

# IBM SPSS Complex Samples 19



Note: Before using this information and the product it supports, read the general information under Notices p. 293 .

This document contains proprietary information of SPSS Inc, an IBM Company. It is provided under a license agreement and is protected by copyright law. The information contained in this publication does not include any product warranties, and any statements provided in this manual should not be interpreted as such.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

**© Copyright SPSS Inc. 1989, 2010.**

---

# はじめに

IBM® SPSS® Statistics は、データ分析の包括的システムです。コンプレックス サンプル は、このマニュアルで説明されている追加の分析手法を提供するオプションのアドオン モジュールです。コンプレックス サンプル アドオン モジュールは SPSS Statistics Core システムと組み合わせて使用し、Core システムに完全に統合されます。

## SPSS Inc., an IBM Company について

SPSS Inc., an IBM Company は、余剰分析ソフトウェアおよびソリューションの世界的なリーディングカンパニーです。当社のデータ収集、統計、モデリング、展開という製品の包括的なポートフォリオによりお客様の考えや意見を収集、見込み客との対話の結果を予測、分析を業務プロセスに組み込むことによりこれらの見解に判断を下すことができます。SPSS Inc. のソリューションにより、分析、IT アーキテクチャ、業務プロセスの収束に焦点を当て、組織全体の相互接続した経営目標に取り組みます。世界中の民間、政府、学術分野のお客様が SPSS Inc. のテクノロジーを包括的に利用しています。お客様の関心を呼び、拡大する一方、不正やリスクを軽減、緩和します。2009 年 10 月、SPSS Inc. は IBM 社に買収されました。詳細は <http://www.spss.com> をご覧ください。

## テクニカル サポート

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 Inc. 製品の使用方法や、対応してい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ることもできます。テクニカル サポートにご連絡するには、<http://support.spss.com> の SPSS Inc. Web サイトを参照いただくか、<http://support.spss.com/default.asp?refpage=contactus.asp> の Web サイトでお近くの営業所にお問い合わせください。連絡の際は、所属団体名、サポート契約などを確認できるよう、あらかじめ手元にご用意ください。

## 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> です)。お問い合わせの際には、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)。

## 追加の出版物

Marija Noruš による『SPSS Statistics: Guide to Data Analysis』、『SPSS Statistics: Statistical Procedures Companion』、『SPSS Statistics: Advanced Statistical Procedures Companion』が Prentice Hall から出版されました。補助的な資料としてご利用いただけます。これらの出版物には、SPSS Statistics Base モジュール、Advanced Statistics モジュール、Regression モジュールの統計的手続きについて記載されています。初めてデータ分析を行う場合、高度なアプリケーションを使用する場合に応じて、この本は IBM® SPSS® Statistics が提供している機能を効率よく使用するための手助けとなります。出版物の内容、サンプルの図表などの詳細は、作者の Web サイトを参照してください。 <http://www.norusis.com>

## パート I: ユーザー ガイド

<b>1</b>	<b>コンプレックス サンプル手続きの概要</b>	<b>1</b>
	コンプレックス サンプル オプションのプロパティ . . . . .	1
	コンプレックス サンプル手続きの使用 . . . . .	2
	計画ファイル . . . . .	2
	参考文献 . . . . .	3
<b>2</b>	<b>コンプレックス デザインからの抽出</b>	<b>4</b>
	新規サンプル計画の作成 . . . . .	5
	サンプリング ウィザード: 計画変数 . . . . .	6
	サンプリング ウィザードのツリー表示の操作方法 . . . . .	7
	サンプリング ウィザード: 抽出方法 . . . . .	8
	サンプリング ウィザード: サンプル サイズ . . . . .	10
	等しくないサイズの定義 . . . . .	11
	サンプリング ウィザード: 出力変数 . . . . .	12
	サンプリング ウィザード: 計画の要約 . . . . .	13
	サンプリング ウィザード: サンプルの抽出: 選択オプション . . . . .	14
	サンプリング ウィザード: サンプルの抽出: 出力ファイル . . . . .	15
	サンプリング ウィザード: 終了 . . . . .	16
	既存のサンプル計画の修正 . . . . .	17
	サンプリング ウィザード: 計画の要約 . . . . .	18
	既存のサンプル計画の実行 . . . . .	18
	CSPLAN および CSSELECT コマンドの追加機能 . . . . .	19
<b>3</b>	<b>コンプレックス サンプルの分析準備</b>	<b>20</b>
	新規分析計画の作成 . . . . .	21

分析準備ウィザード: 計画変数	22
分析ウィザードのツリー表示の操作方法	23
分析準備ウィザード: 推定方法	23
分析準備ウィザード: サイズ	24
等しくないサイズの定義	25
分析準備ウィザード: 計画の要約	26
分析準備ウィザード: 終了	27
既存の分析計画の修正	27
分析準備ウィザード: 計画の要約	28
<b>4   コンプレックス サンプルの計画</b>	<b>29</b>
<b>5   コンプレックス サンプル度数分布表</b>	<b>30</b>
コンプレックス サンプルの度数分布表の統計量	31
コンプレックス サンプルの欠損値	32
コンプレックス サンプルのオプション	33
<b>6   コンプレックス サンプル記述統計</b>	<b>34</b>
コンプレックス サンプルの記述統計の統計量	36
コンプレックス サンプルの記述統計の欠損値	37
コンプレックス サンプルのオプション	37
<b>7   コンプレックス サンプルのクロス集計表</b>	<b>38</b>
コンプレックス サンプルのクロス集計表の統計量	40
コンプレックス サンプルの欠損値	41
コンプレックス サンプルのオプション	42
<b>8   コンプレックス サンプル比率</b>	<b>43</b>
コンプレックス サンプルの比率分析の統計量	44

コンプレックス サンプルの比率分析の欠損値 . . . . .	45
コンプレックス サンプルのオプション . . . . .	46
<b>9   コンプレックス サンプルの一般線型モデル</b>	<b>47</b>
コンプレックス サンプルの一般線型モデルの統計量 . . . . .	50
コンプレックス サンプルの仮説の検定 . . . . .	52
コンプレックス サンプルの一般線型モデルの推定平均 . . . . .	53
コンプレックス サンプルの一般線型モデルの保存 . . . . .	54
コンプレックス サンプルの一般線型モデルのオプション . . . . .	55
CSGLM コマンドの追加機能 . . . . .	56
<b>10  コンプレックス サンプルのロジスティック回帰</b>	<b>57</b>
コンプレックス サンプルのロジスティック回帰の参照カテゴリ . . . . .	59
コンプレックス サンプルのロジスティック回帰のモデル . . . . .	60
コンプレックス サンプルのロジスティック回帰の統計量 . . . . .	62
コンプレックス サンプルの仮説の検定 . . . . .	63
コンプレックス サンプルのロジスティック回帰のオッズ比 . . . . .	64
コンプレックス サンプルのロジスティック回帰の保存 . . . . .	65
コンプレックス サンプルのロジスティック回帰のオプション . . . . .	66
CSLOGISTIC コマンドの追加機能 . . . . .	67
<b>11  コンプレックス サンプルの順序回帰</b>	<b>69</b>
コンプレックス サンプルの順序回帰の応答確率 . . . . .	71
コンプレックス サンプルの順序回帰のモデル . . . . .	72
コンプレックス サンプルの順序回帰の統計量 . . . . .	74
コンプレックス サンプルの仮説の検定 . . . . .	76
コンプレックス サンプルの順序回帰のオッズ比 . . . . .	77
コンプレックス サンプルの順序回帰の保存 . . . . .	78
コンプレックス サンプルの順序回帰のオプション . . . . .	80
CSORDINAL コマンドの追加機能 . . . . .	81

## 12 コンプレックス サンプル Cox 回帰 82

事象の定義	85
予測変数	86
時間依存の予測変数の定義	87
サブグループ	89
モデル	90
統計量	92
作図	94
仮説の検定	95
保存	96
エクスポート	98
オプション	100
CSCOXREG コマンドの追加機能	102

## パート II: 例

## 13 コンプレックス サンプルのサンプリング ウィザード 104

完全サンプリング枠からのサンプル抽出	104
ウィザードの使用方法	104
計画の要約	114
抽出の要約	114
抽出結果	116
部分サンプリング枠からのサンプル抽出	116
ウィザードによる最初の部分枠からの抽出	117
抽出結果	130
ウィザードによる 2 番目の部分枠からの抽出	130
抽出結果	135
確率比例 (PPS) 法を使用した抽出	135
ウィザードの使用方法	136
計画の要約	147
抽出の要約	147
抽出結果	149
関連手続き	151



