとバッチ ジョブ

SPSS の OUTPUT コマンド (OUTPUT SAVE、OUTPUT ACTIVATE、OUTPUT NEW など) は、バッチ ジョブに優先して実行されます。バッチ ジョブの実行中に OUTPUT SAVE コマンドを実行すると、指定した出力文書の内容が指定した場所に書き込まれます。さらにこれはバッチ ジョブにより作成された出力ファイルです。OUTPUT NEW を使用して新しい出力文書を作成する場合は、OUTPUT SAVE コマンドを使用して文書を明示的に保存することをお勧めします。

バッチ ジョブによる出力ファイルには、ジョブ終了時点でのアクティブな出力文書の内容が含まれています。OUTPUT コマンドを含むジョブの場合、セッション中に作成された出力の一部が出力ファイルに含まれない場合があります。たとえば、バッチ ジョブに多数の SPSS 手続きが記述され、その後に OUTPUT NEW コマンドが 1 つ記述されているとします。その後さらに SPSS 手続きが続きますが、OUTPUT コマンドは出現しません。この OUTPUT NEW コマンドは、新しいアクティブな出力文書を定義します。バッチ ジョブの終了時、出力文書には OUTPUT NEW コマンド以降に実行された手続きによる出力しか含まれません。

HTML オプション

テーブル オプション。HTML 形式ではテーブル オプションは利用できません。ピボット テーブルは HTML テーブルに変換されます。

イメージ オプション。使用できるイメージ書式は EPS、JPEG、TIFF、PNG および BMP です。Windows オペレーティング システムでは、EMF (拡張メタファイル) 形式も利用できます。イメージのサイズは 1% から 200% までの間で変更できます。

PowerPoint オプション

テーブル オプション。スライド表題としてビューアのアウトラインのエントリを使用できます。各スライドは単一出力項目を含んでいます。表題はビューアのアウトラインのウィンドウ枠のアウトライン項目から作成されます。

イメージ オプション。イメージのサイズは 1% から 200% までの間で変更できます。(すべてのイメージはは TIFF 形式で PowerPoint にエクスポートされます。)

注: PowerPoint 形式は、Windows オペレーティング システムでのみ使用でき、PowerPoint 97 以降が必要です。

PDF オプション

ブックマークを埋め込む。このオプションを選択すると、ビューアのアウトラインのエントリに対応するブックマークが PDF 文書に挿入されます。ブックマークを埋め込むことにより、ビューアのアウトライン ウィンドウ枠と同様に、多数の出力オブジェクトを含む文書内の移動が簡単になります。

フォントを埋め込む。フォントを埋め込むと、PDF 文書をどのコンピュータでも同じように表示させることができます。フォントが埋め込まれていない場合、PDF 文書を表示 (または印刷) するとき、コンピュータで利用できない一部のフォントに対して代用フォントが使用されます。ただし、この場合は、最適な結果が得られないことがあります。

テキスト オプション

テーブル オプション。ピボット テーブルは、タブ区切り書式またはスペース区切り書式でエクスポートできます。スペース区切り書式では、次のことがコントロールできます。

- **列幅。**[Autofit] は、表示内容を折り返さなく、各列は列の最大幅のラベルまたは値と同じ幅です。[**カスタム**] は、列の次の行に折り返す幅を超える値およびテーブルのすべての列に適用される最大列幅を設定します。
- **行と列の罫線。**行と列の罫線を作成するのに使用された文字を制御します。行と列の罫線を非表示にするには、値にスペースを入力します。

イメージ オプション。使用できるイメージ書式は EPS、JPEG、TIFF、PNG および BMP です。Windows オペレーティング システムでは、EMF (拡張メタファイル) 形式も利用できます。イメージのサイズは 1% から 200% までの間で変更できます。

ランタイム値

ランタイム値はバッチ ジョブ ファイルで定義し、コマンド シンタックス ファイルで使用します。ランタイム値を使用すると、たとえば、異なったデータ ファイルを使用して同じような分析を実行したり、異なった変数の組み合せで同じコマンドを実行することが簡単になります。たとえば、ランタイム値 @datafile を定義し、バッチ ジョブ を実行するたびにデータ ファイル名を要求するプロンプトを表示させることができます。このようにするには、文字列 @datafile をファイル名の代わりにコマンド シンタックス ファイルで使用します。

図 47-2
[バッチ ジョブ] ダイアログ ボックス、[ランタイム値] タブ

マーカー。コマンド シンタックス ファイルの文字列は値のユーザーをプロンプトするバッチ ジョブを起動します。シンボル名はすべて @ で始まり、変数の命名規則に従う必要があります。 詳細は、85 ページの 5 章の変数名を参照してください。

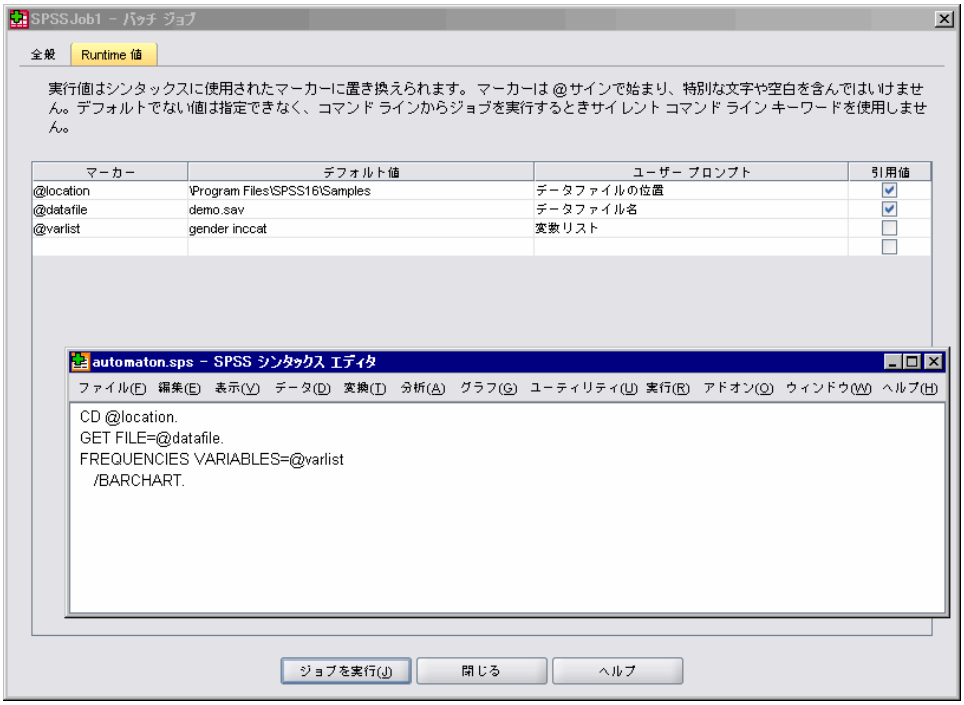
デフォルト値。別の値を入力しなかった場合にデフォルトによりバッチ ジョブが使用する値。この値はバッチ ジョブが情報を要求するプロンプトを行う場合に表示されます。実行時に値を置き換えたり、変更することができます。デフォルト

値を提供しないときは、ランタイム値を指定するために -symbol 切り替えを使用しない限り、コマンド ライン切り替えのあるバッチ ジョブを実行しているときは silent キーワードを使用してはいけません。詳細は、528 ページのコマンド ラインからのバッチ ジョブの実行を参照してください。

ユーザー プロンプト。記述的なラベルで、バッチ ジョブが情報の入力を要求する際に表示されます。たとえば、データ ファイル名を必要とするフィールドを示すために、「使用するデータ ファイル名」という記述ラベルを使用することができます。

引用値。引用符にユーザーにより入力された値またはデフォルト値を入れます。たとえば、ファイル指定は必ず引用符で囲みます。

図 47-3
コマンド シンタックス ファイル内のランタイム マーカー



ユーザー プロンプト

バッチ機能は、実行するバッチ ジョブに定義されたランタイム記号があると、値を要求するプロンプトを表示します。表示されているデフォルト値を置き換えたり、変更することができます。それらの値はそのバッチ ジョブに関連する全コマンド シンタックス ファイルのランタイム記号を入力された値で置き換えます。

図 47-4
[バッチ ユーザー プロンプト] ダイアログ ボックス

コマンド ラインからのバッチ ジョブの実行

コマンド ラインのスイッチでバッチ ジョブをスケジュールしておき、一定の時間に自動的に実行することができます。これにはオペレーティング システムにあるようなスケジューリング ユーティリティを使用します。コマンド ライン引数の基本書式は次のとおりです。

```
spss filename.spj -production
```

バッチ ジョブをどのように使用するかにより、spss 実行ファイル (アプリケーションがインストールされたディレクトリにある) そして/またはバッチ ジョブ ファイルのディレクトリ パスを含む必要があります。

バッチ ジョブをコマンド ラインから実行する場合に次のスイッチを使用します。

-production [prompt|silent]。バッチ モードのアプリケーションを開始します。prompt および silent キーワードは、ジョブ内で指定されたらランタイム値をプロンプトするダイアログ ボックスを表示するかを指定します。prompt キーワードはデフォルトでダイアログ ボックスを表示します。silent キーワードはダイアログ ボックスを抑制します。silentキーワードを使用すると、-symbol スイッチを持つランタイム記号を定義できます。または、デフォルト値を使用します。-switchserver および -singleseat スイッチは、-production スイッチを使用すると無視されます。

-symbol <values>。バッチ ジョブに使用された記号値ペアのリスト。各記号名は、@ で始まります。スペースを含む値は必ず引用符で囲みます。文字列リテラル中にアポストロフィや引用符を含めるときの規則はオペレーティング システムによって異なりますが、一重引用符を含む文字列を二重引用符で囲むことは通常有効です (「"'a quoted value'"」など)。

ディストリビュート アナリシス モードのリモート サーバーのバッチ ジョブを実行するには、サーバー ログイン情報を指定する必要があります。

-server <inet:hostname:port >。サーバーの名前または IP アドレスとポート番号。Windows のみ。

-user <name>。有効なユーザー名。ドメイン名を指定する必要がある場合は、ユーザー名の前にドメイン名とバックスラッシュ (¥) を記述します。Windows のみ。

-password <password>。ユーザー パスワード。Windows のみ。

例

```
spss \production_jobs\prodjob1.spj -production silent
  -symbol @datafile /data/July_data.sav
```

- この例は、インストール ディレクトリからコマンド ラインを実行していることを想定しているため、**spss** 実行ファイルのパスは必要ありません。
- この例は、バッチ ジョブが **@datafile** の値が引用符で囲まれることを想定しているため ([ランタイム値] タブの [引用値] チェックボックス)、コマンド ラインのデータ ファイルを指定するときに引用符は必要ありません。そうでない場合は、コマンド シンタックスによるとこの指定はで引用符で囲む必要があるため、`"'/data/July_data.sav'"` などと指定する必要があります。
- バッチ ジョブの場所のディレクトリ パスは、Windows バック スラッシュを使用します。Macintosh と Linux ではフォワード スラッシュ (/) を使用します。引用符で囲まれた文字列はコマンド シンタックス ファイルに挿入され、また、すべてのオペレーティング システムでフォワード スラッシュはファイル指定を含むコマンド (`GET FILE`、`GET DATA`、`SAVE` など) で受け入れられるため、引用符で囲まれたデータ ファイル指定のフォワード スラッシュは、すべてのオペレーティング システムで機能します。
- `silent` キーワードは、バッチ ジョブのユーザー プロンプトを押さえ、`-symbol` 切り替えがランタイム記号 **@datafile** がバッチ ジョブに含まれるコマンド シンタックス ファイルに表示される場所と引用されたデータ ファイル名に挿入されます。

バッチ ジョブ ファイル変換

16.0 より前のバージョンで作成されたバッチ機能ジョブ ファイル (.spp) は 16.0 以降では使用できません。前のバージョンで作成された Windows および Macintosh バッチ ジョブでは、これらのファイルを新規バッチ ジョブ ファイル (.spj) に変換するためにインストール ディレクトリにある **prodconvert** を使用できます。基本 (唯一の) 指定は以下の項目です。

```
prodconvert filename.spp
```

同じ名前と拡張子 **.spj** を持つ新規ファイルは元のファイルと同じディレクトリに作成されます。

制限

- WMF と EMF 図表形式はサポートされていません。PNG 形式がこれらの形式の場所で使用されます。
- エクスポート オプション **出力文書 (図表なし)**、**図表のみ**、そして **なし** はサポートされていません。選択された形式によりサポートされたすべての出力オブジェクトが含まれています。

- SPSS 表示のオプションとジョブの完了をオープンにすることはサポートされていません。バッチ セッションはジョブ完了後はいつも閉じられています。
- リモート サーバー設定は無視されます。ディストリビュート アナリシスのリモート サーバー設定を指定するには、サーバー設定を指定するためにコマンド ラインを使用しながらコマンド ラインのバッチ ジョブを実行する必要があります。 詳細は、528 ページのコマンド ラインからのバッチ ジョブの実行 を参照してください。
- Web に公開設定は無視されます。

章
48

出力管理システム

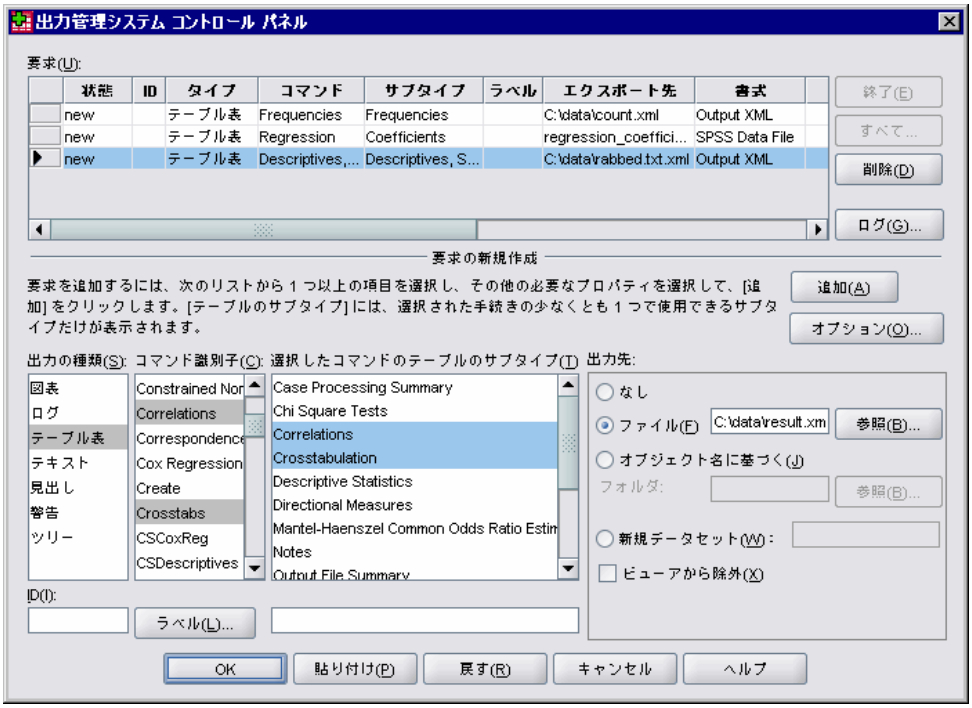
出力管理システム (OMS) は、選択したカテゴリの出力を別の出力ファイルに異なる形式で自動的に書き出す機能を備えています。形式は以下のファイルがあります: SPSS データ ファイル形式 (.sav)、SPSS ビューア ファイル形式 (.spv)、XML、HTML、およびテキスト。詳細は、538 ページのOMS オプションを参照してください。

出力管理システム コントロール パネルを使用するには、次の手順を実行します。

▶ メニューから次の項目を選択します。

ユーティリティ
OMS コントロール パネル

図 48-1
出力管理システム コントロール パネル



コントロール パネルを使用して、さまざまな出力先への出力送信の開始および停止できます。

- 各 OMS 要求は、明示的に終了するか、またはセッションが終了するまで、アクティブな状態が保持されます。
- OMS 要求で指定された出力先ファイルは、OMS 要求が終了されるまで、他の手続きおよび他のアプリケーションでは使用できません。
- OMS 要求がアクティブである間は、指定された出力先ファイルはメモリ (RAM) に格納されます。したがって、外部ファイルに大量の出力を書き込むアクティブな OMS 要求により、大量のメモリが消費される場合があります。
- 複数の OMS 要求は、互いに依存していません。同じ出力が、異なる OMS 要求の指定に基づいて、異なる形式で異なる場所に送られる場合があります。
- 特定の出力先での出力オブジェクトの順序は、作成された順序になっています。この順序は、出力を生成した手続きの順序と操作によって決定されます。