## 14 コンプレックス サンプルの分析準備ウィザード 152

コンプレックス サンプルの分析準備ウィザードによる NHIS 一般使用データの準備	152
ウィザードの使用方法	152
要約表	155
抽出重み付けがデータ ファイル内に存在しない場合の分析準備	155
包含確率と抽出重み付けの計算	155
ウィザードの使用方法	158
要約表	165
関連手続き	166

## 15 コンプレックス サンプル度数分布表 167

コンプレックス サンプルの度数分布表による栄養補給剤使用の分析	167
分析の実行	167
度数分布表	170
部分母集団による度数分布表	170
集計 (報告書 データ列)	171
関連手続き	171

## 16 コンプレックス サンプル記述統計 172

コンプレックス サンプルの記述統計による運動水準の分析	172
分析の実行	172
1 変量統計	175
部分母集団による 1 変量統計	176
集計 (報告書 データ列)	176
関連手続き	176

## 17 コンプレックス サンプルのクロス集計表 178

コンプレックス サンプルのクロス集計によるイベントの相対リスクの測定	178
分析の実行	178
クロス表	182
リスク推定値	182

部分母集団によるリスク推定値 . . . . .	183
集計 (報告書 データ列) . . . . .	184
関連手続き . . . . .	184
<b>18 コМПレックス サンプル比率</b>	<b>185</b>
コМПレックス サンプルの比率分析による財産価値評価の支援 . . . . .	185
分析の実行 . . . . .	185
比率 . . . . .	188
ピボットされた比率分析表 . . . . .	189
集計 (報告書 データ列) . . . . .	189
関連手続き . . . . .	190
<b>19 コМПレックス サンプルの一般線型モデル</b>	<b>191</b>
コМПレックス サンプルの一般線型モデルを使用した二元配置分散分析の適合 . . . . .	191
分析の実行 . . . . .	191
モデルの要約 (ピボットテーブル 回帰) . . . . .	197
モデル効果の検定 . . . . .	197
パラメータ推定値 . . . . .	198
推定周辺平均 (GLM) . . . . .	199
集計 (報告書 データ列) . . . . .	202
関連手続き . . . . .	202
<b>20 コМПレックス サンプルのロジスティック回帰</b>	<b>204</b>
コМПレックス サンプルのロジスティック回帰を使用した信用リスクへのアクセス . . . . .	204
分析の実行 . . . . .	204
擬 R2 乗 . . . . .	208
分類 . . . . .	209
モデル効果の検定 . . . . .	210
パラメータ推定値 . . . . .	210
オッズ比 . . . . .	211
集計 (報告書 データ列) . . . . .	212
関連手続き . . . . .	213

## 21 コンプレックス サンプルの順序回帰 214

コンプレックス サンプルの順序回帰を使用した調査結果の分析	214
分析の実行	214
擬 R2 乗	219
モデル効果の検定	220
パラメータ推定値	220
分類	222
オッズ比	223
一般化累積モデル	224
有意でない予測変数の削除	225
警告	227
モデルの比較	228
集計 (報告書 データ列)	229
関連手続き	229

## 22 コンプレックス サンプルの Cox 回帰 230

時間依存の予測変数の使用 (コンプレックス サンプルの Cox 回帰)	230
データの準備	230
分析の実行	236
サンプル デザインの情報	242
モデル効果の検定	242
ハザードの比例検定	242
時間依存の予測変数の追加	243
コンプレックス サンプルの Cox 回帰の被験者ごとに複数のケース	247
分析用データの準備	248
単純な無作為抽出分析計画の作成	264
分析の実行	268
サンプル デザインの情報	276
モデル効果の検定	277
パラメータ推定値	277
パターン値	278
ログマイナスログ プロット	279
集計 (報告書 データ列)	280

## 付録

A サンプル ファイル	281
B Notices	293
参考文献	296
索引	298

# パート I: ユーザー ガイド



# コンプレックス サンプル手続き の概要

従来のソフトウェア パッケージの分析手続きは、データ ファイル内の観測値は対象母集団の単純な無作為サンプルであるという前提で行われています。構造的な方法でサンプルを得る方がコスト効率が良く便利であると見なす企業や調査機関が増えるにつれ、この考え方は成り立たなくなりつつあります。

コンプレックス サンプル オプションでは、コンプレックス デザインに従ってサンプルを選択し、計画の指定をデータ分析に取り入れることで、確実に有効な結果が得られます。

## コンプレックス サンプル オプションのプロパティ

コンプレックス サンプルは、多くの点で単純な無作為サンプルと異なっています。単純な無作為サンプルでは、個々の抽出単位は置換を行わずに (WOR) 母集団全体から直接等しい確率で無作為に選択されます。これに対して、コンプレックス サンプルは次の特徴のいずれか、またはすべてを持ちます。

**層化。**層化抽出では、母集団内の重なり合わない部分母集団、つまりストラータ内で独立してサンプルが選択されます。たとえば、ストラータには社会経済グループ、職種カテゴリ、年齢グループ、民族グループなどが考えられます。層化を使用することで、対象となるサブグループに適切なサンプル サイズを保証し、全体の推定値の精度を向上し、各ストラータで異なる抽出方法を使用できます。

**クラスタ化。**クラスタ抽出では、抽出単位のグループ、つまりクラスタの選択が行われます。たとえば、クラスタを学校、病院、地域とすると、抽出単位は学生、患者、市民になります。クラスタ化は、複数の段階を持つデザインや地域サンプルで一般的に使用されます。

**複数段階。**複数段階の抽出では、第 1 段階のサンプルをクラスタに基づいて選択します。次に選択したクラスタからサブサンプルを抽出して、第 2 段階を作成します。第 2 段階のサンプルがサブクラスタに基づいている場合、サンプルに第 3 段階を追加できます。たとえば、調査の第 1 段階で都市のサンプルを抽出するとします。次に選択した都市から、世

帯を抽出します。最後に、選択した世帯の個人に対して調査を行います。サンプリングウィザードと分析準備ウィザードを使って、計画の3つの段階を指定できます。

**無作為でない抽出。**無作為に選択することが難しい場合、単位を系統的（固定間隔）または順番に抽出できます。

**不等選択確率。**含まれる単位数が等しくない複数のクラスタを抽出する場合、確率比例（PPS）抽出を使用してクラスタの選択確率をクラスタに含まれる単位の割合に等しくできます。PPS抽出は、単位を選択するための一般的な重み付け方法にも使用できます。

**制限されない抽出。**制限されない抽出では置換（WR）を使用して単位が選択されるため、。各単位が複数回選択されて抽出されることがあります。

**抽出重み付け。**抽出重み付けは、コンプレックス サンプルの抽出中に自動的に計算されます。理論的には、各抽出単位が対象母集団に現れる「頻度」に応じた値になります。したがって、サンプル全体の重み付けの総計から、母集団のサイズが推定できます。コンプレックス サンプルの分析手続きでは、コンプレックス サンプルを適切に分析するために抽出重み付けが必要になります。重み付けはコンプレックス サンプル オプションの内部だけで使用し、重み付けをケースの繰り返しとして扱う [ケースの重み付け] 手続きを利用して他の分析手続きで使用しないように注意してください。

## コンプレックス サンプル手続きの使用

コンプレックス サンプル手続きは、ユーザーの必要に応じて使用します。主に次のようなユーザーが使用します。

- コンプレックス デザインに従って調査を計画し実行した後、サンプルを分析する可能性があるユーザー。調査者の主要なツールは、[サンプリングウィザード](#)です。
- 以前に取得したサンプル データ ファイルをコンプレックス デザインに従って分析するユーザー。コンプレックス サンプル分析手続きを使用する前に、[分析準備ウィザード](#)を使用する必要がある場合があります。

どのタイプのユーザーでも、コンプレックス サンプル手続きに計画情報を指定する必要があります。この情報は、簡単に再利用できるように計画ファイル内に格納されます。

## 計画ファイル

計画ファイルはコンプレックス サンプル（複雑なサンプル）を含んでいます。計画ファイルには次の2つのタイプがあります。



**抽出計画。**サンプリング ウィザードで指定した内容によって、コンプレックス サンプルの抽出に使用されるサンプル計画が定義されます。抽出計画ファイルには、その指定内容が含まれます。また、抽出計画ファイルには、指定されたサンプル計画に適した推定方法を使用するデフォルトの分析計画も含まれています。

**分析計画。**この計画ファイルには、コンプレックス サンプル分析手続きでコンプレックス サンプルの分散の推定値を適切に算出するために必要な情報が含まれます。この計画には、サンプルの構造、各段階での推定方法、および必要な変数への参照（サンプルの重みなど）が含まれます。分析準備ウィザードで、分析計画の作成と編集を行えます。

計画ファイルに指定内容を保存することで、次のような利点が得られます。

- 調査者は複数段階抽出計画の第 1 段階を指定し、第 1 段階を抽出した後、抽出単位の情報を第 2 段階用に収集できます。その後、第 2 段階を含めるように、抽出計画を修正できます。
- 抽出計画ファイルにアクセスできない分析者は、分析計画を指定して、各コンプレックス サンプル分析手続きからその計画を参照できます。
- 大規模な一般使用サンプルを計画する設計者は、抽出計画ファイルを公開できます。これにより分析者への指示を簡略化でき、各分析者は自分で分析計画を指定する必要がなくなります。

## 参考文献

抽出手法の詳細については、次のテキストを参照してください。

Cochran, W. G. 1977. *Sampling Techniques*, 3rd ed. New York: John Wiley and Sons.

Kish, L. 1965. *Survey Sampling*. New York: John Wiley and Sons.

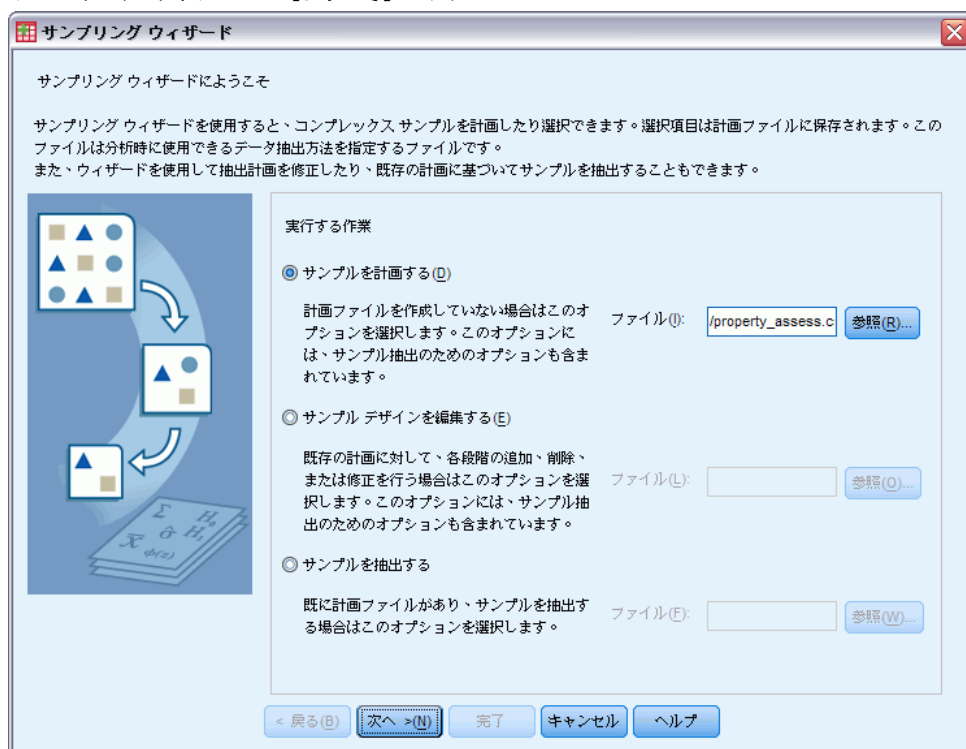
Kish, L. 1987. *Statistical Design for Research*. New York: John Wiley and Sons.

Murthy, M. N. 1967. *Sampling Theory and Methods*. Calcutta, India: Statistical Publishing Society.

Särndal, C., B. Swensson, および J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

# コンプレックス デザインからの抽出

図 2-1  
サンプリング ウィザードの [ようこそ] ステップ



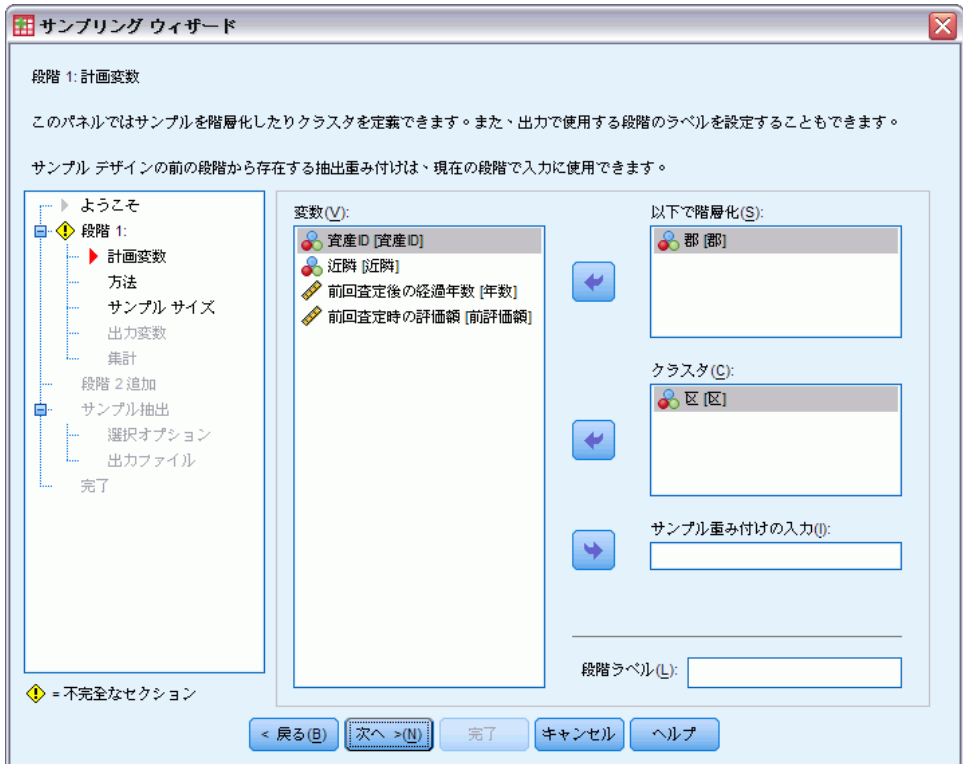
サンプリング ウィザードでは、抽出計画ファイルの作成や修正、実行のステップを、順を追って行います。ウィザードを使用する前に、対象母集団を適切に定義しておき、抽出単位の一覧を準備し、適切なサンプル計画を検討しておく必要があります。

## 新規サンプル計画の作成

- ▶ メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > サンプルの選択...
- ▶ [サンプルを計画する] を選択し、計画ファイル名を選択してサンプル計画を保存します。
- ▶ [次へ] をクリックしてウィザードを進めます。
- ▶ [計画変数] ステップで、ストラータ、クラスタ、入力サンプルの重み付けを定義することもできます。定義が終わったら、[次へ] をクリックします。
- ▶ [抽出方法] ステップで、選択した項目の抽出方法を選択することもできます。  
  
[Brewer (PPS)] または [Murthy (PPS)] を選択した場合、[完了] をクリックしてサンプルを抽出できます。それ以外の場合、[次へ] をクリックして次のステップに進みます。
- ▶ [サンプル サイズ] ステップで、抽出する単位 (ユニット) の数または比率を指定します。
- ▶ ここで [完了] をクリックして、サンプルを抽出できます。  
さらに次のステップ を実行することもできます。
  - 保存する出力変数を選択する。
  - 計画に第 2 または第 3 段階を追加する。
  - どの段階でサンプルを抽出するか、乱数シード、ユーザー指定の欠損値を計画変数の有効な値として扱うかどうかなど、さまざまな選択オプションを設定する。
  - 出力データの保存先を選択する。
  - 選択をコマンド シンタックスとして貼り付ける。

## サンプリング ウィザード: 計画変数

図 2-2  
サンプリング ウィザードの [計画変数] ステップ



このステップでは、層化とクラスタリングの変数の選択と、入力サンプルの重み付けの定義を行えます。また、段階にラベルを指定できます。

**以下で階層化。**層化（ストラータ）変数のクロス分類により、独立した部分母集団、つまりストラータが定義されます。独立したサンプルが各ストラータについて取得されます。推定値の精度を向上するには、ストラータ内の単位を対象特性に合わせて可能な限り均一にする必要があります。

**クラスタ。**クラスタ変数により、観測単位のグループ、つまりクラスタが定義されます。母集団から観測単位を直接抽出することが不可能か、費用がかかりすぎる場合に、クラスタは有効です。または、母集団からクラスタを抽出した後で、選択したクラスタから観測単位を抽出することもできます。しかし、クラスタを使用すると抽出単位間に相関が発生し、精度が下がる場合があります。この影響を最小限にするために、クラスタ内の単位を対象特性に合わせて可能な限り均一にする必要があります。複数の段階を持つデザインを計画するには、1つ以上のクラスタ変数を定義する必要があります。また、異なる複数の抽出方法を使用する