新しい OMS 要求の追加

▶ 含める出力の種類 (テーブル、図表など) を選択します。詳細は、534 ページの出力オブジェクトの種類を参照してください。

▶ 含めるコマンドを選択します。すべての出力を含める場合は、リスト内のすべての項目を選択します。詳細は、535 ページのコマンド識別子とテーブルのサブタイプを参照してください。

▶ ピボット テーブル出力を作成するコマンドに対しては、含める特定のテーブル タイプを選択します。

リストには、選択されたコマンドで使用できるテーブルだけが表示されます。選択されたコマンドのいずれか 1 つ以上で使用できるテーブル タイプが、リストに表示されます。コマンドが選択されない場合、すべてのテーブル タイプが表示されます。詳細は、535 ページのコマンド識別子とテーブルのサブタイプを参照してください。

▶ サブタイプではなくテキスト ラベルに基づいてテーブルを選択するには、**[ラベル]** をクリックします。詳細は、536 ページのラベルを参照してください。

▶ **[オプション]** をクリックして、出力形式 (SPSS データ ファイル、XML、HTML など) を指定します。デフォルトでは、XML 出力形式が使用されます。詳細は、538 ページのOMS オプションを参照してください。

▶ 出力先を指定します。

- **ファイル。** 選択したすべての出力が、単一のファイルに保存されます。

- **オブジェクト名に依存。**出力は、オブジェクト名に応じて複数の出力先ファイルに保存されます。各出力オブジェクトに対して別個のファイルが作成され、テーブルのサブタイプ名またはテーブルのラベルのいずれかに基づいたファイル名が付けられます。保存先フォルダ名を入力します。
- **新規データセット。**SPSS データ ファイル形式出力の場合、出力をデータセットに記録することができます。データセットは、現在のセッションの今後で利用可能ですが、セッション終了前に明示的に保存しない限り、保存されません。このオプションは、SPSS データ ファイル形式出力でのみ利用可能です。データセット名は、変数命名規則に従う必要があります。詳細は、85 ページの 5 章の変数名を参照してください。

▶ 次のオプションが選択できます。

- 選択された出力をビューアから除外します。**[ビューアから除外]** を選択した場合、OMS 要求の出力の種類は [ビューア] ウィンドウに表示されません。複数のアクティブな OMS 要求に同じ出力の種類が含まれている場合、これらの出力の種類をビューアで表示する方法は、その出力の種類を含む最新の OMS 要求によって決定されます。詳細は、543 ページのビューアからの出力表示の除外を参照してください。
- 要求に ID 文字列を割り当てます。すべての要求には、自動的に ID 値が割り当てられます。また、システムのデフォルトの ID 文字列の代わりに、わかりやすい ID を設定できます。これは、複数のアクティブな要求があり、簡単に識別する必要がある場合に役立ちます。ユーザーが割り当てる ID 値は、ドル記号 ($) で開始しないようにします。

リスト内の複数の項目を選択する場合の簡単なヒントを以下に示します。

- リスト内のすべての項目を選択するには、Ctrl キーを押しながら A キーを押します。
- 複数の連続した項目を選択するには、Shift キーを押しながらクリックします。
- 複数の連続していない項目を選択するには、Ctrl キーを押しながらクリックします。

OMS 要求を終了して削除するには

アクティブで新しい OMS 要求は [要求] リストに表示されます。最新の要求が先頭に表示されます。枠線をクリックしてドラッグすると、情報列の幅を変更できます。また、リストを横にドラッグすると、特定の要求に関する詳細情報を見ることができます。

[状況] 列の「**アクティブ**」という語の後にあるアスタリスク (*) は、コントロール パネルで使用できない機能を含むコマンド シンタックスによって作成された OMS 要求を示します。

特定の、アクティブな OMS 要求を終了するには、次の手順を実行します。

▶ [要求] リストで、対象とする要求の行の任意のセルをクリックします。

▶ [**終了**] をクリックします。

すべてのアクティブな OMS 要求を終了するには、次の手順を実行します。

▶ [**すべて終了**] をクリックします。

新しい要求 (追加済みで、まだアクティブではない要求) を削除するには、次の手順を実行します。

▶ [要求] リストで、対象とする要求の行の任意のセルをクリックします。

▶ [**削除**] をクリックします。

注: アクティブな OMS 要求は、[OK] をクリックするまで終了されません。

出力オブジェクトの種類

7 種類の異なる出力オブジェクトがあります。

図表。出力先の形式が XML および HTML である場合に限り、図表オブジェクトが含まれます。HTML 形式の場合、イメージ ファイルは、別個のサブディレクトリ (フォルダ) に保存されます。

ログ。ログ テキスト オブジェクト。ログ オブジェクトには、特定の種類のエラー メッセージと警告メッセージが含まれています。オプションの設定 ([編集] メニュー、[オプション]、[ビューア] タブ) によっては、セッション中に実行されるコマンド シンタックスがログ オブジェクトに含まれる場合もあります。ビューアのアウトライン ウィンドウで、ログ オブジェクトは [**ログ**] と表示されます。

テーブル。ビューアではピボット テーブルになる出力オブジェクト (記録テーブル含む)。テーブルは、SPSS データ ファイル (.sav) 形式に送れる唯一の出力オブジェクトです。

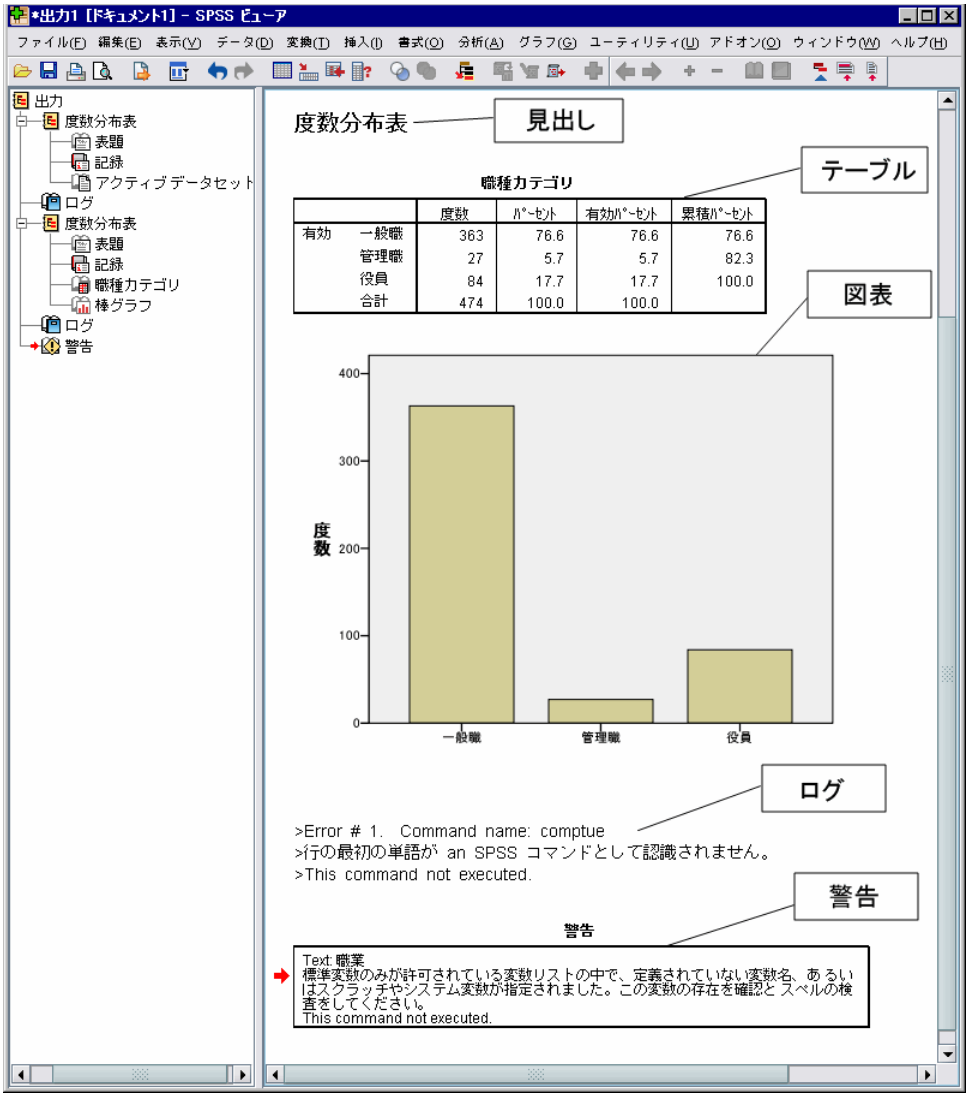
テキスト。ログまたは見出しではないテキスト オブジェクト (ビューアのアウトライン ウィンドウ枠にある [**テキスト出力**] というラベルのオブジェクト含む)。

見出し。ビューアのアウトライン ウィンドウで [**表題**] と表示されるオブジェクト。XML 出力形式には、見出しテキスト オブジェクトは含まれません。

警告。警告オブジェクトには、特定の種類のエラー メッセージと警告メッセージが含まれています。

ツリー。分類ツリー オプションによって作成されたツリー モデル図。出力先の形式が XML および HTML である場合に限り、ツリー オブジェクトが含まれます。

図 48-2
出力オブジェクトの種類



コマンド識別子とテーブルのサブタイプ

コマンド識別子

コマンド識別子は、すべての統計手続きと図表作成手続きに使用できます。また、ビューアのアウトライン ペイン内に独自の識別可能な見出しを持つ出力ブロックを作成する、その他のコマンドにも使用できます。これらの識別子は、メニューやダイアログ ボックスの表題の手続きと同じか似た名前であるのが一般的です (

ただし、常にそうであるとは限りません)。メニューやダイアログ ボックスの表題の手続き名は、基本となるコマンド名に似ていることが一般的です (ただし、常にそうであるとは限りません)。たとえば、度数分布表手続きのコマンド識別子は「Frequencies」で、基本となるコマンド名も同じです。

ただし、一部では、手続き名と、コマンド識別子やコマンド名がそれほど似ていない場合もあります。たとえば、[分析] メニューの [ノンパラメトリック検定] のすべての手続きで、同じ基本コマンドが使用されているので、コマンド識別子は基本コマンド名と同じ Npar Tests です。

テーブルのサブタイプ

テーブルのサブタイプは、生成可能なピボット テーブルのさまざまなタイプです。1 つのコマンドだけで作成されるサブタイプと、複数のコマンドで作成されるサブタイプがあります (ただし、テーブルは似ていない場合があります)。テーブルのサブタイプ名は、説明的なものが通常使用されます。ただし、多数のコマンドを選択した場合などは使用可能な名前が多くなり、また 2 つのサブタイプが非常に似た名前になってしまうということも考えられます。

コマンド識別子とテーブルのサブタイプの検索

疑わしい場合は、[ビューア] ウィンドウでコマンド識別子とテーブルのサブタイプ名を検索できます。

- 手続きを実行して、ビューアに出力を生成します。
- ビューアのアウトライン ペインの項目を右クリックします。
- [OMS コマンド識別子のコピー] または [OMS テーブルのサブタイプのコピー] を選択します。
- コピーされたコマンド識別子またはテーブルのサブタイプ名を、任意のテキスト エディタ (シンタックス エディタ ウィンドウなど) に貼り付けます。

ラベル

テーブルのサブタイプ名の代わりに、ビューアのアウトライン ウィンドウ枠に表示される、テキストを基に作成されるテーブルを選択できます。その他のオブジェクト型をラベルに基き選択することもできます。ラベルは、アウトライン テキストに特定の出力オブジェクトの一部の属性 (変数名やラベルなど) が反映される、同じタイプの複数のテーブルの識別に役立ちます。ただし、ラベル テキストに影響を与える要因がいくつかあります。

- ファイルの分割処理が有効になっている場合は、分割ファイル グループの識別がラベルに追加されます。

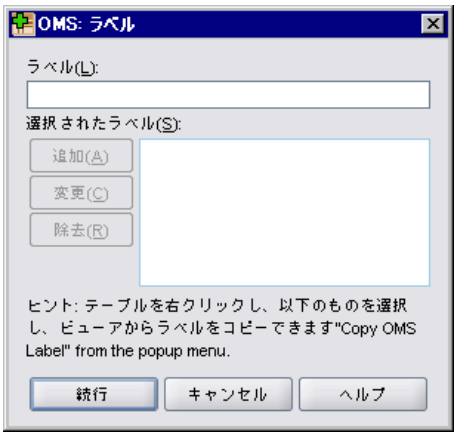
- 変数または値に関する情報を含むラベルは、現在の出力ラベル オプションの設定 ([編集] メニュー -> [オプション] -> [出力ラベル] タブ) の影響を受けます。
- ラベルは、現在の出力言語設定 ([編集] メニュー -> [オプション] -> [全般] タブ) の影響を受けます。

出力テーブルの識別に使用するラベルを指定するには

▶ 出力管理システム コントロール パネルで、1 つまたは複数の出力の種類を選択し、1 つまたは複数のコマンドを選択します。

▶ **[ラベル]** をクリックします。

図 48-3
[OMS テーブルのラベル] ダイアログ ボックス



▶ [ビューア] ウィンドウのアウトライン ウィンドウ枠に表示されている**とおりに**、ラベルを入力します。(アウトラインの項目を右クリックして、**[OMS ラベルのコピー]** を選択し、コピーしたラベルを [ラベル] テキスト フィールドに貼り付けます)。

▶ **[追加]** をクリックします。

▶ テーブルのラベルを含めるごとに、この処理を繰り返します。

▶ **[続行]** をクリックします。

ワイルドカード

ラベルの文字列の最後の文字に、ワイルドカード文字としてアスタリスク (*) を使用できます。指定された文字列 (アスタリスク以外) で始まるすべてのラベルが選択されます。アスタリスクはテーブルのラベルの内部で有効な文字として表示できるので、この処理は、アスタリスクが最後の文字である場合に限り有効です。

OMS オプション

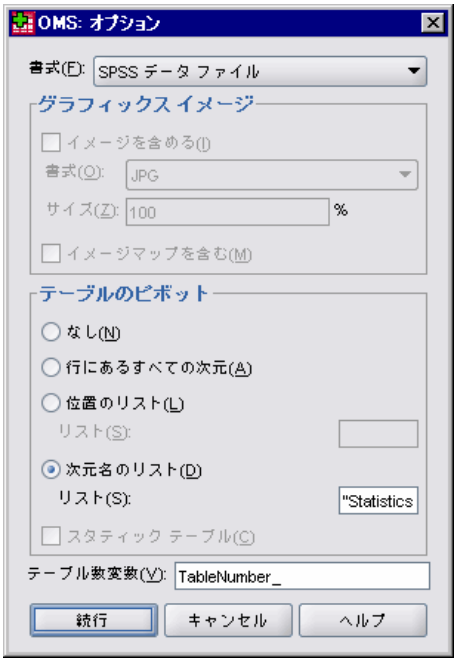
[OMS: オプション] ダイアログを使用して、以下の操作を実行できます。

- 出力形式を指定します。
- イメージ形式 (HTMLおよび出力 XML 出力形式) を選択します。
- 行次元に含めるテーブル次元要素を指定します。
- (SPSS データ ファイル形式では) 各ケースのソースである連続テーブル番号を識別する変数を含めます。

OMS オプションを指定するには

▶ 出力管理システム コントロール パネルで、[**オプション**] をクリックします。

図 48-4
[OMS:オプション] ダイアログ ボックス



書式 (カスタム テーブル)

XML の出力。SPSS出力 スキーマに従った XML。図表は、vizml スキーマに従った XML として含まれます (**xml.spss.com/spss/visualization**)。ツリー図は、pmml スキーマに従った XML として含まれます (*www.dmg.org*)。XML 形式のツリー図および図表を含む代わりに、選択したグラフィック形式に別のファイルとして保存できます。

HTML。ビューアでピボット テーブルになる出力オブジェクトは、簡単な HTML テーブルに変換されます。テーブルルック属性 (フォント特性、枠線のスタイル、色など) はサポートされていません。HTML では、テキスト出力オブジェクトには <PRE> というタグが付きます。図表を含めることを選択した場合、図表は選択されたグラフィック形式で、別個のファイルとしてエクスポートされ、参照により埋め込まれます。

SPSS データ ファイル。このフォーマットはバイナリ ファイル形式です。テーブル以外のすべての出力オブジェクトの種類は除外されます。テーブルの各列はデータ ファイルの変数になります。同じセッションの OMS で作成されたデータ ファイルを使用するには、データ ファイルを開く前に、アクティブな OMS 要求を終了する必要があります。詳細は、543 ページのSPSS データ ファイルへの出力の送信を参照してください。

SPV。SPSS ビューア ファイル形式。これは [ビューア] ウィンドウの内容を保存するときに使用した同じ形式です。

テキスト。スペース区切りのテキスト。出力はテキストとして書き込まれ、固定ピッチ フォントのスペースに合わせて表の出力が配置されます。図表はすべて除外されます。

タブ付きテキスト。タブ区切りテキスト。ビューアでピボット テーブルになる出力では、テーブル列要素がタブで区切られます。テキスト ブロック行はそのまま書き込まれます。タブを使用して有効な位置で区切られることはありません。図表はすべて除外されます。