場合にも、クラスタは必要です。詳細は、p.8 [サンプリング ウィザード: 抽出方法](#) を参照してください。

**サンプル重み付けの入力。**現在のサンプル デザインが、より大きなサンプル デザインの一部である場合、大きな計画の前の段階でサンプルの重み付けを使った可能性があります。現在の計画の第1段階で、このような重み付けを含む数値変数を指定することができます。現在の計画の以降の段階では、サンプルの重み付けは自動的に算出されます。

**段階ラベル。**各段階に、オプションとして文字列ラベルを指定できます。このラベルは出力で使用され、段階に関する情報を識別するのに役立ちます。

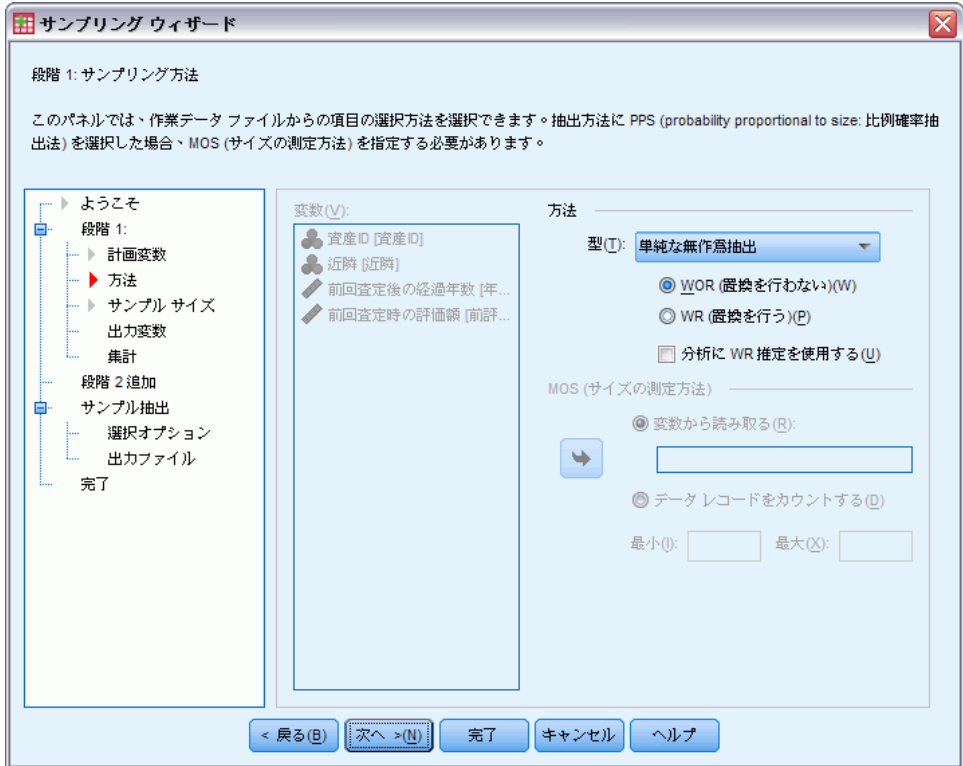
**注:** ソース変数リストは、ウィザードのすべてのステップで同じ内容が含まれます。つまり、特定のステップでソース リストから変数を削除すると、すべてのステップでその変数が削除されます。変数をソース リストに戻すと、すべてのステップでその変数が表示されます。

## サンプリング ウィザードのツリー表示の操作方法

サンプリング ウィザードの各ステップの左側には、すべてのステップのアウトラインが表示されます。アウトライン内で有効になっているステップの名前をクリックすることで、そのステップを表示することができます。ステップは、その前にあるすべてのステップが有効、つまり前のステップで必要最低限の指定が行われていれば、有効です。ステップが無効になる原因の詳細については、各ステップのヘルプを参照してください。

## サンプリング ウィザード: 抽出方法

図 2-3  
サンプリング ウィザードの [サンプリング方法] ステップ



このステップでは、アクティブなデータセットからケースを選択する方法を指定できます。

**方法。**このグループのコントロールは、選択方法を選択するために使用されます。抽出のタイプによっては、置換を行う (WR) か、置換を行わない (WOR) かを選択できます。詳細については、各タイプの説明を参照してください。いくつかの確率比例 (PPS) タイプは、クラスタが定義されている場合にのみ利用することに注意してください。また、どの PPS タイプも計画の第 1 段階でのみ使用できます。さらに、WR 方法は計画の最後の段階でのみ使用できます。

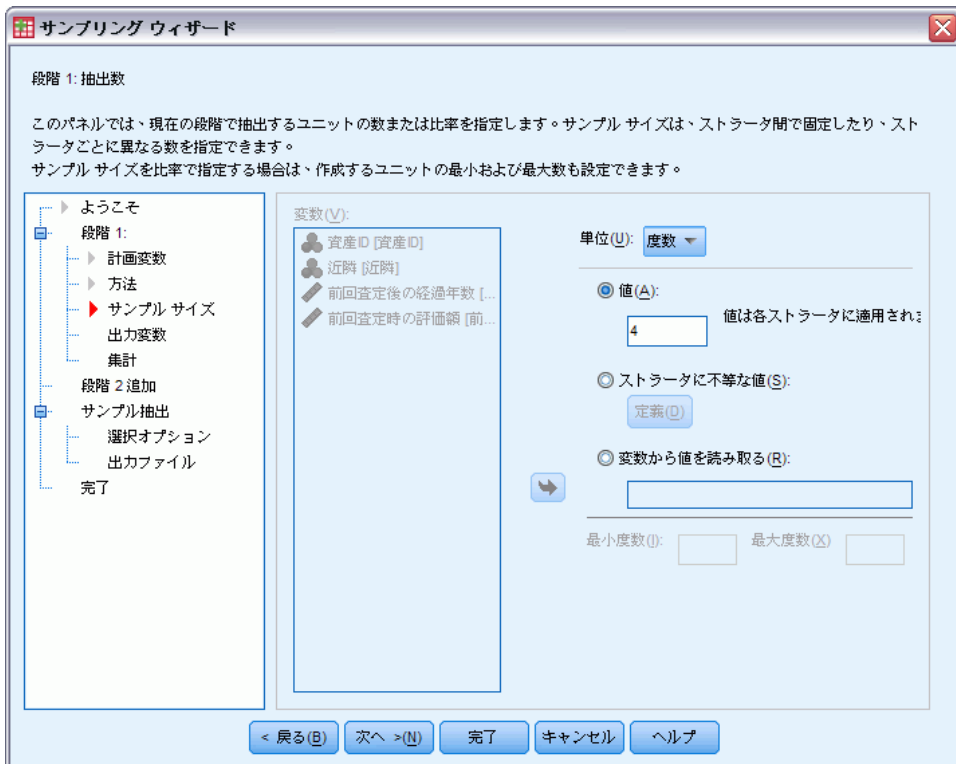
- **単純な無作為抽出。**等しい確率で単位が選択されます。置換を使用しても、使用しなくても選択されます。
- **単純な系統的抽出。**抽出枠 (ストラータが指定されている場合、ストラータ) 全体で単位が固定間隔で選択され、置換なしに抽出されます。最初の間隔内で無作為に選択された単位が開始点になります。
- **単純な連続値。**単位は等しい確率で順番に選択され、置換は行われません。

- **PPS**。サイズに比例した確率で無作為に単位を選択する、第 1 ステージの方法。置換を行って任意の単位を選択できます。置換を行わない場合、クラスタのみ抽出できます。
- **系統的 (PPS)**。サイズに比例した確率で系統的に単位を選択する、第 1 ステージの方法。置換なしに選択されます。
- **連続的 (PPS)**。置換なしに、クラスタ サイズに比例した確率で順番に単位を選択する、第 1 ステージの方法。
- **Brewer (PPS)**。置換なしに、クラスタ サイズに比例した確率で各ストラータから 2 つのクラスタを選択する、第 1 ステージの方法。この方法を使用するには、クラスタ変数を指定する必要があります。
- **Murthy (PPS)**。置換なしに、クラスタ サイズに比例した確率で各ストラータから 2 つのクラスタを選択する、第 1 ステージの方法。この方法を使用するには、クラスタ変数を指定する必要があります。
- **Sampford (PPS)**。置換なしに、クラスタ サイズに比例した確率で各ストラータから 3 つ以上のクラスタを選択する、第 1 ステージの方法。これは Brewer 法を拡張したものです。この方法を使用するには、クラスタ変数を指定する必要があります。
- **分析に WR 推定を使用する**。デフォルトで、推定方法は選択した抽出方法が一貫する計画ファイル内で指定されます。これにより、抽出方法に WOR 推定が含まれている場合でも、置換を行う推定を使用できます。このオプションは第 1 段階でのみ使用できます。

**MOS (サイズの測定方法)**。PPS 法を選択した場合、各単位のサイズを定義するサイズの測定方法を指定する必要があります。これらのサイズは変数で明示的に定義するか、データから計算できます。オプションとして、MOS に上限と下限を設定し、MOS 変数のすべての値またはデータから計算される値を上書きできます。これらのオプションは第 1 段階でのみ使用できます。

## サンプリング ウィザード: サンプル サイズ

図 2-4  
サンプリング ウィザードの [サンプル サイズ] ステップ



このステップでは、現在の段階内で抽出する単位数または単位の比率を指定できます。抽出サイズは固定することも、ストラータごとに変えることもできます。サンプル サイズを指定する際には、前の段階で選択されたクラスタをストラータの定義に使用できます。

**単位。** 抽出する単位の正確なサンプル サイズまたは比率を指定できます。

- **値。** 1つの値がすべてのストラータに適用されます。単位の測定基準として [度数] が選択されている場合、正の整数値を入力する必要があります。[比率] が選択されている場合、負でない値を入力する必要があります。また、置換を行う抽出でない場合、比率の値は 1 より大きな値にしないでください。
- **ストラータに不等な値。** [等しくないサイズの定義] ダイアログ ボックスを使って、ストラータごとにサイズの値を入力できます。
- **変数から値を読み取る。** ストラータのサイズ値を含む数値型変数を選択できます。



[比率] が選択されている場合、抽出される単位数に下限と上限を設定することもできます。

## 等しくないサイズの定義

図 2-5  
[等しくないサイズの定義] ダイアログ ボックス



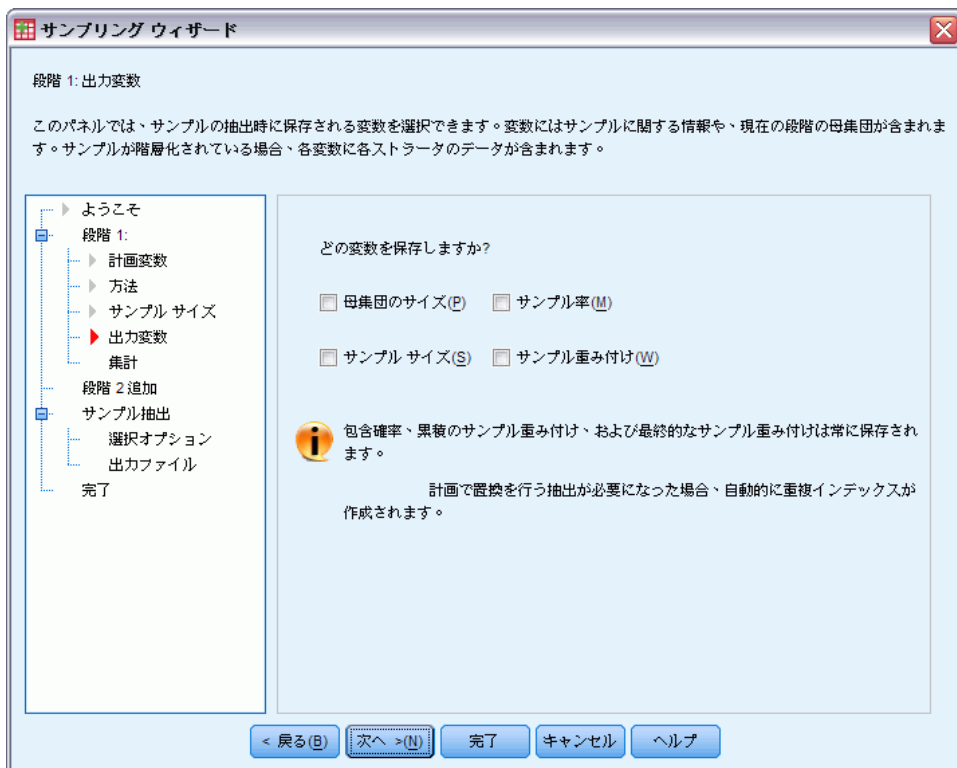
[等しくないサイズの定義] ダイアログ ボックスでは、サイズをストラータごとに入力できます。

**[サイズ指定] グリッド。**このグリッドには、最大 5 個のストラータまたはクラスター変数のクロス分類が表示されます。1 行に 1 つのストラータとクラスターの組み合わせが表示されます。適切なグリッド変数は、現在および以前の段階のすべての層化 (ストラータ) 変数と、以前の段階のすべてのクラスター変数です。グリッド内部で変数を並べ替えたり、[除外] リストに移動したりできます。右端の列にサイズを入力します。[ラベル] または [値] をクリックすると、グリッドのセル内の表示が値ラベルまたは層化変数およびクラスター変数のデータ値に切り替わります。ラベルのない値を含むセルには常に値が表示されます。[ストラータを再表示] をクリックすると、グリッド内の変数に対するラベル付きデータ値の組み合わせがグリッドに再入力されます。

**除外。**ストラータとクラスターの組み合わせのサブセットにサイズを指定するには、1 つ以上の変数を [除外] リストに移動します。移動した変数はサンプルサイズの定義に使用されません。

## サンプリング ウィザード: 出力変数

図 2-6  
サンプリング ウィザードの [出力変数] ステップ



このステップでは、サンプルを抽出するときに保存する変数を選択できます。

**母集団のサイズ。**その段階における母集団の推定単位数。保存される変数のルート名は、PopulationSize\_ になります。

**サンプル率。**その段階における抽出比率。保存される変数のルート名は、SamplingRate\_ になります。

**サンプル サイズ。**その段階で抽出される単位数。保存される変数のルート名は、SampleSize\_ になります。

**サンプル重み付け。**包含確率の逆数です。保存される変数のルート名は、SampleWeight\_ になります。

段階に関する変数には、自動的に生成されるものもあります。次にその内容を示します。

**包含確率。**その段階で抽出される単位の比率。保存される変数のルート名は、InclusionProbability\_ になります。

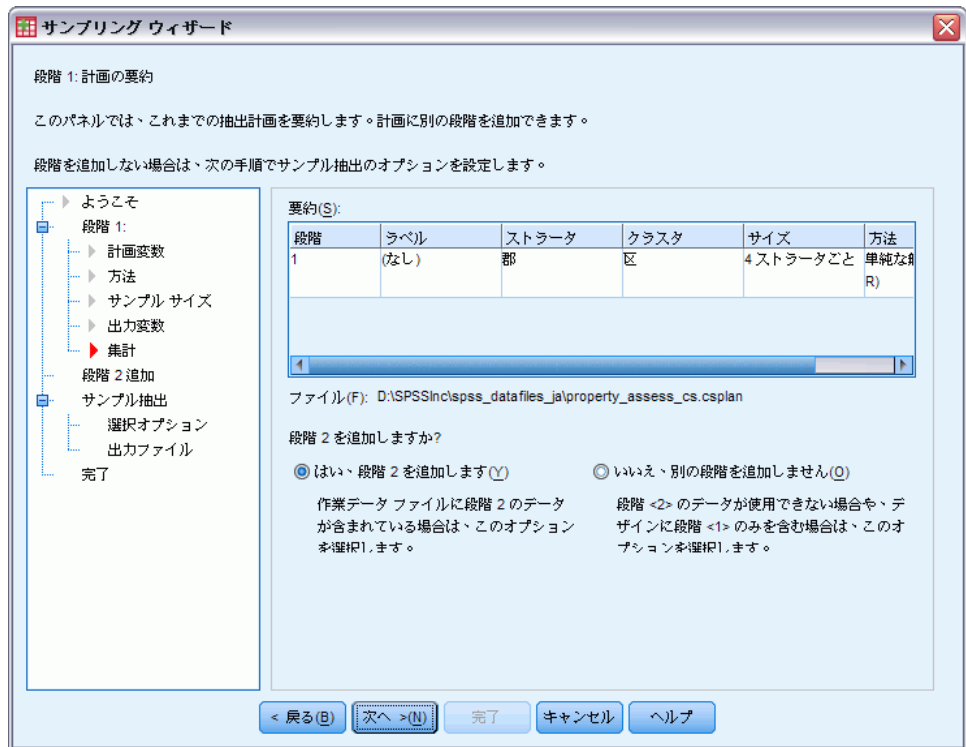
**累積重み付け。**現在の段階および以前の段階における、累積サンプル重み付け。保存される変数のルート名は、SampleWeightCumulative\_ になります。