グラフィックス イメージ

HTML および出力 XML 形式では、図表とツリー モデル図をイメージ ファイルとして含めることができます。各図表および/またはツリーごとに個別のイメージが作成されます。

- HTML 文書形式では、標準 <IMG SRC='filename'> タグが各イメージ ファイルに対して HTML 文書に含まれます。
- 出力 XML 文書形式では、XML ファイルは各イメージ ファイルに対して一般的な <hart imageFile="filepath/filename"/> の **イメージ ファイル** 属性を持つ **図表** 要素を含みます。
- イメージ ファイルは、別個のサブディレクトリ (フォルダ) に保存されます。サブディレクトリ名は、拡張子を含まない保存先ファイルの名前で、末尾に _files が追加されます。たとえば、保存先ファイルが julydata.htm である場合、イメージのサブディレクトリの名前は julydata_files になります。

書式。利用可能なイメージ形式は、PNG、JPG、EMF、BMP および VML です。

- EMF (拡張メタファイル)形式は、Windows オペレーティング システムでのみ使用できます。
- VML イメージ形式は、HTML 文書形式でのみ利用できます。

- VML イメージ形式は別のイメージ ファイルを作成しません。イメージを表す VML コードは HTML に埋め込まれます。
- VML イメージ形式はツリー図を含みません。

サイズ。イメージのサイズは 10% から 200% までの間で変更できます。

イメージマップを含む。HTML 文書形式では、このオプションが折れ線グラフの選択したポイントまたは棒グラフの棒の値のような図表要素の情報を表示するイメージ マップ ToolTips を作成します。

テーブルのピボット

ピボット テーブルの出力では、列に表示する次元要素を指定できます。その他のすべての次元要素は、行に表示されます。SPSS データ ファイル形式では、テーブル列が変数になり、行がケースになります。

- 複数の次元要素を指定した場合、列内では一覧された順序で入れ子になります。SPSS データ ファイル形式では、入れ子になった列要素によって変数名が作成されます。詳細は、550 ページのOMS で生成されたデータ ファイルの変数名を参照してください。
- 一覧された次元要素がテーブルに含まれていない場合は、そのテーブルのすべての次元要素が行に表示されます。
- ここで指定されたテーブルのピボットは、ビューアに表示されたテーブルには影響しません。

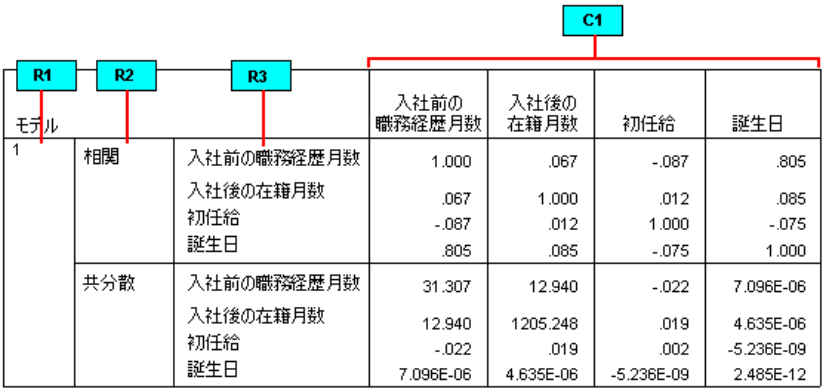
テーブルの行、列、層の各次元には、要素が含まれていないことも、複数の要素が含まれていることもあります。たとえば、単純な 2 次元クロス表に、1 つの行次元要素と 1 つの列次元要素が含まれ、それぞれにテーブルで使用する変数が 1 つずつ含まれているとします。位置の引数または次元要素のいずれかの「名前」を使用して、列次元に配置する次元要素を指定できます。

すべての次元を行に。各テーブルに単一の行を作成します。SPSS フォーマットのデータ ファイルでは、このことは、各テーブルが単一ケースであり、全テーブル要素が変数であることを表します。

位置のリスト。位置の引数の一般的書式は、要素のデフォルトの位置を示す文字です。列は C、行は R、層は L で、この後に次元内でのデフォルトの位置を示す正の整数が続きます。たとえば、R1 は最も外側の行次元要素を示します。

- 複数の次元から複数の要素を指定するには、「R1 C2」など、各次元をスペースで区切ります。
- 後に ALL が付いた次元文字は、その次元のすべての要素をデフォルトの順序で示します。たとえば、CALL はデフォルトの動作と同様で、すべての列要素をデフォルトの順序で使用して列を作成します。
- CALL RALL LALL (または RALL CALL LALL など) では、すべての次元要素が列に配置されます。SPSS データ ファイル形式では、この操作により、データ ファイルのテーブルごとに 1 つの行またはケースが作成されます。

図 48-5
行と列の位置の引数



次元名のリスト。位置の引数の代わりに、テーブルに表示されるテキスト ラベルである次元名の「名前」を使用できます。たとえば、単純な 2 次元クロス表には、1 つの行次元要素と 1 つの列次元要素が含まれ、それぞれに、これらの次元の変数に基づいたラベルが付きます。さらに、(出力言語が英語の場合)「**統計**」というラベルが付いた 1 つの層次元要素が含まれます。

- 次元要素名は、テーブルの変数名やラベルの表示に影響する、出力言語や設定によって異なります。
- 各次元要素名は、一重引用符または二重引用符で囲む必要があります。複数の次元要素名を指定するには、引用符で囲んだ各要素名の間に、スペースを含めます。

次元要素と関連付けられたラベルは、常に明確であるとは限りません。

ピボット テーブルのすべての次元要素とラベルを表示するには

▶ ビューアのテーブルをアクティブ化 (ダブルクリック) します。

▶ メニューから次の項目を選択します。

表示
すべて表示

および/または

▶ ピボット トレイが表示されていない場合は、メニューから次の項目を選択します。

ピボット
ピボット トレイ

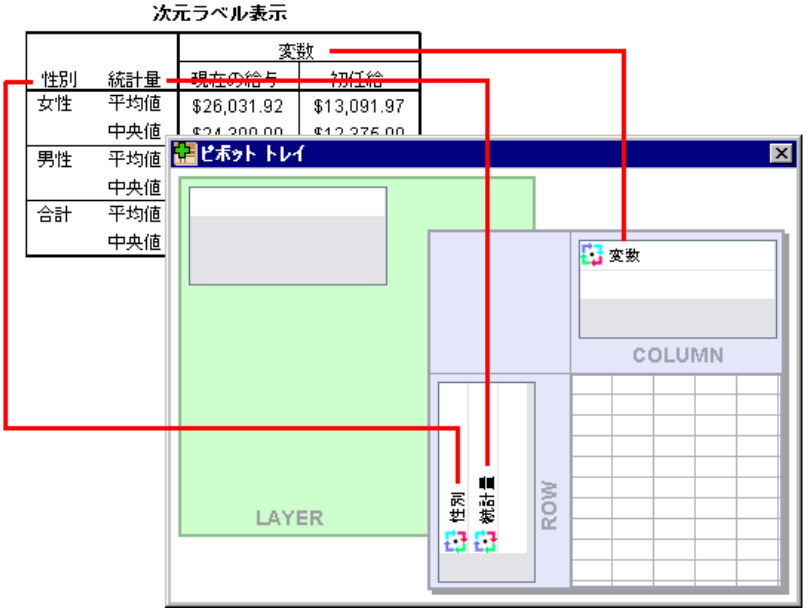
要素ラベルはピボット トレイに表示されます。

図 48-6
テーブルとピボット トレイに表示された次元要素名

次元ラベル非表示

		現在の給与	初任給
女性	平均値	$26,031.92	$13,091.97
	中央値	$24,300.00	$12,375.00
男性	平均値	$41,441.78	$20,301.40
	中央値	$32,850.00	$15,750.00
合計	平均値	$34,419.57	$17,016.09
	中央値	$28,875.00	$15,000.00

次元ラベル表示



ログ

XML またはテキスト形式のログに、OMS の動作状況を記録できます。

- ログにより、セッションのすべての新しい OMS 要求が追跡されます。ログを要求する前にすでにアクティブだった OMS 要求は含まれません。
- 新しいログ ファイルを指定した場合、または [**OMS 操作をログに記録**] をオフにした場合は、現在のログ ファイルは終了されます。

OMS ログを指定するには、次の手順を実行します。

▶ 出力管理システム コントロール パネルで、[**ログ**] をクリックします。

ビューアからの出力表示の除外

[ビューアから除外] チェック ボックスをオンにすると、OMS 要求で選択されているすべての出力が [ビューア] ウィンドウから表示されなくなります。多数の出力を生成するバッチ ジョブの場合や、ビューア ドキュメント (..spv ファイル) 形式の結果が必要ない場合に、この処理が役立つことが多くあります。また、再び参照する必要がない、特定の出力オブジェクトの表示を抑制するためにもこの機能を使用できます。その他の出力が外部のファイルと形式に送られることはありません。

その他の出力を外部ファイルに送ることなく、特定の出力オブジェクトの表示を抑制するには、次の手順を実行します。

- 不要な出力を識別する OMS 要求を作成します。
- [ビューアから除外] をオンにします。
- 出力先を選択するには、[ファイル] を選択してください。ただし、[ファイル] フィールドは空欄にしておいてください。
- [追加] をクリックします。

  選択された出力は、ビューアから除外されます。その他のすべての出力は、通常どおりビューアに表示されます。

SPSS データ ファイルへの出力の送信

SPSS データ ファイルは、列の変数と行のケースによって構成されています。これはピボット テーブルがデータ ファイルに変換される場合の基本的な仕様です。

- テーブルの列は、データ ファイルの変数です。有効な変数名は、列ラベルから作成されます。
- データ ファイルのテーブルの行ラベルは、汎用的な変数名 (Var1、Var2、Var3 など) を持つ変数になります。これらの変数の値は、テーブルの行ラベルです。
- データ ファイルには、Command_、Subtype_、Label_ という 3 つのテーブル識別子変数が自動的に含まれます。いずれも文字列変数です。最初の 2 変数はコマンド識別子およびサブタイプ識別子に関連しています。詳細は、535 ページのコマンド識別子とテーブルのサブタイプを参照してください。Label_ には、テーブル表題テキストが含まれます。
- テーブルの行は、データ ファイルのケースになります。

例: 単一の 2 次元テーブル

最も単純なケースである、単一の 2 次元テーブルでは、データ ファイルでテーブルの列が変数に、行がケースになります。

図 48-7
単一の 2 次元テーブル

報告書

性別		現在の給与	初任給
女性	平均値	$26,031.92	$13,091.97
	中央値	$24,300.00	$12,375.00
	最小値	$15,750	$9,000
	最大値	$58,125	$30,000
男性	平均値	$41,441.78	$20,301.40
合計			

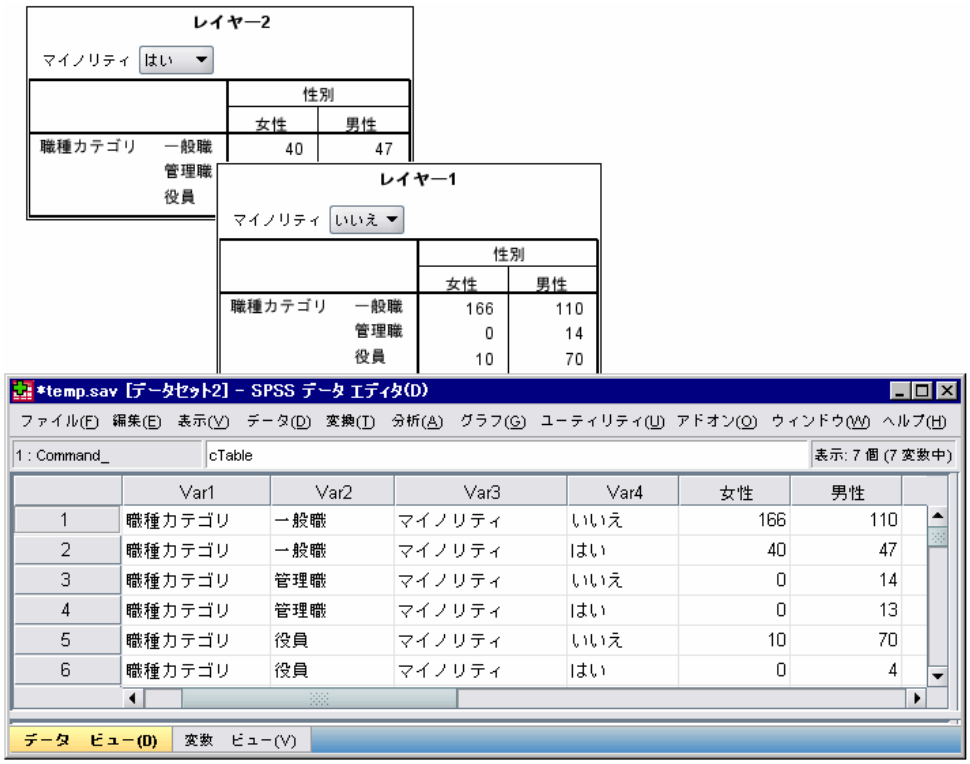
	command_	Subtytle_	Label_	Var1	Var2	現在の給与	初任給
1	Means	報告書	報告書	女性	平均値	$26,031.92	$13,091.97
2	Means	報告書	報告書	女性	中央値	$24,300.00	$12,375.00
3	Means	報告書	報告書	女性	最小値	$15,750.00	$9,000.00
4	Means	報告書	報告書	女性	最大値	$58,125.00	$30,000.00
5	Means	報告書	報告書	男性	平均値	$41,441.78	$20,301.40
6	Means	報告書	報告書	男性	中央値	$32,850.00	$15,750.00
7	Means	報告書	報告書	男性	最小値	$19,650.00	$9,000.00
8	Means	報告書	報告書	男性	最大値	$135,000.00	$79,980.00
9	Means	報告書	報告書	合計	平均値	$34,419.57	$17,016.09

- 最初の 3 つの変数では、コマンド、サブタイプ、およびラベルごとにソース テーブルを識別します。
- テーブルの行を定義した 2 つの要素である、変数「**性別**」の値と統計測定量には、一般的な変数名である **Var1** と **Var2** が割り当てられます。これらは両方とも文字型変数です。
- テーブルの列ラベルは、有効な変数名の作成に使用されます。ここでは、これらの変数名は、テーブルでまとめられた 3 つのスケール変数の変数ラベルに基づいています。変数によって変数ラベルが定義されていない場合、またはテーブルの列ラベルとして、変数ラベルではなく変数名を表示するように選択する場合は、新しいデータ ファイルでの変数名はソース データ ファイルでの変数名と同じになります。

例: 層のあるテーブル

テーブルには、行と列に加えて、3 次元である層次元を含めることができます。

図 48-8
層のあるテーブル表



- テーブルでは、[人種] というラベルが付いた変数により、層が定義されます。これにより、データ ファイルに 2 つの変数が作成されます。層要素を識別する変数と、層要素のカテゴリを識別する変数です。
- 行要素から作成された変数と同様に、層要素から作成された変数は、汎用的な変数名 (接頭辞 Var に連続番号が続く形式) が付いた文字型変数です。

複数のテーブルから作成されたデータ ファイル

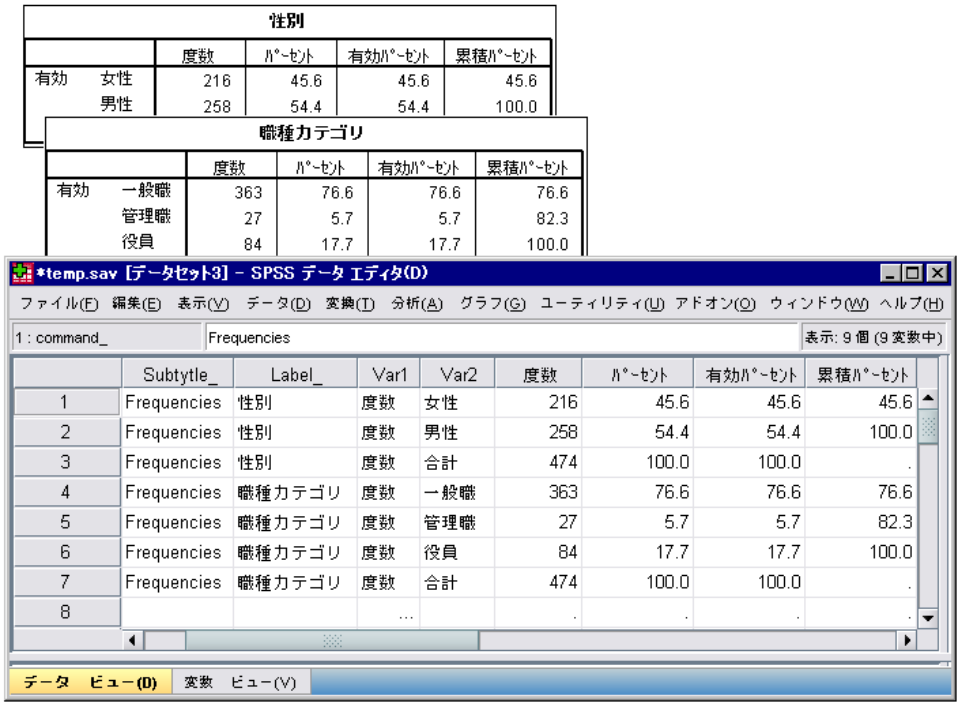
複数のテーブルが同じデータファイルに送られる場合は、各テーブルがデータ ファイルに追加されます ([データ] メニュー、[ファイルの結合]、[ケースの追加])。この方法は、あるデータ ファイルのケースを別のデータ ファイルに追加してデータ ファイルを結合する方法と同様です。