**インデックス。**その段階で複数回選択される単位を識別します。保存される変数のルート名は、Index\_ になります。

**注:** 保存される変数のルート名には、段階の番号が含まれます。たとえば第 1 段階で保存される母集団サイズの場合、PopulationSize\_1\_ になります。

## サンプリング ウィザード: 計画の要約

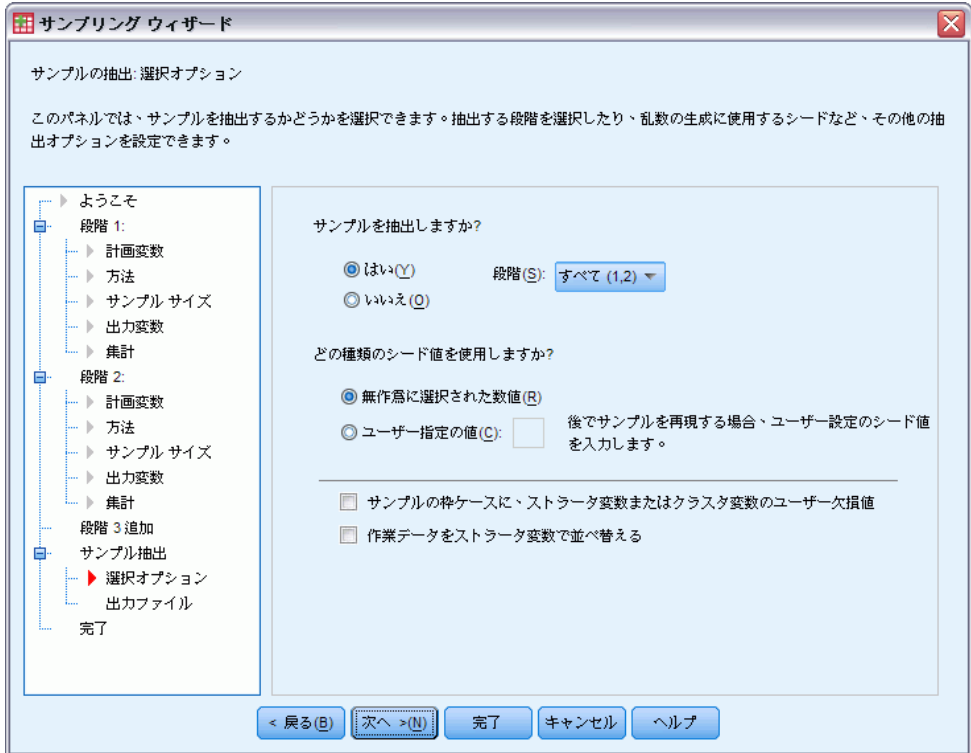
図 2-7  
サンプリング ウィザードの [計画の要約] ステップ



これは各段階における最後のステップです。現在の段階におけるサンプルデザインの指定の要約が得られます。ここから、次の段階（必要に応じて作成）に進むか、サンプル抽出のオプションを設定します。

## サンプリング ウィザード: サンプルの抽出: 選択オプション

図 2-8  
サンプリング ウィザードの [サンプルの抽出: 選択オプション] ステップ



このステップでは、サンプルを抽出するかどうかを選択できます。乱数のシードや欠損値の処理など、他の抽出オプションも設定できます。

**サンプル抽出。**サンプルを抽出するかどうかを選択できるだけでなく、抽出計画を実行する段階も選択できます。段階では順番に抽出する必要があります。つまり、第2段階で抽出するには第1段階でも抽出する必要があります。計画を編集または実行するときには、ロックされた段階で再抽出することはできません。

**シード。**ここでは、乱数生成のシード値を選択できます。

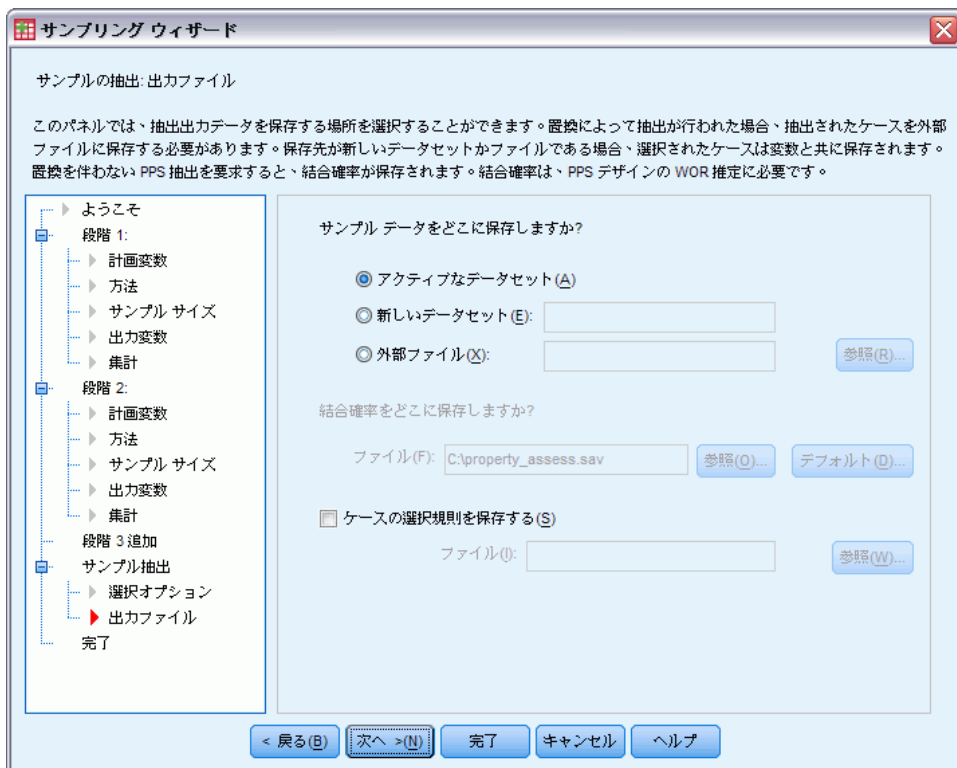
**ユーザー欠損値を含める。**ここで、ユーザー欠損値が有効かどうかを指定します。有効にした場合、ユーザー欠損値は別のカテゴリとして処理されます。

**データは並べ替え済み。**サンプル枠が層化変数の値によってあらかじめソートされている場合、このオプションを指定することで選択プロセスを高速化することができます。

## サンプリング ウィザード: サンプルの抽出: 出力ファイル

図 2-9

サンプリング ウィザードの [サンプルの抽出: 出力ファイル] ステップ



このステップでは、抽出したケース、重み付け変数、結合確率、ケース選択規則の出力先を選択できます。

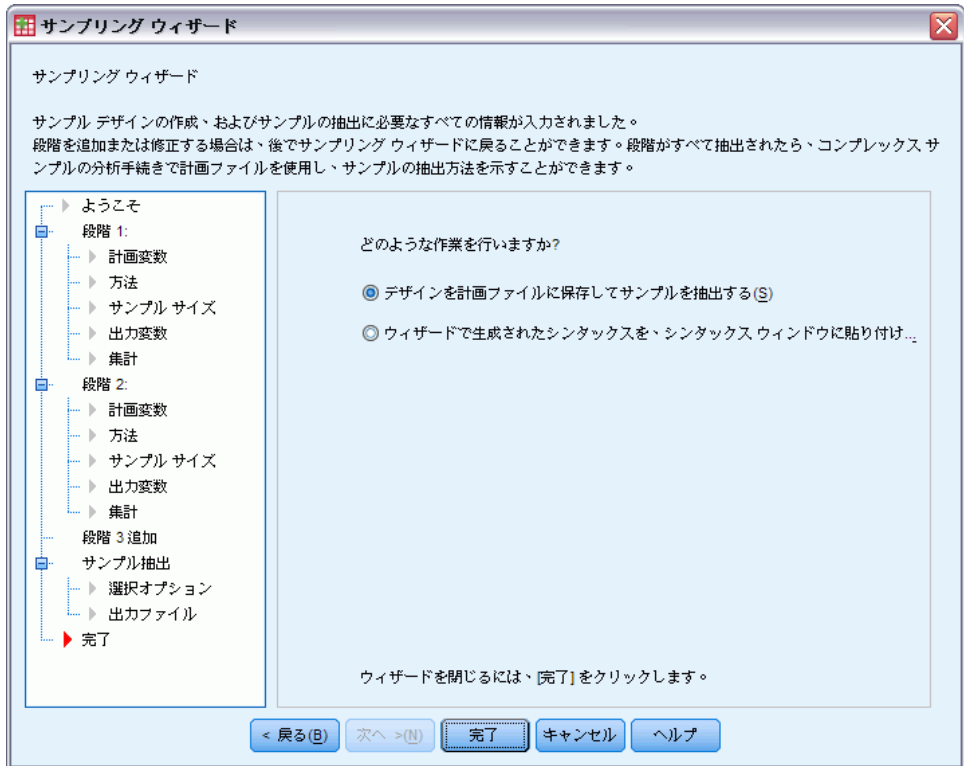
**サンプル データ。**ここにあるオプションでは、サンプル出力の書き込み先を指定できます。アクティブなデータセットに追加するか、新しいデータセットに書き込むか、IBM® SPSS® Statistics データ ファイルに保存します。データセットが利用可能なのは現在のセッション中ですが、今後のセッションで利用するには、データセットをデータ ファイルとして明示的に保存する必要があります。データセット名は、変数命名規則に従う必要があります。外部ファイルまたは新しいデータセットを指定すると、選択したケースに対する抽出出力変数と作業データ ファイルの変数が書き込まれます。

**結合確率。**ここにあるオプションでは、結合確率の書き込み先を指定できます。これらは外部 SPSS Statistics データ ファイルに保存され、[WOR (PPS)]、[Brewer (PPS)]、[Sampford (PPS)]、[Murthy (PPS)] のいずれかを選択し、WR 推定を指定しない場合、結合可能性が生成されます。

**ケースの選択規則。** サンプルを1つの段階で一度に作成する場合、ケース選択規則をテキストファイルに保存してもよいでしょう。これは後の段階でサブフレームを作成するときに役立ちます。

## サンプリング ウィザード: 終了

図 2-10  
サンプリング ウィザードの [完了] ステップ



これは最後のステップです。計画ファイルを保存してすぐにサンプルを抽出するか、選択したものをシンタックス ウィンドウに貼り付けます。

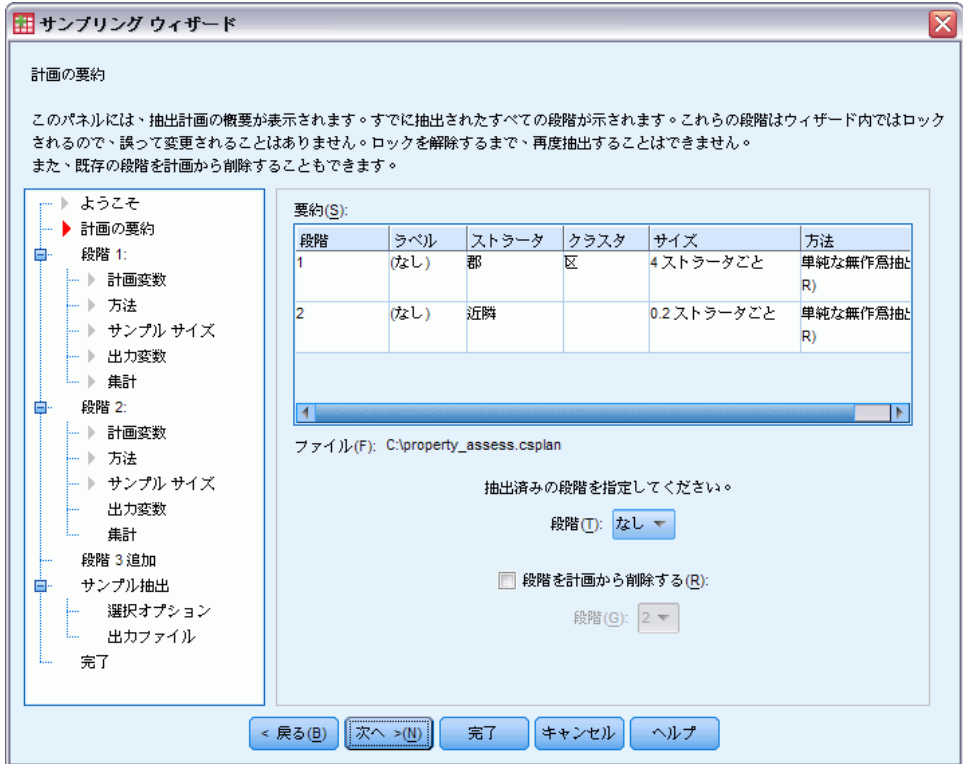
既存の計画ファイルの段階に変更を加えた場合、編集した計画を新しいファイルに保存するか、既存のファイルに上書きできます。既存の段階に変更を加えずに段階を追加した場合は、既存の計画ファイルが自動的に上書きされます。計画を新しいファイルに保存したい場合は、[ウィザードで生成されたシンタックスを、シンタックス ウィンドウに貼り付ける]を選択し、シンタックス コマンドでファイル名を変更します。

## 既存のサンプル計画の修正

- ▶ メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > サンプルの選択...
- ▶ [サンプル デザインを編集する] を選択し、編集する計画ファイルを選択します。
- ▶ [次へ] をクリックしてウィザードを進めます。
- ▶ [計画の要約] ステップで抽出計画を確認した後、[次へ] をクリックします。  
以降のステップは、新規計画の場合とほとんど同じです。詳細については、各ステップのヘルプを参照してください。
- ▶ [完了] ステップを開き、編集した計画ファイルに新しい名前を指定するか、既存の計画ファイルを上書きします。  
オプションとして、次の選択が可能です。
  - すでに抽出を行った段階を指定する。
  - 段階を計画から削除する。

## サンプリング ウィザード: 計画の要約

図 2-11  
サンプリング ウィザードの [計画の要約] ステップ



このステップでは、抽出計画を確認し、すでに抽出が行われた段階を指定できます。計画を編集している場合、計画から段階を削除することもできます。

**以前に抽出された段階。** 拡張抽出枠が使用できない場合、複数の段階を持つ抽出デザインでは一度に 1 段階ずつ実行する必要があります。すでに抽出が行われた段階を、ドロップダウン リストから選択します。実行済みの段階はすべてロックされ、[サンプルの抽出: 選択オプション] ステップでも利用できません。また、計画の編集時にも変更できません。

**段階の削除。** 複数の段階を持つデザインから、段階 2 と 3 を削除できます。

## 既存のサンプル計画の実行

- ▶ メニューから次の項目を選択します。  
分析 > コМПЛЕКС サンプル > サンプルの選択...
- ▶ [サンプルを抽出する] を選択し、実行する計画ファイルを選択します。



- ▶ [次へ] をクリックしてウィザードを進めます。
  - ▶ [計画の要約] ステップで抽出計画を確認した後、[次へ] をクリックします。
  - ▶ 段階に関する情報を含む個別のステップは、サンプル計画を実行するときにはスキップされます。いつでも [完了] ステップに進めます。
- オプションとして、すでに抽出を行った段階を指定することもできます。

## CSPLAN および CSSELECT 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。

- 出力変数にユーザー定義の名前を指定する。
- ビューアの出力を制御する。たとえば、サンプルの計画時または修正時に表示される計画の段階ごとの要約や、サンプル デザインの実行時に表示されるストラータごとの抽出ケースの区分の要約を表示しないようにしたり、ケース処理の要約を要求したりできます。
- アクティブなデータセットにある変数のサブセットを選択し、外部サンプル ファイルや別のデータセットに書き込む。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

# コンプレックス サンプルの分析準備

図 3-1  
分析準備ウィザードの [よろこ] ステップ



分析準備ウィザードでは、さまざまなコンプレックス サンプル分析手続きで使用される分析計画の作成と修正のステップを、順を追って行います。ウィザードを使用する前に、複合計画に従ってサンプルを抽出しておきます。