- 後続のテーブルが追加されるごとに、データ ファイルにケースが追加されます。
- また、テーブルの列ラベルが異なる場合、各テーブルによりデータ ファイルに変数が追加される場合があります。このとき、同じラベルが付いた列を持たない、他のテーブルのケースの欠損値も追加されます。

例: 同じ列ラベルが設定された複数のテーブル

同じ列ラベルを含む複数のテーブルでは、通常、最も即座に役立つデータファイル (追加の操作が不要なファイル) が作成されます。たとえば、度数分布表手続きからの度数分布表が 2 つ以上ある場合、すべての度数分布表に同一の列ラベルが存在します。

図 48-9
同一の列ラベルを含む 2 つのテーブル



性別

		度数	ﾊﾟｰｾﾝﾄ	有効ﾊﾟｰｾﾝﾄ	累積ﾊﾟｰｾﾝﾄ
有効	女性	216	45.6	45.6	45.6
	男性	258	54.4	54.4	100.0

職種カテゴリ

		度数	ﾊﾟｰｾﾝﾄ	有効ﾊﾟｰｾﾝﾄ	累積ﾊﾟｰｾﾝﾄ
有効	一般職	363	76.6	76.6	76.6
	管理職	27	5.7	5.7	82.3
	役員	84	17.7	17.7	100.0

*temp.sav [データセット3] - SPSS データ エディタ(D)

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) グラフ(G) ユーティリティ(U) アドオン(O) ウィンドウ(W) ヘルプ(H)

1 : command_ | Frequencies | 表示: 9 個 (9 変数中)

	Subtytle_	Label_	Var1	Var2	度数	ﾊﾟｰｾﾝﾄ	有効ﾊﾟｰｾﾝﾄ	累積ﾊﾟｰｾﾝﾄ
1	Frequencies	性別	度数	女性	216	45.6	45.6	45.6
2	Frequencies	性別	度数	男性	258	54.4	54.4	100.0
3	Frequencies	性別	度数	合計	474	100.0	100.0	.
4	Frequencies	職種カテゴリ	度数	一般職	363	76.6	76.6	76.6
5	Frequencies	職種カテゴリ	度数	管理職	27	5.7	5.7	82.3
6	Frequencies	職種カテゴリ	度数	役員	84	17.7	17.7	100.0
7	Frequencies	職種カテゴリ	度数	合計	474	100.0	100.0	.
8			...		.	.	.	.

データ ビュー(D) 変数 ビュー(V)

- 2 番目のテーブルでは、追加のケース (行) がデータ ファイルに追加されますが、列ラベルが完全に同じなので、新しい変数は追加されません。したがって、大規模な欠損値のパッチはありません。
- Command_ と Subtype_ の値は同じですが、2 つの度数分布表には異なる表題が付いているので、Label_ 値により、ケースの各グループのソース テーブルが識別されます。

例: 異なる列ラベルが設定された複数のテーブル

データ ファイルに送られたテーブルの一意な列ラベルに対して、データ ファイル内で新しい変数が作成されます。この処理の結果、テーブルに異なる列ラベルが含まれている場合、欠損値のブロックが発生します。

図 48-10
異なる列ラベルを含む 2 つのテーブル

テーブル1

性別		現在の給与	初任給
女性	平均値	$26,031.92	$13,091.97
	中央値	$24,300.00	$12,375.00
男性	平均値	$41,441.78	$20,301.40
合計			

テーブル2

性別		就学年数	入社後の在籍月数
女性	平均値	12.37	80.38
	中央値	12.00	81.00
男性	平均値	14.43	81.72

	Subtytle_	Label_	Var1	Var2	現在の給与	初任給	就学年数	入社後の在籍月数
1	Report	Report	女性	平均値	$26,032	$13,092.0	.	.
2	Report	Report	女性	中央値	$24,300	$12,375.0	.	.
3	Report	Report	男性	平均値	$41,442	$20,301.4	.	.
4	Report	Report	男性	中央値	$32,850	$15,750.0	.	.
5	Report	Report	合計	平均値	$34,420	$17,016.1	.	.
6	Report	Report	合計	中央値	$28,875	$15,000.0	.	.
7	Report	Report	女性	平均値	.	.	12.37	80.38
8	Report	Report	女性	中央値	.	.	12.00	81.00
9	Report	Report	男性	平均値	.	.	14.43	81.72
10	Report	Report	男性	中央値	.	.	15.00	82.00
11	Report	Report	合計	平均値	.	.	13.49	81.11
12	Report	Report	合計	中央値	.	.	12.00	81.00

- 1 番目のテーブルには [**初任給**] と [**給与**] という列が存在しますが、2 番目のテーブルには存在しません。したがって、2 番目のテーブルからのケースでは、これらの変数の欠損値が発生します。
- 逆に、2 番目のテーブルには [**就学年数**] と [**在籍月数**] という列が存在しますが、1 番目のテーブルには存在しません。したがって、1 番目のテーブルからのケースでは、これらの変数の欠損値が発生します。
- この例のような変数の不一致は、同じサブタイプのテーブルでも発生する場合があります。実際に、この例でも、両方のテーブルが同じサブタイプです。

例: 複数のテーブルから作成されなかったデータ ファイル

他のテーブルと同じ数の行要素が含まれていないテーブルでは、データ ファイルは作成されません。行の数は同じである必要はありませんが、データ ファイルで変数になる行**要素**の数は同じである必要があります。たとえば、2 つの変数を含

むクロス表に含まれる行要素の数と 3 つの変数を含むクロス表に含まれる行要素の数は、実際には、デフォルトの 3 つの変数を含むクロス表の表示では行変数内で「層」変数が入れ子になっているため、異なります。

図 48-11
行要素の数が異なるテーブル

職種カテゴリ と 性別 のクロス表

度数

		性別		
		女性	男性	合計
職種カテゴリ	事務	206	157	363
	中間管理	0	27	27
	管理	10	74	84
合計		216	258	474

職種カテゴリ と 性別 と 人種 のクロス表

度数

人種			性別		
			女性	男性	合計
白人	職種カテゴリ	事務	166	110	276
		中間管理	0	14	14
		管理	10	70	80
	合計		176	194	370
白人以外	職種カテゴリ	事務	40	47	87
		中間管理	0	13	13
		管理	0	4	4
	合計		40	64	104

列要素を制御してデータ ファイルの変数を制御する

出力管理コントロール パネルの [オプション] ダイアログ ボックスで、列に含める次元要素を指定できます。これらの次元要素は、生成されたデータ ファイルでの変数の作成に使用されます。この処理は、ビューアでのテーブルのピボットに相当します。

たとえば、度数分布表手続きからは統計が行に格納された記述統計テーブルが作成され、記述統計手続きからは統計が列に格納された記述統計テーブルが作成されます。両方の種類のテーブルを有意な方法で同一のデータ ファイルに含めるには、一方のテーブルの種類の列の次元を変更する必要があります。

両方のテーブル タイプで、統計量次元に要素名「Statistics」を使用しているため、OMS [オプション] ダイアログ ボックスの次元名リストで「Statistics」(引用符で囲みます) を指定するだけで、度数分布の統計量テーブルからの統計量を、列に設定できます。

図 48-12
[OMS: オプション] ダイアログ ボックス

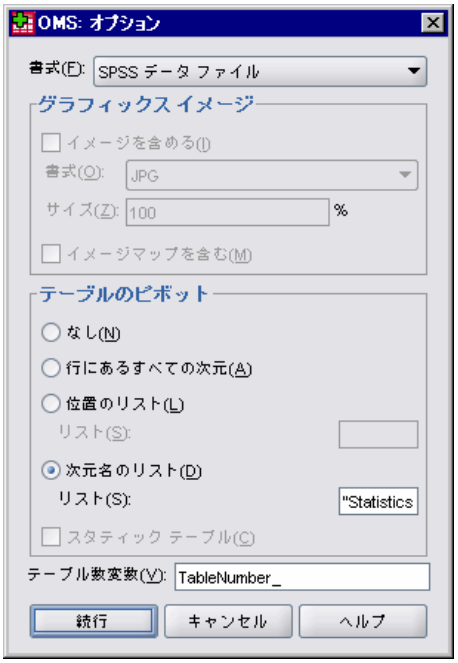
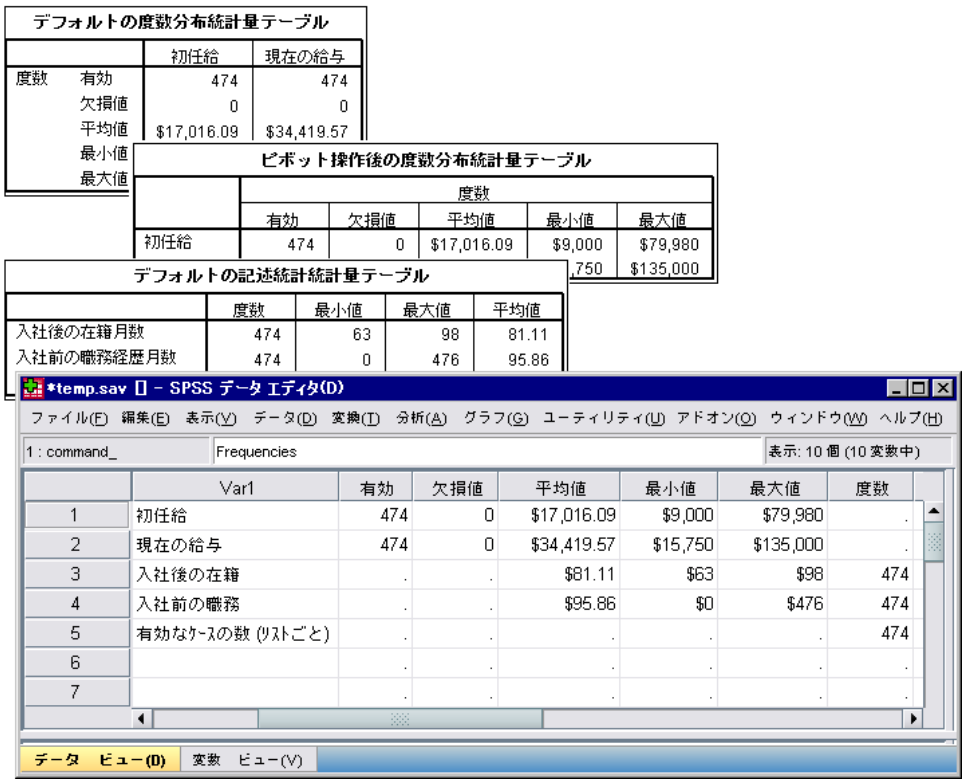


図 48-13
次元要素をピボットしてデータ ファイル内で異なるテーブルの種類を組み合わせる



デフォルトの度数分布統計量テーブル

		初任給	現在の給与
度数	有効	474	474
	欠損値	0	0
	平均値	$17,016.09	$34,419.57
	最小値		
	最大値		

ピボット操作後の度数分布統計量テーブル

	度数				
	有効	欠損値	平均値	最小値	最大値
初任給	474	0	$17,016.09	$9,000	$79,980
				,750	$135,000

デフォルトの記述統計統計量テーブル

	度数	最小値	最大値	平均値
入社後の在籍月数	474	63	98	81.11
入社前の職務経歴月数	474	0	476	95.86

*temp.sav [] – SPSS データ エディタ(D)

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) グラフ(G) ユーティリティ(U) アドオン(O) ウィンドウ(W) ヘルプ(H)

1 : command_ | Frequencies | 表示: 10 個 (10 変数中)

	Var1	有効	欠損値	平均値	最小値	最大値	度数
1	初任給	474	0	$17,016.09	$9,000	$79,980	.
2	現在の給与	474	0	$34,419.57	$15,750	$135,000	.
3	入社後の在籍	.	.	$81.11	$63	$98	474
4	入社前の職務	.	.	$95.86	$0	$476	474
5	有効なｹｰｽの数 (ﾘｽﾄごと)	.	.	.	.	.	474
6		.	.	.	.	.	.
7		.	.	.	.	.	.

データ ビュー(D) 変数 ビュー(V)

テーブルの構造が列の統計量とまだ完全に同じではないため、一部の変数に欠損値が発生します。

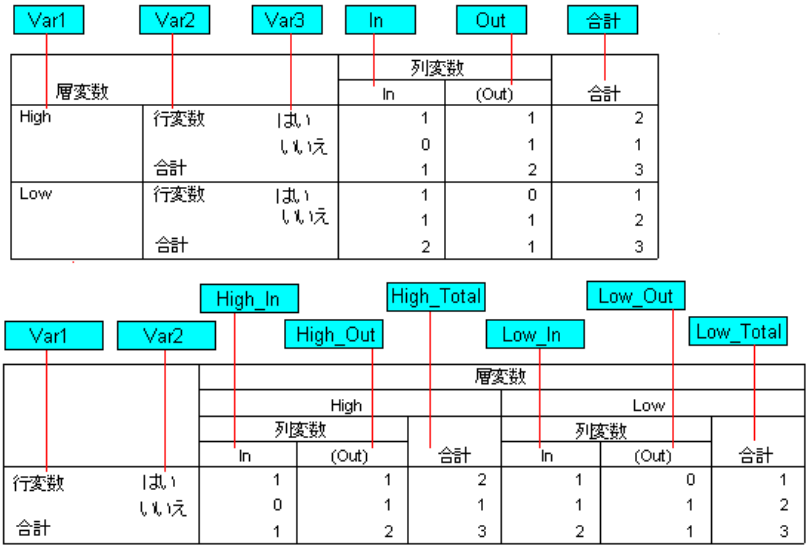
OMS で生成されたデータ ファイルの変数名

OMS により、列ラベルから有効で一意な変数名が作成されます。

- 行要素と層要素には、汎用的な変数名 (Var に連続番号が続く形式) が割り当てられます。
- 変数名に使用できない文字 (スペースやかっこなど) は削除されます。たとえば、「This (Column) Label。」は ThisColumnLabel という変数になります。
- 変数名に使用できても、最初の文字としては使用できない文字 (数字など) で始まるラベルには、接頭辞として「@」が挿入されます。たとえば、「2nd」は「@2nd」という変数になります。

- ラベルの末尾の下線やピリオドは、作成される変数名から削除されます。(自動的に生成された変数 Command_、Subtype_、および Label_ の下線は、削除されません)。
- 列次元に複数の要素が存在する場合は、カテゴリ ラベルを下線でつないで組み合わせることで変数名が作成されます。グループ ラベルは含まれません。たとえば、列の VarA の下で VarB が入れ子になっている場合は、CatA1_CatB1 という変数が生成されます。VarA_CatA1_VarB_CatB1 にはなりません。

図 48-14
テーブル要素から作成される変数名



OXML テーブルの構造

XML の出力 (OXML) は spss-output スキーマに従った XML です。スキーマの詳細な説明は、ヘルプ システムで出力スキーマに関するセクションを参照してください。

- OMS コマンド識別子とサブタイプ識別子は、OXML の command 属性と subType 属性の値として使用されます。一例を以下に挙げます。

```
<command text="Frequencies" command="Frequencies"...>
  <pivotTable text="Gender" label="Gender" subType="Frequencies"...>
```

- OMS の command 属性値と subType 属性値は、出力言語の影響、または変数名/ラベルや値/値ラベルの表示設定の影響を受けません。
- XML では、大文字と小文字が区別されます。「frequencies」の subType 属性値は、「Frequencies」の subType 属性値と同じでは**ありません**。

- テーブルに表示されるすべての情報は、OXML の属性値に含まれています。個別のセル レベルでは、OXML は「空」の要素で構成されています。空の要素には属性が含まれますが、属性値に含まれるもの以外の「内容」は含まれません。
- OXML のテーブルの構成は、行単位で表されます。列を表す要素は行の内部で入れ子になっていて、個別のセルは列要素の内部で入れ子になっています。

```
<pivotTable...>
 <dimension axis='row'...>
   <dimension axis='column'...>
     <category...>
       <cell text='...' number='...' decimals='...'/>
     </category>
     <category...>
       <cell text='...' number='...' decimals='...'/>
     </category>
   </dimension>
 </dimension>
    ...
 </pivotTable>
```

上の例は、これらの要素の子孫/祖先の関係を示す構造を簡略化した表現です。しかし、通常は入れ子になった要素レベルが介在するので、親/子の関係を示すとは限りません。

以下の図は、簡単な度数分布表と、その表の完全な XML 出力表現を示します。

図 48-15
簡単な度数分布表

性別

		度数	パーセント	有効パーセント	累積パーセント
有効	女性	3179	49.7	49.7	49.7
	男性	3221	50.3	50.3	100.0
	合計	6400	100.0	100.0	

図 48-16
簡単な度数分布表の XML 出力

```
<?xml version="1.0" encoding="UTF-8" ?>
<outputTreeoutputTree xmlns="http://xml.spss.com/spss/oms"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://xml.spss.com/spss/oms
  http://xml.spss.com/spss/oms/spss-output-1.0.xsd">
 <command text="Frequencies" command="Frequencies"
  displayTableValues="label" displayOutlineValues="label"
```

```
displayTableVariables="label" displayOutlineVariables="label">
<pivotTable text="Gender" label="Gender" subType="Frequencies"
 varName="gender" variable="true">
 <dimension axis="row" text="Gender" label="Gender"
  varName="gender" variable="true">
  <group text="Valid">
   <group hide="true" text="Dummy">
    <category text="Female" label="Female" string="f"
     varName="gender">
     <dimension axis="column" text="Statistics">
      <category text="Frequency">
       <cell text="216" number="216"/>
      </category>
      <category text="Percent">
       <cell text="45.6" number="45.569620253165" decimals="1"/>
      </category>
      <category text="Valid Percent">
       <cell text="45.6" number="45.569620253165" decimals="1"/>
      </category>
      <category text="Cumulative Percent">
       <cell text="45.6" number="45.569620253165" decimals="1"/>
      </category>
     </dimension>
    </category>
    <category text="Male" label="Male" string="m" varName="gender">
     <dimension axis="column" text="Statistics">
      <category text="Frequency">
       <cell text="258" number="258"/>
      </category>
      <category text="Percent">
       <cell text="54.4" number="54.430379746835" decimals="1"/>
      </category>
      <category text="Valid Percent">
       <cell text="54.4" number="54.430379746835" decimals="1"/>
      </category>
      <category text="Cumulative Percent">
       <cell text="100.0" number="100" decimals="1"/>
      </category>
     </dimension>
    </category>
   </group>
   <category text="Total">
    <dimension axis="column" text="Statistics">
     <category text="Frequency">
      <cell text="474" number="474"/>
     </category>
     <category text="Percent">
      <cell text="100.0" number="100" decimals="1"/>
     </category>
```

```
        <category text="Valid Percent">
         <cell text="100.0" number="100" decimals="1"/>
        </category>
       </dimension>
      </category>
     </group>
    </dimension>
   </pivotTable>
  </command>
</outputTree>
```

お気付きのように、簡単で小規模な度数分布表でも、かなりの量の XML が生成されます。XML には、元の表では簡単にわからなかった情報、元の表では表示できなかった情報、およびある程度の冗長性が含まれるのが理由の一つです。

- ビューアのピボット テーブルで表現されるテーブルの内容は、テキスト属性に含まれています。一例を以下に挙げます。

```
<command text="Frequencies" command="Frequencies"...>
```

- テキスト属性は、出力言語の影響、および変数名/ラベルと値/値ラベルの表示設定の影響の両方を受ける場合があります。この例では、text 属性値は出力言語によって異なりますが、command 属性値は出力言語に関係なく同じになります。
- 行ラベルまたは列ラベルで変数や変数の値が使用される場面では、XML に text 属性および 1 つ以上の追加の属性値が含まれます。一例を以下に挙げます。

```
<dimension axis="row" text="Gender" label="Gender" varName="gender">
    ...<category text="Female" label="Female" string="f" varName="gender">
```

- 数値型変数には、string 属性ではなく number 属性が含まれます。変数または値によってラベルが定義されている場合に限り、label 属性が含まれます。
- 数値のセル値を含む <cell> 要素には、text 属性および 1 つ以上の追加の属性値が含まれます。一例を以下に挙げます。

```
<cell text="45.6" number="45.569620253165" decimals="1"/>
```

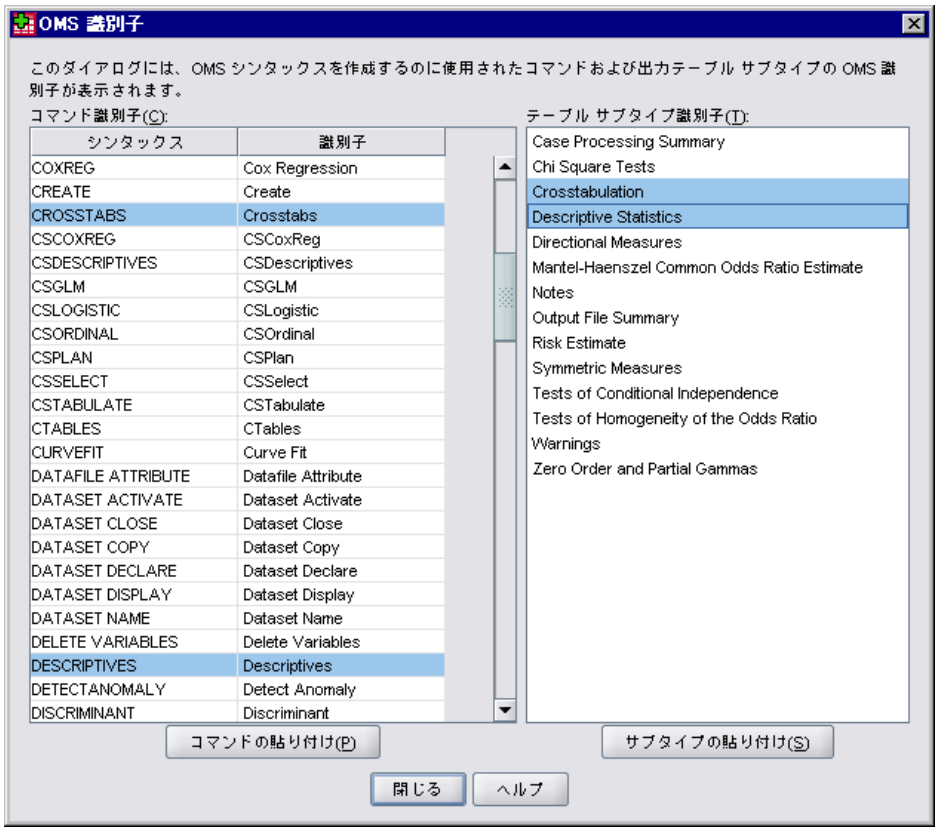
number 属性は丸められていない実際の数値で、decimals 属性はテーブルに表示される小数桁数を示します。

- 行の内部で列が入れ子になっているので、各列を識別するカテゴリ要素が、行ごとに繰り返されます。たとえば、統計量は列に表示されるので、要素 <category text="Frequency"> は XML で male 行、female 行、および total 行の 3 か所に表示されています。

OMS 識別子

[OMS 識別子] ダイアログ ボックスは、OMS コマンド シンタックスの記述を支援するために設計されています。このダイアログ ボックスを使用して、選択したコマンド識別子とサブタイプ識別子をコマンド シンタックス ウィンドウに貼り付けることができます。

図 48-17
[OMS 識別子] ダイアログ ボックス



[OMS 識別子] ダイアログ ボックスの使用

▶ メニューから次の項目を選択します。

ユーティリティ
OMS 識別子

▶ 1 つ以上のコマンド識別子またはサブタイプ識別子を選択します。(Ctrl キーを押しながらクリックすることにより、各リストから複数の識別子を選択できます。)

▶ **[コマンドの貼り付け]** や **[サブタイプの貼り付け]** をクリックします。

- 利用可能なサブタイプのリストは、現在選択されているコマンドに基づいています。複数のコマンドを選択している場合、利用可能なサブタイプのリストには、選択した各コマンドで利用可能なすべてのサブタイプが表示されます。コマンドを選択しなかった場合は、すべてのサブタイプが一覧表示されます。
- 識別子は、指定したコマンド シンタックス ウィンドウ内のカーソルの現在位置に貼り付けられます。コマンド シンタックス ウィンドウを開いていない場合は、新しいシンタックス ウィンドウが自動的に開かれます。
- OMS コマンド シンタックスの要件に従い、各コマンド識別子およびサブタイプ識別子は引用符で囲まれて貼り付けられます。
- COMMANDS および SUBTYPES キーワードの識別子リストは、次のようにかっこで囲む必要があります。

```
/IF COMMANDS=['Crosstabs' 'Descriptives']
    SUBTYPES=['Crosstabulation' 'Descriptive Statistics']
```

ビューアのアウトラインからの OMS 識別子のコピー

ビューアのアウトライン ウィンドウ枠でも、OMS コマンド識別子およびサブタイプ識別子のコピーや貼り付けを行えます。

▶ アウトライン ペインで項目のアウトライン エントリを右クリックします。

▶ **[OMS コマンド識別子のコピー]** または **[OMS テーブルのサブタイプのコピー]** を選択します。

[OMS 識別子] ダイアログ ボックスの方法と異なり、コピーされた識別子がコマンド シンタックス ウィンドウに自動的に貼り付けられることはありません。識別子は単にクリップボードにコピーされるだけなので、任意の場所に貼り付けることができます。コマンドおよびサブタイプの識別子の値は、XML 形式の出力 (OXML) の対応するコマンドおよびサブタイプの属性値と同じであり、このコピー/貼り付け方式は XSLT 変換を書き出す場合に有用です。

OMS ラベルのコピー

識別子ではなく、ラベルをコピーして LABELS キーワードで使用します。ラベルを使用すると、アウトライン テキストに特定の出力オブジェクトのいくつかの属性 (変数名やラベルなど) が反映される、同じタイプの複数のグラフまたは複数のテーブルを識別できます。ただし、ラベル テキストに影響を与える要因がいくつかあります。

- ファイルの分割処理が有効になっている場合は、分割ファイル グループの識別がラベルに追加されます。
- 変数または値に関する情報を格納しているラベルは、アウトライン ペインの変数名/ラベルおよび値/値ラベルの表示設定 ([編集] メニュー -> [オプション] -> [出力ラベル] タブ) の影響を受けます。
- ラベルは、現在の出力言語設定 ([編集] メニュー -> [オプション] -> [全般] タブ) の影響を受けます。

OMS ラベルをコピーするには:

▶ アウトライン ペインで項目のアウトライン エントリを右クリックします。

▶ [OMS ラベルのコピー] を選択します。

次に示すように、コマンド識別子およびサブタイプ識別子の場合と同様、ラベルを引用符で囲み、リスト全体を角かっこで囲む必要があります。

```
/IF LABELS=['Employment Category' 'Education Level']
```

索引

0 次相関
  偏相関, 353
1 サンプルによる Kolmogorov-Smirnov 検定, 429
  オプション, 431
  欠損値, 431
  検定分布, 429
  コマンドの追加機能, 431
  統計量, 431
1 サンプルの t 検定, 325
  オプション, 326
  欠損値, 326
  コマンドの追加機能, 327
  信頼区間, 326
2 項検定, 425
  2 分変数, 425
  オプション, 426
  欠損値, 426
  コマンドの追加機能, 427
  統計量, 426
2 個の対応サンプルの検定, 434
  オプション, 436
  欠損値, 436
  検定の種類, 435
  コマンドの追加機能, 436
  統計量, 436
2 個の独立サンプルの検定, 431
  オプション, 434
  グループ化変数, 433
  グループの定義, 433
  欠損値, 434
  検定の種類, 432
  コマンドの追加機能, 434
  統計量, 434
2 サンプルの *t* 検定
  独立サンプルの t 検定, 320
2 次モデル
  曲線推定, 380
2 変量の相関分析
  オプション, 350
  欠損値, 350
  コマンドの追加機能, 350
  相関係数, 348
  統計量, 350
  有意確率, 348
3 次モデル
  曲線推定, 380
4 分位
  度数分布表, 283
5% トリム平均
  探索的, 293

Access (Microsoft), 20
Anderson-Rubin 因子得点, 399
Andrews のウェイブ推定量
  探索的, 293
Bartlett 因子得点, 399
Bartlett の球面性の検定
  因子分析, 396
Blom 推定値, 158
BMP ファイル, 237, 244
  図表のエクスポート, 237, 244
Bonferroni の方法
  GLM, 341
  一元配置分散分析, 330
Box の M 検定
  判別分析, 389
Brown-Forsythe 統計量
  一元配置分散分析, 333
Chebychev の距離
  距離, 357
Cochran 統計量
  クロス集計表, 299
Cochran の *Q*
  複数の対応サンプルの検定, 440
Cohen のカッパ
  クロス集計表, 299
Cook の距離
  GLM, 343
  線型回帰, 365
Cox と Snell の R^2 乗
  順序回帰, 373
Cramer の *V*
  クロス集計表, 299
Cronbach のアルファ
  信頼性分析, 463, 465
CSV 形式
  データの保存, 48
  データの読み込み, 33
d
  クロス集計表, 299
DATA LIST, 70
  GET DATA コマンドとの比較, 70
dBASE ファイル, 16, 18, 48
  保存, 48
  読み込み, 16, 18

DfBeta
線型回帰, 365
DfFit
線型回帰, 365
Dimensions データ, 42
保存, 69
DOT 形式, 87, 89
Duncan の多重範囲検定
GLM, 341
一元配置分散分析, 330
Dunnett の *C*
GLM, 341
一元配置分散分析, 330
Dunnett の T3
GLM, 341
一元配置分散分析, 330
Dunnett の *t* 検定
GLM, 341
一元配置分散分析, 330
Durbin-Watson の統計量
線型回帰, 367
EPS ファイル, 237, 244
図表のエクスポート, 237, 244
Euclid 距離
距離, 357
Excel ファイル, 16–17, 48, 515
Excel
にデータを送信するためのメニュー項目を追加, 515
値の代わりに値ラベルを保存, 48
開く, 16–17
保存, 48
Excel 形式
出力をエクスポートする, 237, 239
EXECUTE (コマンド)
ダイアログ ボックスから貼り付け, 281
Fisher の LSD
GLM, 341
Fisher の直接法
クロス集計表, 299
Friedman の検定
複数の対応サンプルの検定, 440
Gabriel と Howell のペアごとの比較検定
GLM, 341
一元配置分散分析, 330
Gabriel のペアごとの比較検定
GLM, 341
一元配置分散分析, 330
GET DATA, 70
DATA LIST コマンドとの比較, 70
GET CAPTURE コマンドとの比較, 70
GLM
行列の保存, 343
その後の検定, 341
プロファイル プロット, 340
平方和, 337
変数の保存, 343
モデル, 337
GLM – 1 変量, 335, 346
オプション, 345
診断, 345
推定周辺平均, 345
対比, 339–340
表示, 345
Goodman と Kruskal のガンマ
クロス集計表, 299
Goodman と Kruskal のタウ
クロス集計表, 299
Goodman と Kruskal のラムダ
クロス集計表, 299
Guttman のモデル
信頼性分析, 463, 465
Hampel の M-推定量
探索的, 293
Helmert 対比
GLM, 339–340
Hochberg の GT2
GLM, 341
一元配置分散分析, 330
Hotelling の T^2
信頼性分析, 463, 465
HTML, 237, 239
出力をエクスポートする, 237, 239
Huber の M-推定量
探索的, 293
ICC。「級内相関係数」を**参照**, 465
JPEG ファイル, 237, 243
図表のエクスポート, 237, 243
Kendall の *W*
複数の対応サンプルの検定, 440
Kendall のタウ *b*
2 変量の相関分析, 348
クロス集計表, 299
Kendall のタウ *c*, 299
クロス集計表, 299
Kolmogorov-Smirnov の *Z*
1 サンプルによる Kolmogorov-Smirnov 検定, 429
2 個の独立サンプルの検定, 432
KR20
信頼性分析, 465
Kruskal-Wallis の *H*
2 個の独立サンプルの検定, 436
Kruskal のタウ
クロス集計表, 299
Kuder-Richardson 20 (KR20)
信頼性分析, 465
LAG (関数), 181

Lance と Williams の非類似度測定方法, 357
　距離, 357
Levene の検定
　GLM 1 変量, 345
　一元配置分散分析, 333
　探索的, 294
Lilliefors の検定
　探索的, 294
Lotus 1-2-3 ファイル, 16, 48, 515
　Lotus
　にデータを送信するためのメニュー項目を追加, 515
　開く, 16
　保存, 48
M-推定量
　探索的, 293
Mahalanobis の距離
　線型回帰, 365
　判別分析, 390
Mann-Whitney の *U*
　2 個の独立サンプルの検定, 432
Mantel - Haenszel 統計量
　クロス集計表, 299
McFadden の R^2 乗
　順序回帰, 373
McNemar の検定
　2 個の対応サンプルの検定, 434
　クロス集計表, 299
Microsoft Access, 20
Minkowski の距離
　距離, 357
Moses の外れ値反応検定
　2 個の独立サンプルの検定, 432
mrInterview, 42
Nagelkerke の R^2 乗
　順序回帰, 373
Newman-Keuls
　GLM, 341
[OK] ボタン, 9
OLAP キューブ, 314
　統計量, 316
　表題, 319
OMS, 531, 555
　OXML に対する XSLT の使用, 556
　SAV ファイル形式, 538, 543
　SAV ファイルの変数名, 550
　SPSS データ ファイル形式, 538, 543
　XML, 538, 551
　コマンド識別子, 535
　出力オブジェクトの種類, 534
　テーブルのサブタイプ, 535
　テーブルのピボットの制御, 538, 548
　テキスト形式, 538
　ビューアから出力を除外, 543
OMS によるビューアからの出力の除外, 543
OMS によるビューアからの出力の非表示 (除外), 543
OXML, 556
PDF
　出力をエクスポートする, 237, 240
Pearson 残差
　順序回帰, 373
Pearson のカイ 2 乗
　クロス集計表, 299
　順序回帰, 373
Pearson の相関係数
　2 変量の相関分析, 348
　クロス集計表, 299
PLUM
　順序回帰, 371
PNG ファイル, 237, 244
　図表のエクスポート, 237, 244
PostScript ファイル (EPS), 237, 244
　図表のエクスポート, 237, 244
PowerPoint, 240
　PowerPoint として出力をエクスポートする, 240
PowerPoint 形式
　出力をエクスポートする, 237
Quancept, 42
Quanvert, 42
R^2
　R^2 の変化量, 367
　グループの平均, 311
　線型回帰, 367
R-E-G-W の *F*
　GLM, 341
　一元配置分散分析, 330
R-E-G-W の *Q*
　GLM, 341
　一元配置分散分析, 330
Rankit 推定値, 158
Rao の *V*
　判別分析, 390
ROC 曲線, 478
　統計量と作図, 480
Ryan-Einot-Gabriel-Welsch 多重 *F*
　GLM, 341
　一元配置分散分析, 330
Ryan-Einot-Gabriel-Welsch 多重範囲
　GLM, 341
　一元配置分散分析, 330
R 統計量
　グループの平均, 311
　線型回帰, 367
r の相関係数
　2 変量の相関分析, 348
　クロス集計表, 299

SAS ファイル
開く, 16
保存, 48
SAV ファイル形式
出力を SPSS データ ファイルに送信, 538, 543
Scheffé の検定
GLM, 341
一元配置分散分析, 330
Shapiro-Wilk の検定
探索的, 294
Sidak の *t* 検定
GLM, 341
一元配置分散分析, 330
Somers の *d*
クロス集計表, 299
Spearman-Brown の信頼性
信頼性分析, 465
Spearman の相関係数
2 変量の相関分析, 348
クロス集計表, 299
spp ファイル
spj ファイル変換, 529
SPSSTMPDIR 環境変数, 511
SPSS データ ファイル形式
出力をデータファイルに送信, 538, 543
Stata ファイル, 18
開く, 18
保存, 48
読み込み, 16
Student-Newman-Keuls の検定
GLM, 341
一元配置分散分析, 330
SYSTAT ファイル, 16
開く, 16
S ストレス
多次元尺度法, 468
S モデル
曲線推定, 380
T4253H 平滑化, 181
Tamhane の T2
GLM, 341
一元配置分散分析, 330
TIFF ファイル, 244
図表のエクスポート, 237, 244
Tukey 推定値, 158
Tukey の *b* 検定
GLM, 341
一元配置分散分析, 330
Tukey の HSD 検定
GLM, 341
一元配置分散分析, 330
Tukey の加法性の検定
信頼性分析, 463, 465
Tukey のバイウェイト推定量
探索的, 293
TwoStep クラスタ分析, 403
オプション, 406
外部ファイルに保存, 409
作業ファイルに保存, 409
作図, 408
統計量, 409
t 検定
1 サンプルの t 検定, 325
GLM 1 変量, 345
対応のあるサンプルの t 検定, 323
独立サンプルの t 検定, 320
Unicode, 16, 47
Unicode コマンド シンタックス ファイル, 280
V
クロス集計表, 299
Van der Waerden 推定値, 158
Wald-Wolfowitz のラン検定
2 個の独立サンプルの検定, 432
Waller-Duncan の *t* 検定
GLM, 341
一元配置分散分析, 330
Welch 統計量
一元配置分散分析, 333
Wilcoxon の符号付き順位検定
2 個の対応サンプルの検定, 434
Wilks のラムダ
判別分析, 390
Word 形式
出力をエクスポートする, 237, 239
幅の広いテーブル, 237
XML
OMS からの OXML 出力, 556
OXML でのテーブルの構造, 551
XML として出力を保存, 531
出力を XML に送信, 538
XSLT
OXML に使用, 556
Yates の連続修正
クロス集計表, 299
Z 得点
記述統計, 287
ケースのランク付け, 158
変数として保存, 287

アイコン
ダイアログ ボックス, 9
アウトライン, 232
閉じる, 232
ビューア, 232
開く, 232
レベルの変更, 232

アクティブ ウィンドウ, 6
アクティブ ファイル, 70, 72
  仮想アクティブ ファイル, 70
  キャッシュ, 72
  テンポラリ アクティブ ファイルの作成, 72
値の再割り当て, 135, 152–153, 155–156, 160
値ラベル, 90, 100, 105, 114, 504
  Excel ファイルを保存, 48
  アウトライン ウィンドウ枠, 504
  改行の挿入, 91
  結合したデータ ファイル, 195
  コピー, 119
  データ エディタ, 105
  データ入力に使用, 100
  ピボット テーブル, 504
  複数の変数への適用, 119
アルゴリズム, 12
アルファ因子法, 397
アルファ係数
  信頼性分析, 463, 465

イータ
  グループの平均, 311
  クロス集計表, 299
イータの 2 乗
  GLM 1 変量, 345
  グループの平均, 311
位置合わせ, 92, 231, 500
  出力, 231, 500
  データ エディタ, 92
一元配置分散分析, 328
  因子変数, 328
  オプション, 333
  欠損値, 333
  コマンドの追加機能, 334
  その後の検定, 330
  対比, 329
  多項式の対比, 329
  多重比較, 330
  統計量, 333
位置モデル
  順序回帰, 375
一致するペアの研究
  対応のあるサンプルの t 検定, 323
一般化した最小 2 乗法
  因子分析, 397
移動中央値関数, 181
イメージ因子法, 397
入れ替える, 272
  ピボット テーブルでの選択, 272
印刷, 106, 245–246, 248, 261, 265, 273
  隠す, 106
  出力項目間のスペース, 248
  図表, 245
  図表サイズ, 248
  層, 245, 261, 265
  テーブルのサイズ変更, 261, 265
  テーブル分割の制御, 273
  テキスト出力, 245
  ピボット テーブル, 245
  プレビューの印刷, 245
  ページの番号付け, 248
  ヘッダーとフッター, 246
印刷タイトル, 265
  ピボット テーブル, 265
因子得点, 399
因子分析, 394
  因子抽出法, 397
  因子得点, 399
  回転法, 398
  概要, 394
  記述統計, 396
  係数の表示書式, 400
  ケースの選択, 395
  欠損値, 400
  コマンドの追加機能, 400
  収束, 397–398
  統計量, 394, 396
  プロットのロード, 398
  例, 394
インタラクティブ グラフ, 236
  他のアプリケーションにコピーする, 236

ウィンドウ, 4
  アクティブ ウィンドウ, 6
  指定ウィンドウ, 6
ウィンドウの区切り線
  データ エディタ, 105

エカマックス回転
  因子分析, 398
円グラフ
  度数分布表, 285

オートスクリプト, 512
オプション, 498, 500–501, 503–504, 506, 509, 511–512
  2 桁の年数, 501
  隠す, 501
  出力ラベル, 504
  スクリプト, 512
  図表, 506
  全般, 498
  通貨, 503
  テンポラリ ディレクトリ, 511

ピボット テーブル ルック, 509
ビューア (一般オプション), 500
変数ビュー, 503
重み付き最小 2 乗法
線型回帰, 360
重み付きデータ, 227
再構成データ ファイル, 227
重み付き平均値
比率統計量, 476
重み付き予測値
GLM, 343
重み付けのない最小 2 乗法
因子分析, 397
オンライン ヘルプ, 12
統計コーチ, 11

カイ 2 乗, 422
1 サンプル検定, 422
Fisher の直接法, 299
Pearson の相関係数, 299
Yates の連続修正, 299
オプション, 424
期待値, 424
期待範囲, 424
クロス集計表, 299
欠損値, 424
線型と線型による連関, 299
統計量, 424
独立性, 299
尤度比, 299
カイ 2 乗距離
距離, 357
回帰
作図, 363
線型回帰, 360
複数の回帰, 360
回帰係数
線型回帰, 367
改行
変数ラベルと値ラベル, 91
解説, 269–270
階層クラスタ分析, 411
距離行列, 414
距離測度, 413
クラスタ化の方法, 413
クラスタ凝集経過工程, 414
ケースのクラスタ化, 411
変換測定方法, 413
変換値, 413
コマンドの追加機能, 416
所属クラスタ, 414–415
新変数の保存, 415
つららプロット, 415
デンドログラム, 415
統計量, 411, 414
プロット方向, 415
変数のクラスタ化, 411
類似度の測定方法, 413
例, 411
階層的分解法, 338
外部結合, 25
価格関連格差 (PRD)
比率統計量, 476
科学的表記法, 87, 498
出力での抑制, 498
格子線, 271
ピボット テーブル, 271
拡大と縮小
ピボット テーブル, 261, 265
仮想アクティブ ファイル, 70
カッパ
クロス集計表, 299
カテゴリ データ, 117
区間データを個別カテゴリに変換する, 135
カテゴリを閉じる, 135
環境変数, 511
SPSSTMPDIR, 511
関数, 147
欠損値の処理, 147
観測度数
クロス集計表, 302
順序回帰, 373
観測平均値
GLM 1 変量, 345
ガンマ
クロス集計表, 299
カンマ区切りファイル, 33
カンマ書式, 87, 89
幹葉図
探索的, 294

幾何平均
OLAP キューブ, 316
グループの平均, 311
要約, 306
記述統計記述統計, 287
z 得点の保存, 287
コマンドの追加機能, 289
統計量, 288
表示順, 288
記述統計量
GLM 1 変量, 345
TwoStep クラスタ分析, 409
記述統計, 287
探索的, 293
度数分布表, 283

比率統計量, 476
要約, 306
季節差分関数, 181
期待度数
クロス集計表, 302
順序回帰, 373
基本ステップ, 10
逆 Helmert 対比
GLM, 339–340
逆数モデル
曲線推定, 380
脚注, 261, 269–271
図表, 488
番号の付け直し, 271
マーカー, 261, 270
キャッシュ, 72
アクティブ ファイル, 72
[キャンセル] ボタン, 9
級内相関係数 (ICC)
信頼性分析, 465
共線性の診断
線型回帰, 367
行と列の移動, 252
行と列の入れ替え, 253
行と列の順序の変更, 252
行パーセント
クロス集計表, 302
共分散比
線型回帰, 365
行または列のグループ化, 253
曲線推定, 378
残差の保存, 381
定数を含む, 378
分散分析, 378
モデル, 380
予測, 381
予測区間の保存, 381
予測値の保存, 381
許容度
線型回帰, 367
距離, 355
ケース間の距離の計算, 355
変換測定方法, 357–358
変換値, 357–358
コマンドの追加機能, 358
統計量, 355
非類似度の測定方法, 357
変数間の距離の計算, 355
類似度の測定方法, 358
例, 355
距離測度
階層クラスタ分析, 413
距離, 357

近接
階層クラスタ分析, 411

クォーティマックス回転
因子分析, 398
クラスタの度数
TwoStep クラスタ分析, 409
クラスタ分析
階層クラスタ分析, 411
効率, 419
大規模ファイルのクラスタ分析, 417
クラスタリング
手続きの選択, 402
グラフの保存, 237, 243–244
BMP ファイル, 237, 244
EMF ファイル, 237
EPS ファイル, 237, 244
JPEG ファイル, 237, 243
PICT ファイル, 237
PNG ファイル, 244
PostScript ファイル, 244
TIFF ファイル, 244
メタファイル, 237
グループ化変数, 211
作成する, 211
グループ間の差分
OLAP キューブ, 318
グループ中央値
OLAP キューブ, 316
グループの平均, 311
要約, 306
グループの比較
OLAP キューブ, 318
グループ平均, 309, 314
グループ ラベル, 253
グループ ラベルの削除, 254
グループ ラベルの挿入, 253
グループ ラベルの追加, 253
クロス集計表, 297
クラスタ棒グラフ, 299
小数表記の重み付け, 209
書式, 303
制御変数, 299
セル表示, 302
層, 299
統計量, 299
表の抑制, 297
クロス表
クロス集計表, 297
多重回答, 446

傾向化除去正規プロット
探索的, 294

罫線, 264, 271
  隠れた罫線を表示する, 271
ケース, 102, 211
  新しいケースの挿入, 102
  重み付け, 209
  サブセットの選択, 204, 206, 208
  重複の検索, 132
  データ エディタ内での検索, 103
  並べ替え, 189
  変数への再構成, 211
ケースごとの図
  線型回帰, 367
ケースコントロール研究
  対応のある T 検定, 323
ケース選択変数
  線型回帰, 363
ケースの一覧を表示, 304
ケースの重み付け, 209
  クロス集計表での小数表記の重み付け, 209
ケースの数
  OLAP キューブ, 316
  グループの平均, 311
  要約, 306
ケースのサブセット
  選択, 204, 206, 208
  無作為抽出, 207
ケースの選択, 204
  ケースの範囲, 208
  時刻の範囲, 208
  選択基準に基づく, 206
  日付の範囲, 208
  無作為抽出, 207
ケースの並べ替え, 189
ケースのランク付け, 157
  サベージ スコア, 158
  小数点つき順位, 158
  同順位値, 159
  パーセンタイル, 158
欠損値, 91, 489
  1 サンプルによる Kolmogorov-Smirnov 検定, 431
  1 サンプルの t 検定, 326
  2 項検定, 426
  2 個の対応サンプルの検定, 436
  2 個の独立サンプルの検定, 434
  2 変量の相関分析, 350
  ROC 曲線, 480
  一元配置分散分析, 333
  因子分析, 400
  カイ 2 乗検定, 424
  関数, 147
  探索的, 295
  時系列データの置き換え, 183
  図表, 489
  線型回帰, 369
  対応のあるサンプルの t 検定, 325
  多重回答のクロス集計表, 448
  多重回答の度数表, 444
  定義, 91
  独立サンプルの t 検定, 323
  複数の独立サンプルの検定, 438
  偏相関, 353
  報告書: 行の集計, 454
  文字型変数, 91
  ラン検定, 429
  列の集計報告書, 461
欠損値の置き換え
  系列平均, 184
  周囲値中央値, 184
  周囲値平均, 184
  線型トレンド, 184
  線型補間, 184
言語
  出力言語の変更, 498
検索テーブル, 196
探索的分析, 291
  オプション, 295
  欠損値, 295
  コマンドの追加機能, 296
  作図, 294
  統計量, 293
  べき乗の変換, 295
検索と置換
  ビューア文書, 234
厳密平行モデル
  信頼性分析, 463, 465

効果サイズの推定
  GLM 1 変量, 345
変換行列
  因子分析, 394
合計
  OLAP キューブ, 316
  記述統計, 288
  グループの平均, 311
  度数分布表, 283
  要約, 306
合計パーセント
  クロス集計表, 302
交互作用項, 337, 376
項目の構築, 337, 376
極値
  探索的, 293
固定書式, 33
コマンド言語, 275
コマンド言語を使用したプログラミング, 275
コマンド識別子, 535

コマンド シンタックス, 275, 280, 515, 519, 521
『SPSS Command Syntax Reference』の利用, 12
実行, 280
ジャーナル ファイル, 281
出力ログ, 278
シンタックスの規則, 275
ツールバー ボタンを使用して実行, 519
バッチ ジョブの実行, 521
貼り付け, 277
メニューへの追加, 515
コマンド シンタックス ファイル, 280
コマンド ライン スイッチ, 528
バッチ ジョブ, 528
固有値
因子分析, 396–397
線型回帰, 367

サーバー, 75–76
追加, 76
名前, 76
編集, 76
ポート番号, 76
ログイン, 75
サーバーへのログイン, 75
最後
OLAP キューブ, 316
グループの平均, 311
要約, 306
最初
OLAP キューブ, 316
グループの平均, 311
要約, 306
最小値
OLAP キューブ, 316
記述統計, 288
グループの平均, 311
探索的, 293
度数分布表, 283
比率統計量, 476
要約, 306
列の報告書の比較, 459
最小有意差
GLM, 341
一元配置分散分析, 330
サイズ, 233
アウトライン, 233
サイズの差異測定
距離, 357
最大
OLAP キューブ, 316
記述統計, 288
グループの平均, 311
探索的, 293
度数分布表, 283
比率統計量, 476
要約, 306
列の報告書の比較, 459
最大枝数
TwoStep クラスタ分析, 406
最頻値
度数分布表, 283
最尤法
因子分析, 397
削除ケース残差
GLM, 343
線型回帰, 365
サブグループ平均, 309, 314
サブタイプ, 535
ラベルとの対比, 536
差分関数, 181
サベージ スコア, 158
残差
曲線推定で保存, 381
クロス集計表, 302
線型回帰での保存, 365
残差プロット
GLM 1 変量, 345
参照カテゴリ
GLM, 339–340
散布図
線型回帰, 363

時系列データ
欠損値の置き換え, 183
変換関数, 181
時系列変数の新規作成, 180
データの変換, 178
日付変数の定義, 179
時系列分析
ケースの予測, 381
予測, 381
辞書, 46
指数成長モデル
曲線推定, 380
指定ウィンドウ, 6
自動バッチ, 521
ジャーナル ファイル, 511
尺度, 87, 117
ダイアログ ボックスでのアイコン, 9
定義, 87
尺度モデル
順序回帰, 376
主因子法, 397
集計列
報告書, 459

自由書式, 33
重相関係数
  線型回帰, 367
収束
  因子分析, 397–398
  大規模ファイルのクラスタ分析, 419
従属 *t* 検定
  対応のあるサンプルの t 検定, 323
重複ケース (レコード)
  検索および除外, 132
周辺等質性検定
  2 個の対応サンプルの検定, 434
重要度の図表
  TwoStep クラスタ分析, 408
主成分分析, 394, 397
出現数の計算, 149
出力, 229–231, 236–237, 249, 500
  位置合わせ, 231, 500
  移動, 230
  エクスポート, 237
  コピー, 230
  削除, 230–231
  出力言語の変更, 498
  中心化 (センタリング), 231, 500
  非表示, 230
  ビューア (一般オプション), 229
  表示, 230
  他のアプリケーションにコピーする, 236
  他のアプリケーションへの貼り付け, 236
  保存, 249
出力オブジェクトの種類
  OMS, 534
出力管理システム (OMS), 531, 555
出力の削除, 231
出力の保存, 237, 240, 242
  Excel 形式, 237, 239
  HTML, 237, 239
  HTML 形式, 237
  PDF 形式, 237, 240
  PowerPoint 形式, 237, 240
  Word 形式, 237, 239
  テキスト形式, 237, 242
出力をエクスポートする, 237, 240, 242
  Excel 形式, 237, 239
  HTML, 239
  HTML 形式, 237
  OMS, 531
  PDF 形式, 237, 240
  PowerPoint 形式, 237
  Word 形式, 237, 239
出力を中央に配置, 231, 500
順位相関係数
  2 変量の相関分析, 348
順序, 87
  尺度, 117
  測定レベル, 87
順序回帰 , 371
  位置モデル, 375
  オプション, 372
  コマンドの追加機能, 377
  尺度モデル, 376
  統計量, 371
  リンク, 372
条件付変換, 146
乗算
  列の報告書全体の乗算, 459
小計
  列の集計報告書, 460
初期しきい値
  TwoStep クラスタ分析, 406
除算
  列の報告書全体の除算, 459
書式化
  報告書の列, 452
シンタックス, 275, 280, 518, 521
  『SPSS Command Syntax Reference』の利用, 12
  Unicode コマンド シンタックス ファイル, 280
  実行, 280
  ジャーナル ファイル, 281
  出力ログ, 278
  シンタックスの規則, 275
  ツールバー ボタンを使用してコマンド シンタックスを実行, 518
  バッチ ジョブの実行, 521
  貼り付け, 277
シンタックス ファイル内の複数の EXECUTE を削除, 281
信頼区間
  1 サンプルの t 検定, 326
  GLM, 339, 345
  ROC 曲線, 480
  一元配置分散分析, 333
  探索的, 293
  線型回帰, 367
  線型回帰での保存, 365
  対応のあるサンプルの t 検定, 325
  独立サンプルの t 検定, 323
信頼性分析, 463
  Hotelling の T^2, 465
  Kuder-Richardson 20, 465
  Tukey の加法性の検定, 465
  記述統計, 465
  級内相関係数, 465
  項目間の相関係数および共分散, 465
  コマンドの追加機能, 466
  統計量, 463, 465

分散分析表 (カテゴリ/カテゴリの回帰), 465
例, 463

水準と広がりの図
GLM 1 変量, 345
探索的, 294
推定周辺平均
GLM 1 変量, 345
数値書式, 87, 89
スクリプト, 515, 519
ツールバー ボタンを使用して実行, 519
デフォルト言語。, 512
メニューへの追加, 515
スケール, 87
尺度, 117
信頼性分析, 463
測定レベル, 87
多次元尺度法, 468
スケール変数
ビン分割によるカテゴリ変数の作成, 135
スコアリング
エクスポートとスコアリングに対応するモデル, 184
保存したモデルの読み込み, 186
読み込まれたモデルの表示, 188
スチューデント *t* 検定, 320
スチューデント化された残差
線型回帰, 365
ステータス バー, 7
ステップワイズ法
線型回帰, 362
ストレス
多次元尺度法, 468
図表, 230, 237, 274, 481, 506
ROC 曲線, 478
エクスポート, 237
概要, 481
ケースのラベル, 378
欠損値, 489
サイズ, 489
作成する, 481
図表ビルダー, 481
縦横比, 506
テンプレート, 489, 506
パネルの折り返し, 489
非表示, 230
ピボット テーブルからの作成, 274
他のアプリケーションにコピーする, 236
図表エディタ, 485
プロパティ, 486
図表オプション, 506
図表出力, 236
図表のエクスポート, 237, 243–244, 521
自動バッチ, 521
図表の作成, 481
図表ビルダー, 481
ギャラリ, 482
スプレッドシート ファイル, 16–17, 51
範囲の読み込み, 17
開く, 17
変数名の書き込み, 51
変数名の読み込み, 17
スペース区切りデータ, 33
すべての因子によるモデル
GLM, 337
スペリング, 98
辞書, 501

正規確率プロット
探索的, 294
線型回帰, 363
正規スコア
ケースのランク付け, 158
正規性検定
探索的, 294
制御変数
クロス集計表, 299
先行移動平均関数, 181
成長モデル
曲線推定, 380
セッション ジャーナル, 511
折半法の信頼性
信頼性分析, 463, 465
セル プロパティ, 267–268
線型回帰, 360
重み付きと成分負荷, 360
ケース選択変数, 363
欠損値, 369
コマンドの追加機能, 370
作図, 363
残差, 365
新変数の保存, 365
統計量, 367
ブロック, 360
変数選択方法, 362, 369
モデル情報のエクスポート, 365
線型性の検定
グループの平均, 311
線型と線型による連関
クロス集計表, 299
線型モデル
曲線推定, 380
選択方法, 272
ピボット テーブルでの行と列の選択, 272

尖度
OLAP キューブ, 316
記述統計, 288
グループの平均, 311
探索的, 293
度数分布表, 283
報告書: 列の集計, 458
報告書: 行の集計, 452
要約, 306
尖度の標準誤差
OLAP キューブ, 316
グループの平均, 311
要約, 306
全平均との対比
GLM, 339–340

層, 245, 254, 256, 261, 265
印刷, 245, 261, 265
クロス集計表, 299
作成する, 254
ピボット テーブル, 254
表示する, 254, 256
相関行列
因子分析, 394, 396
順序回帰, 373
判別分析, 389
相関係数
0 次, 353
2 変量の相関分析, 348
クロス集計表, 299
偏相関, 352
総計
列の集計報告書, 461
相対リスク
クロス集計表, 299
属性
ユーザー指定の変数属性, 94
測定システム, 498
速度, 72
データのキャッシュ, 72
その後の多重比較, 330

ダイアログ ボックス, 9, 494–495, 498
項目, 9
変数, 7
変数グループの使用, 495
変数グループの定義, 494
変数情報, 10
変数のアイコン, 9
変数の選択, 9
変数の表示順, 498
変数名の表示, 7, 498
変数ラベルの表示, 7, 498
目標リストの並べ替え, 496
対応サンプル, 434, 439
大規模ファイルのクラスタ分析
概要, 417
クラスタ距離, 420
クラスタ情報の保存, 420
欠損値, 420
効率, 419
コマンドの追加機能, 421
収束基準, 419
所属クラスタ, 420
統計量, 417, 420
反復回数, 419
方法, 417
例, 417
対数モデル
曲線推定, 380
対比
GLM, 339–340
一元配置分散分析, 329
タウ *b*
クロス集計表, 299
タウ *c*
クロス集計表, 299
多項式の対比
GLM, 339–340
一元配置分散分析, 329
多次元尺度法, 468
基準, 472
距離行列の作成, 470
距離測度, 470
変換値, 470
コマンドの追加機能, 473
次元, 471
尺度モデル, 471
尺度レベル, 471
条件付け, 471
データ行列の定義, 470
統計量, 468
表示オプション, 472
例, 468
多重回答
コマンドの追加機能, 449
多重回答グループ
多重 2 分変数, 121
多重カテゴリ, 121
定義, 121
多重回答グループを定義, 443
2 分変数, 443
カテゴリ, 443
変数グループの名前, 443
変数グループのラベル, 443
多重回答のクロス集計表, 446
値の範囲の定義, 448

回答に基づいたパーセント, 448
グループ間で変数を順に整合, 448
ケースに基づいたパーセント, 448
欠損値, 448
セルのパーセント, 448
多重回答の度数表, 444
欠損値, 444
多重回答分析
クロス表, 446
多重回答のクロス集計表, 446
多重回答の度数表, 444
度数分布表, 444
多重比較
一元配置分散分析, 330
縦横比, 506
縦ラベルのテキスト, 254
タブ区切りファイル, 16–17, 33, 48, 51
開く, 16
変数名の書き込み, 51
変数名の読み込み, 17
保存, 48
単純対比
GLM, 339–340

中央値
OLAP キューブ, 316
グループの平均, 311
探索的, 293
度数分布表, 283
比率統計量, 476
要約, 306
抽出
無作為抽出, 207
中心化移動平均関数, 181
中心傾向の測定
探索的, 293
度数分布表, 283
比率統計量, 476
調整済み R^2
線型回帰, 367
調和平均
OLAP キューブ, 316
グループの平均, 311
要約, 306
直接オブリミン回転
因子分析, 398
散らばり係数 (COD)
比率統計量, 476
散らばりの測定
記述統計, 288
探索的, 293
度数分布表, 283
比率統計量, 476
通貨の書式, 503
ツールバー, 516, 518–519
新しいツールの作成, 519
カスタマイズ, 516, 518
作成する, 516, 518
表示と非表示, 516
別のウィンドウで表示, 518
つららプロット
階層クラスタ分析, 415
ツリーの深さ
TwoStep クラスタ分析, 406

ディスク スペース, 70, 72
テンポラリ, 70, 72
ディストリビュート モード, 75–76, 79–81
使用可能な手続き, 81
相対パス, 81
データ ファイル アクセス, 79–80
データ エディタ, 83, 85, 92, 99–103, 105–106, 515
新しいケースの挿入, 102
新しい変数の挿入, 102
位置合わせ, 92
印刷, 106
数値データの入力, 99
データ値の制限, 100
データの型の変更, 103
データの入力, 99
データの編集, 100–101
データ ビュー, 83
非数値データの入力, 100
表示オプション, 105
複数のオープン データ ファイル, 108, 498
複数のビュー/ペイン, 105
分析から除外されたケース, 105
変数の移動, 103
変数の定義, 85
変数ビュー, 84
他のアプリケーションにデータを送信, 515
列幅, 92
データ辞書
別のファイルからの適用, 124
データセット
名前の変更, 111
データセットの名前の変更, 111
データ入力, 99
データのインポート, 16, 19
データのエクスポート, 48, 515
データをエクスポートするためのメニュー項目を追加, 515
データの型, 87, 89, 103, 503
定義, 87
入力書式, 89
表示書式, 89

変更, 103
ユーザーの指定による通貨, 87, 503
データのグループ集計, 199
集計関数, 202
変数名とラベル, 202
データの変換, 501
値の再割り当て, 152–153, 155–156, 160
関数, 147
ケースのランク付け, 157
時系列, 178, 180
実行時間を遅らせる, 501
条件付変換, 146
変数の計算, 144
文字型変数, 147
データの再構成, 210–211, 213–215, 217–225, 227
重み付きデータ, 227
概要, 210
ケースから変数へ再構成するデータの選択, 223
ケースから変数へ再構成する例, 214
ケースから変数へのオプション, 225
ケースから変数への再構成するデータの並べ替え, 224
再構成の種類, 211
変数からケースに再構成する変数グループ, 214
変数からケースへ再構成するインデックス変数の作成, 217
変数からケースへ再構成するインデックス変数を 1 つ作成する, 220
変数からケースへ再構成するインデックス変数を複数作成する, 221
変数からケースへ再構成するデータの選択, 215
変数からケースへ再構成する場合のオプション, 222
変数からケースへ再構成する例, 213
変数からケースへの再構成で 1 つのインデックスを作成する例, 218
変数からケースへの再構成で 2 つのインデックスを作成する例, 219
データの入力, 99–100
値ラベルの使用, 100
数値型, 99
非数値型, 100
データの編集, 100–101
データ ビュー, 83
データ ファイル, 16, 33, 46–48, 54, 72, 79–80, 211
mrInterview, 42
Quancept, 42
Quanvert, 42
SPSS 形式のデータ ファイルとして出力を保存, 531
入れ換え, 192
行と列の入れ換え, 192
コメントの追加, 493
再構成, 211
次元, 42
辞書情報, 46
大規模ファイルを使用する場合のパフォーマンスの向上, 72
テキスト, 33
開く, 16
ファイル情報, 46
複数のオープン データ ファイル, 108, 498
変数のサブグループの保存, 54
保護, 70
保存, 47–48
リモート サーバー, 79–80
データ ファイルの結合
異なるケースのファイル, 193
異なる変数のファイル, 196
辞書情報, 195
変数名の変更, 195
データ分析, 10
基本ステップ, 10
データベース, 19–20, 23, 25–26, 29–30, 32
Microsoft Access, 20
SQL シンタックス, 32
値のプロンプト, 29
既存フィールドの値の置換, 62
クエリーの保存, 32
結果の確認, 32
更新する, 55
条件式, 26
条件節, 26
条件の指定, 26
新規テーブルを作成する, 66
データ ソースの選択, 20
データ フィールドの選択, 23
テーブル結合, 25
テーブルに新規フィールドを追加する, 63
テーブルを置換する, 66
パラメータ クエリー, 26, 29
変数の定義, 30
保存, 55
無作為抽出, 26
文字列の数値型変数への変換, 30
読み込み, 19–20, 23
リレーションシップの作成, 25
レコード (ケース) をテーブルに追加する, 64
テーブル, 273
位置合わせ, 268
セル プロパティ, 267–268
テーブル分割の制御, 273
背景色, 267
フォント, 267
余白, 268

テーブルから作成する図表, 274
テーブルのサブタイプ, 535
  ラベルとの対比, 536
テーブル分割, 273
テーブルルック, 258–260
  作成する, 260
  適用, 259
テーブルを分割する, 273
  テーブル分割の制御, 273
適合度
  順序回帰, 373
テキスト, 33, 233–234, 237, 242
  データ ファイル, 33
  テキストとして出力をエクスポートする, 237, 242
  ビューアにテキスト ファイルを追加する, 234
  ビューアへの追加, 233
てこ比の値
  GLM, 343
  線型回帰, 365
デフォルト ファイル位置, 511
デンドログラム
  階層クラスタ分析, 415
テンプレート, 92–93, 489, 506
  外部の SPSS データ ファイルをテンプレートとして使用, 124
  図表, 489, 506
  変数の定義, 92–93
テンポラリ アクティブ ファイル, 72
テンポラリ ディスク スペース, 70, 72
テンポラリ ディレクトリ, 511
  SPSSTMPDIR 環境変数, 511
  ローカル モードでの位置の設定, 511

統計コーチ, 11
等分散性の検定
  GLM 1 変量, 345
  一元配置分散分析, 333
独立したサンプルの t 検定, 320
  オプション, 323
  グループ化変数, 322
  グループの定義, 322
  欠損値, 323
  信頼区間, 323
  文字型変数, 322
独立性の検定
  カイ 2 乗, 299
度数 (プロビット分析), 282
  書式, 286
  図表, 285
  統計量, 283
  表示順, 286
  表の抑制, 286
度数分布表
  探索的, 293
  度数分布表, 282
ドル書式, 87, 89

内部結合, 25
並べ替え
  変数, 190

入力書式, 89

年, 501
  2 桁の年数, 501

対応のある T 検定, 323
  オプション, 325
  欠損値, 325
  対応のある変数の選択, 323
ノイズ処理
  TwoStep クラスタ分析, 406
濃度インデックス
  比率統計量, 476
ノンパラメトリック検定
  1 サンプルによる Kolmogorov-Smirnov 検定, 429
  2 個の対応サンプルの検定, 434
  2 個の独立サンプルの検定, 431
  カイ 2 乗, 422
  複数の対応サンプルの検定, 439
  複数の独立サンプルの検定, 436
  ラン検定, 427

パーセンタイル
  探索的, 293
  度数分布表, 283
パーセント
  クロス集計表, 302
背景色, 267
配置, 231, 500
  出力, 231, 500
箱ひげ図
  因子の水準の比較, 294
  探索的, 294
  変数の比較, 294
外れ値
  TwoStep クラスタ分析, 406
  探索的, 293
  線型回帰, 363
パターン行列
  因子分析, 394

パターンの差異測定
距離, 357
バッチ機能, 498
ジャーナル ファイルからコマンド シンタックスを使用する, 498
ファイルをバッチ ジョブに変換, 529
バッチ ジョブ, 521, 526–528
コマンド ライン スイッチ, 528
出力ファイル, 521
シンタックスの規則, 521
シンタックス ファイルの値の置き換え, 526
図表のエクスポート, 521
バッチ ジョブのスケジュール, 528
バッチ ジョブ ファイル変換, 529
複数のバッチ ジョブの実行, 528
幅の広いテーブル
Microsoft Word への貼り付け, 236
パフォーマンス, 72
データのキャッシュ, 72
パラメータ推定値
GLM 1 変量, 345
順序回帰, 373
[貼り付け] ボタン, 9
バリマックス回転
因子分析, 398
範囲
OLAP キューブ, 316
記述統計, 288
グループの平均, 311
度数分布表, 283
比率統計量, 476
要約, 306
反復回数
因子分析, 397–398
大規模ファイルのクラスタ分析, 419
反復測定
GLM, 339–340
反復の記述
順序回帰, 373
判別分析, 386
Mahalanobis の距離, 390
Rao の *V*, 390
Wilks のラムダ, 390
関数係数, 389
記述統計量, 389
基準, 390
行列, 389
グループ化変数, 386
ケースの選択, 388
欠損値, 391
コマンドの追加機能, 393
作図, 391
事前確率, 391
ステップワイズ法, 386
統計量, 386, 389
独立変数, 386
範囲の定義, 388
判別分析法, 390
表示オプション, 390–391
分散共分散行列, 391
分類変数の保存, 393
モデル情報のエクスポート, 393
例, 386

比率推定値
ケースのランク付け, 158
比率統計量, 474
統計量, 476
ヒストグラム
探索的, 294
線型回帰, 363
度数分布表, 285
日付書式変数, 87, 89, 501
日付時刻変数の一部の抽出, 163
日付時刻変数の加算または減算, 163
変数のグループからの日付時刻変数の作成, 163
文字列からの日付時刻変数の作成, 163
日付の書式
2 桁の年数, 501
日付変数
時系列データの定義, 179
非表示, 230, 257–258, 270, 516
解説, 270
脚注, 270
行と列, 257
次元ラベル, 258
ツールバー, 516
手続きの出力結果, 230
表題, 258
ピボット
エクスポートした出力の OMS による制御, 548
ピボット テーブル, 230, 236–237, 245, 251–254, 257–258, 260–262, 264–265, 271–274, 509
HTML としてエクスポート, 237
アイコンの使用方法, 252
新しいテーブルのデフォルト ルック, 509
位置合わせ, 268
印刷タイトル, 265
大きなテーブルの印刷, 273
解説, 269–270
階層の印刷, 245
格子線, 271
隠れた罫線を表示する, 271
脚注, 269–270
脚注プロパティ, 261
行と列の移動, 252

行と列の入れ替え, 253
行と列の選択, 272
行または列のグループ化, 253
行または列のグループ解除, 254
グループ ラベルの削除, 254
グループ ラベルの挿入, 253
罫線, 264
セルの書式, 262
セルの表示と非表示, 257
セル幅, 271
セル プロパティ, 267–268
全般的なプロパティ, 261
層, 254
操作, 251
テーブルからの図表の作成, 274
テーブルとして貼り付け, 236
テーブル分割の制御, 273
デフォルトの列幅の調整, 509
背景色, 267
非表示, 230
ピボット, 251–252
表示順の変更, 252
フォント, 267
プロパティ, 260
ページに合わせたサイズ変更, 261, 265
編集, 251
他のアプリケーションにコピーする, 236
他のアプリケーションへの貼り付け, 236
余白, 268
ラベルの回転, 254
ルックの変更, 258
ピボット テーブルの色, 264
罫線, 264
ピボット テーブルのセル, 262, 270–271
書式, 262
幅, 271
非表示, 257
表示, 257
ビューア
情報の検索と置換, 234
ビューア (一般オプション), 229–233, 248–249, 500, 504
OMS による出力の種類の除外, 543
アウトライン, 232
アウトライン ウィンドウ枠, 229
アウトライン サイズの変更, 233
アウトライン フォントの変更, 233
アウトライン レベルの変更, 232
アウトラインを閉じる, 232
アウトラインを開く, 232
値ラベルの表示, 504
出力結果のウィンドウ枠, 229
出力結果を隠す, 230
出力項目間のスペース, 248
出力の移動, 230
出力の削除, 231
データ値の表示, 504
表示オプション, 500
文書の保存, 249
変数名の表示, 504
変数ラベルの表示, 504
表示, 230, 258, 270, 516
解説, 270
脚注, 270
行または列, 258
次元ラベル, 258
出力結果, 230
ツールバー, 516
表題, 258
表示順, 252
表示書式, 89
標準化
TwoStep クラスタ分析, 406
標準回帰係数
線型回帰, 367
標準化されていない残差
GLM, 343
標準化残差
GLM, 343
線型回帰, 365
標準化した値
記述統計, 287
標準誤差
GLM, 343, 345
ROC 曲線, 480
記述統計, 288
探索的, 293
度数分布表, 283
標準偏差
GLM 1 変量, 345
OLAP キューブ, 316
記述統計, 288
グループの平均, 311
探索的, 293
度数分布表, 283
比率統計量, 476
報告書: 列の集計, 458
報告書: 行の集計, 452
要約, 306
表題, 233
OLAP キューブ, 319
図表, 488
ビューアへの追加, 233
ビン, 135
ビン分割, 135

ファイ
クロス集計表, 299
ファイ 2 乗距離
距離, 357
ファイル, 234
ビューアにテキスト ファイルを追加する, 234
開く, 16
ファイル位置
デフォルト ファイル位置の制御, 511
ファイル変換, 210–211
ケースの重み付け, 209
ケースの並べ替え, 189
データのグループ集計, 199
データの再構成, 210–211
データ ファイルの結合, 193, 196
ファイルの分割処理, 203
変数とケースの入れ換え, 192
ファイル情報, 46
ファイルの分割処理, 203
ファイルの保存, 47–48
SPSS データ ファイル, 47
データ ファイル, 48
データベース ファイル クエリー, 32
デフォルト ファイル位置の制御, 511
ファイルを開く, 16–19, 33
dBASE ファイル, 16, 18
Excel ファイル, 16–17
Lotus 1-2-3 ファイル, 16
Stata ファイル, 18
SYSTAT ファイル, 16
スプレッドシート ファイル, 16–17
タブ区切りファイル, 16
データ ファイル, 16
テキスト データ ファイル, 33
デフォルト ファイル位置の制御, 511
フォント, 105, 233, 267
アウトライン ウィンドウ枠, 233
データ エディタ, 105
不確定性係数
クロス集計表, 299
複合成長モデル
曲線推定, 380
複数のオープン データ ファイル, 108, 498
抑制, 112
複数の回帰
線型回帰, 360
複数の対応サンプルの検定, 439
検定の種類, 440
コマンドの追加機能, 441
統計量, 441
複数の独立サンプルの検定, 436
オプション, 438
グループ化変数, 438
欠損値, 438
検定の種類, 437
コマンドの追加機能, 439
統計量, 438
範囲の定義, 438
複数のビュー/ペイン
データ エディタ, 105
副題
図表, 488
符号検定
2 個の対応サンプルの検定, 434
フッター, 246
ブレーク変数
集計データ, 199
ブロック距離
距離, 357
プロットのロード
因子分析, 398
プロパティ, 260–261
テーブル, 261
ピボット テーブル, 260
プロファイル プロット
GLM, 340
分割係数
クロス集計表, 299
分割表, 297
分散
OLAP キューブ, 316
記述統計, 288
グループの平均, 311
探索的, 293
度数分布表, 283
報告書: 列の集計, 458
報告書: 行の集計, 452
要約, 306
分散拡大係数
線型回帰, 367
分散共分散行列
GLM, 343
順序回帰, 373
線型回帰, 367
判別分析, 389, 391
分散分析
一元配置分散分析, 328
曲線推定, 378
グループの平均, 311
線型回帰, 367
分散分析 (分散成分 オプション)
GLM 1 変量, 335
一元配置分散分析, 328
グループの平均, 311
モデル, 337
分析から除外されたケース, 105
データ エディタ, 105

分布の測定
記述統計, 288
度数分布表, 283
分類
ROC 曲線, 478

平滑化関数, 181
平均絶対偏差 (AAD)
比率統計量, 476
平均値
OLAP キューブ, 316
一元配置分散分析, 333
記述統計, 288
グループの平均, 311
探索的, 293
サブグループ, 309, 314
度数分布表, 283
比率統計量, 476
複数の列の報告書の, 459
報告書: 列の集計, 458
報告書: 行の集計, 452
要約, 306
平均値 (信頼性分析), 309
オプション, 311
統計量, 311
平均値の標準誤差
OLAP キューブ, 316
グループの平均, 311
要約, 306
平行線の検定
順序回帰, 373
平行モデル
信頼性分析, 463, 465
平方ユークリッド距離
距離, 357
平方和, 338
GLM, 337
ペインの区切り線
データ エディタ, 105
ページ設定, 246, 248
図表サイズ, 248
ヘッダーとフッター, 246
ページの指定
行の集計報告書, 454
列の集計報告書, 460
ページの番号付け, 248
行の集計報告書, 454
列の集計報告書, 461
べき乗推定値
GLM 1 変量, 345
べき乗モデル
曲線推定, 380
ヘッダー, 246

[ヘルプ] ウィンドウ, 12
[ヘルプ] ボタン, 9
偏 2 段階最小 2 乗回帰, 382
変数をエクスポート, 385
モデル, 384
変数, 9, 85, 102–103, 211, 492, 494, 498
新しい変数の挿入, 102
移動, 103
ケースへの再構成, 211
結合したデータ ファイルの名前変更, 195
再割り当て, 152–153, 155–156, 160
ダイアログ ボックス, 7
ダイアログ ボックスでの選択, 9
ダイアログ ボックスでの表示順, 498
ダイアログ ボックスでの変数情報, 10
定義, 85
定義情報, 492
データ エディタ内での検索, 103
並べ替え, 190
変数グループの定義, 494
変数増加法
線型回帰, 362
変数間の差分
OLAP キューブ, 318
変数グループ, 494–495
使用, 495
定義, 494
変数情報, 492
変数とケースの入れ換え, 192
変数の計算, 144
新しい文字型変数の計算, 147
変数の重要度の作図
TwoStep クラスタ分析, 408
変数の属性, 92–93
コピーと貼り付け, 92–93
ユーザーの指定による, 94
変数の定義, 85, 87, 89–93, 114
値ラベル, 90, 114
欠損値, 91
属性のコピーと貼り付け, 92–93
データ辞書の適用, 124
データの型, 87
テンプレート, 92–93
変数ラベル, 89
変数の並べ替え, 190
変数の比較
OLAP キューブ, 318
変数ビュー, 84
カスタマイズ, 98, 503
変数ペア, 211
作成する, 211
変数減少法
線型回帰, 362

変数名, 85, 498
OMS による生成, 550
従来のバージョンにおける長い変数名の切り詰め, 48
出力における長い変数名の改行, 85
ダイアログ ボックス, 7, 498
複合ケース変数名, 85
ポータブル ファイル, 48
ルール, 85
変数ラベル, 89, 498, 504
アウトライン ウィンドウ枠, 504
改行の挿入, 91
結合したデータ ファイル, 195
ダイアログ ボックス, 7, 498
ピボット テーブル, 504
変数リスト, 496
目標リストの並べ替え, 496
偏相関, 352
0 次相関, 353
オプション, 353
欠損値, 353
コマンドの追加機能, 354
線型回帰, 367
統計量, 353
偏相関プロット
線型回帰, 363
変動係数 (COV)
比率統計量, 476

棒グラフ
度数分布表, 285
報告書
行の集計報告書, 450
合計の列, 459
集計列, 459
列の値の乗算, 459
列の値の除算, 459
列の集計報告書, 456
列の比較, 459
報告書の行の集計, 450
欠損値, 454
コマンドの追加機能, 461
ソート順, 450
データ列, 450
表題, 455
表題の変数, 455
フッター, 455
ブレーク間隔, 453
ブレーク列, 450
ページの指定, 453
ページの設計, 454
ページの番号付け, 454
列の書式, 452
報告書の列の集計, 456
欠損値, 461
コマンドの追加機能, 461
集計列, 459
小計, 460
総計, 461
ページの指定, 460
ページの設計, 454
ページの番号付け, 461
列の書式, 452
ポータブル ファイル
変数名, 48
ポート番号, 76

無作為抽出, 26
選択, 207
データベース, 26
乱数のシード, 148

名義, 87
尺度, 117
測定レベル, 87
メタファイル, 237
図表のエクスポート, 237
メディアン検定
2 個の独立サンプルの検定, 436
メニュー, 515
カスタマイズ, 515
メモリー, 498
メモリ割り当て
TwoStep クラスタ分析, 406

目標リスト, 496
文字型変数, 91, 100
新しい文字型変数の計算, 147
欠損値, 91
従来のバージョンにおける長い文字型の細分化, 48
ダイアログ ボックス, 7
データの入力, 100
連続整数への再割り当て, 160
文字書式, 87
モデル ファイル
データのスコアリング、保存したモデルの読み込み, 186

ユーザー欠損値, 91
ユーザー指定の属性, 94
ユーザー指定の変数属性, 94
ユーザーの指定による通貨の書式, 87, 503

ユーザーの指定によるモデル
GLM, 337
尤度比カイ 2 乗
クロス集計表, 299
順序回帰, 373

要約, 304
オプション, 306
統計量, 306
予測
曲線推定, 381
予測区間
曲線推定で保存, 381
線型回帰での保存, 365
予測値
曲線推定で保存, 381
線型回帰での保存, 365

ラップ, 261
変数ラベルと値ラベル, 91
ラップされたテキストの列幅のコントロール, 261
ラベル, 253–254
OMS のサブタイプ名との対比, 536
グループ ラベルの挿入, 253
削除, 254
ラベルの回転, 254
ラムダ
クロス集計表, 299
ラン検定
オプション, 429
欠損値, 429
コマンドの追加機能, 429
統計量, 429
分割点, 427–428
乱数のシード, 148

リード関数, 181
リスク
クロス集計表, 299
[リセット] ボタン, 9
リモート サーバー, 75–76, 79–81
使用可能な手続き, 81
相対パス, 81
追加, 76
データ ファイル アクセス, 79–80
編集, 76
ログイン, 75
リンク
順序回帰, 372

類似度の測定方法
階層クラスタ分析, 413
距離, 358
累積合計関数, 181
累積度数
順序回帰, 373

列, 271–272
ピボット テーブルでの選択, 272
ピボット テーブルでの幅の変更, 271
列の集計報告書, 456
列パーセント
クロス集計表, 302
列幅, 92, 261, 271, 509
最大幅のコントロール, 261
データ エディタ, 92
デフォルトの幅をコントロールする, 509
ピボット テーブル, 271
ラップされたテキストの幅のコントロール, 261
連続変数のカテゴリ化, 135

ロー
2 変量の相関分析, 348
クロス集計表, 299
ロケールのエンコード, 280
ロジスティック モデル
曲線推定, 380

歪度
OLAP キューブ, 316
記述統計, 288
グループの平均, 311
探索的, 293
度数分布表, 283
報告書: 列の集計, 458
報告書: 行の集計, 452
要約, 306
歪度の標準誤差
OLAP キューブ, 316
グループの平均, 311
要約, 306