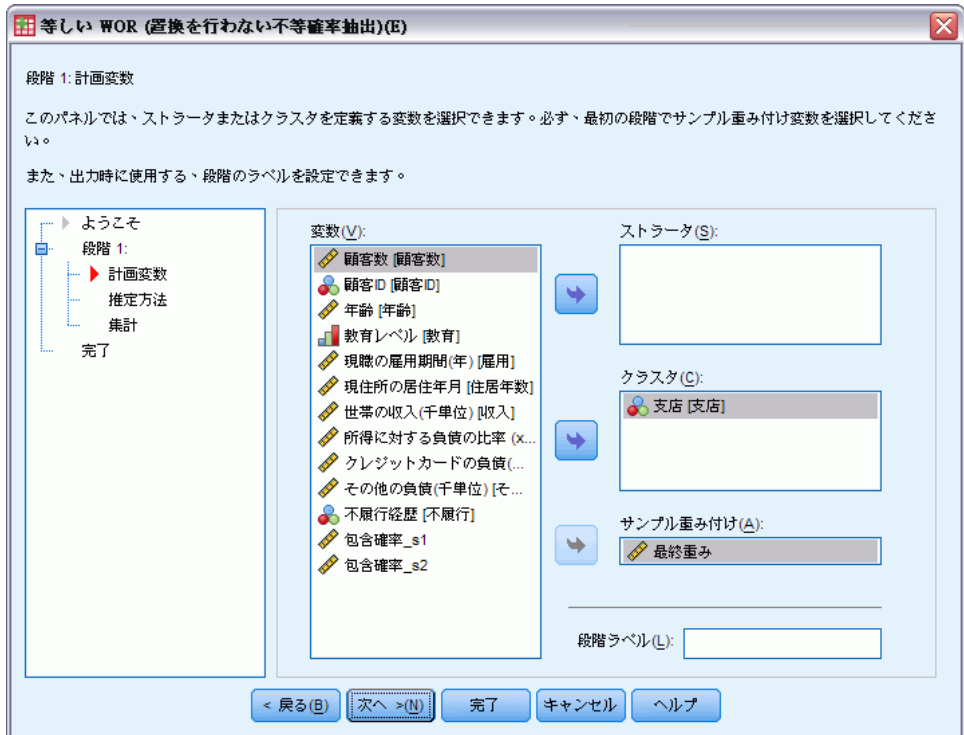
サンプルの抽出に使用する抽出計画ファイルにアクセスできない場合、新しい計画を作成することを推奨します。抽出計画にはデフォルト分析計画が含まれます。サンプルの抽出に使用する抽出計画ファイルにアクセスできる場合、抽出計画ファイルに含まれるデフォルトの分析計画を使用するか、デフォルトの分析指定を上書きして、変更を新しいファイルに保存できます。

## 新規分析計画の作成

- ▶ メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > 分析の準備...
- ▶ [計画ファイルの作成] を選択し、分析計画を保存する計画ファイル名を選択します。
- ▶ [次へ] をクリックしてウィザードを進めます。
- ▶ [計画変数] ステップで、サンプル重み付けが含まれる変数を指定します。  
オプションでストラータとクラスタを定義できます。
- ▶ [完了] をクリックして計画を保存します。  
さらに次のステップ を実行することもできます。
  - [推定方法] ステップで、標準誤差の推定方法を選択する。
  - [サイズ] ステップで、抽出される単位数または単位ごとの包含確率を指定する。
  - 計画に第 2 または第 3 段階を追加する。
  - 選択をコマンド シンタックスとして貼り付ける。

## 分析準備ウィザード: 計画変数

図 3-2  
分析準備ウィザードの [計画変数] ステップ



このステップでは、層化とクラスタリングの変数の指定と、サンプルの重み付けの定義を行えます。また、段階にラベルを設定できます。

**ストラータ。**層化（ストラータ）変数のクロス分類により、独立した部分母集団、つまりストラータが定義されます。すべてのサンプルが、各ストラータからの独立したサンプルの組み合わせを示します。

**クラスタ。**クラスタ変数により、観測単位のグループ、つまりクラスタが定義されます。複数の段階で抽出されたサンプルによって、初期の段階でクラスタが選択された後、選択されたクラスタからサブサンプル単位が選択されます。置換を使用してクラスタを抽出することで得られたデータ ファイルを分析するときには、重複インデックスをクラスタ変数として使用する必要があります。

**サンプル重み付け。**第 1 段階でサンプルの重み付けを指定する必要があります。現在の計画の以降の段階では、サンプルの重み付けは自動的に算出されます。

**段階ラベル。**各段階に、オプションとして文字列ラベルを指定できます。このラベルは出力で使用され、段階に関する情報を識別するのに役立ちます。

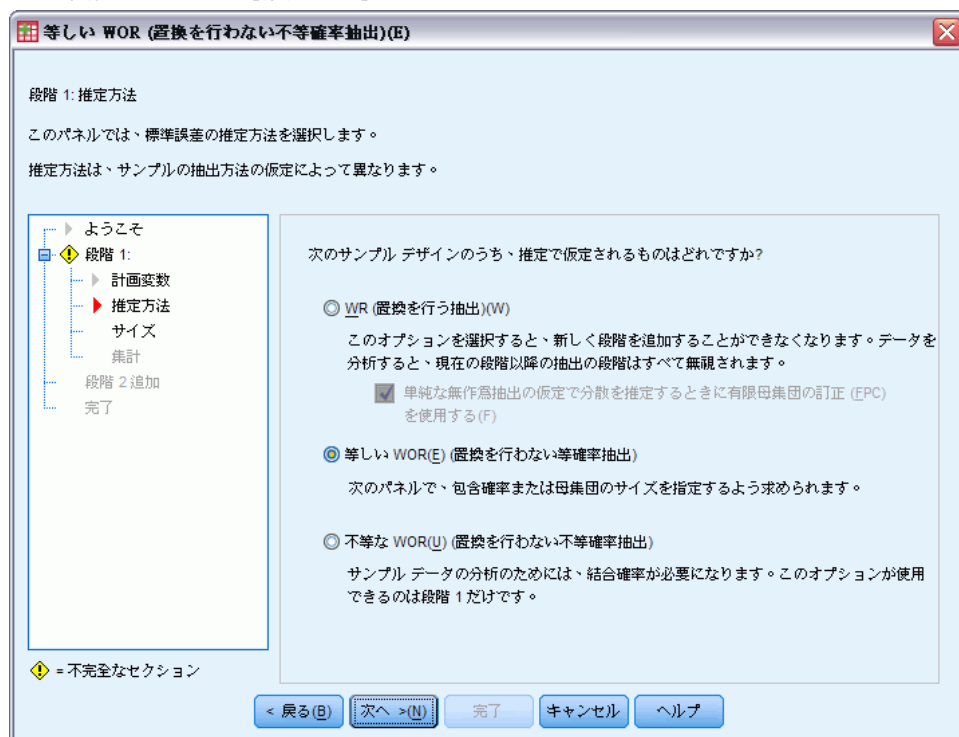
**注:** ソース変数リストは、ウィザードのすべてのステップで同じ内容が含まれます。つまり、特定のステップでソース リストから変数を削除すると、すべてのステップでその変数が削除されます。変数をソース リストに戻すと、すべてのステップでその変数が表示されます。

## 分析ウィザードのツリー表示の操作方法

分析ウィザードの各ステップの左側には、全ステップのアウトラインが表示されます。アウトライン内で有効になっているステップの名前をクリックすることで、そのステップを表示することができます。ステップは、その前にあるすべてのステップが有効、つまり前にあるステップで必要最低限の指定が行われていれば、有効になります。ステップが無効になる原因の詳細については、各ステップのヘルプを参照してください。

## 分析準備ウィザード: 推定方法

図 3-3  
分析準備ウィザードの [推定方法] ステップ



このステップでは、ステージの推定方法を指定できます。

**WR (置換を行う抽出)**。コンプレックス サンプル計画での分散の推定を行うとき、WR 推定には有限母集団からの抽出の修正 (FPC) が含まれません。単純な無作為抽出 (SRS) での分散の推定を行うときは、FPC を含めるか除外するかを選択できます。

分析の重み付けが行われていて、全ケースを合計しても母集団のサイズに達しない場合は、SRS の分散推定に FPC を含めないようにすることをお勧めします。SRS の分散推定値は、デザイン効果などの統計量の計算に使用されます。WR 推定は計画の最後の段階にのみ指定できます。ウィザードで WR 推定を選択した場合、段階を追加することはできません。

**Equal WOR (置換を行わない等確率抽出)**。等しい WOR 推定には有限母集団の修正が含まれ、単位が等しい確率で抽出されると仮定されます。等しい WOR は、計画の任意の段階に指定できます。

**不等な WOR (置換を行わない不等確率抽出)**。不等な WOR では有限母集団の訂正を使用するだけでなく、不等確率で選択された単位 (通常はクラスタ) の抽出も行われます。この推定方法は第 1 段階でのみ使用できます。

## 分析準備ウィザード: サイズ

図 3-4  
分析準備ウィザード [サイズ] ステップ

The screenshot shows a software window titled "等しい WOR (置換を行わない不等確率抽出)(E)". The main area is labeled "段階 1: Size" and contains the following text: "このパネルでは、現在の段階の包含確率または母集団のサイズを指定します。" and "サイズは、ストラタ間で固定したり、ストラタごとに異なるサイズを指定できます。"

On the left, a navigation tree shows the following steps: ようこそ, 段階 1: (selected), 計画変数, 推定方法, サイズ (selected), 集計, 段階 2 追加, 完了.

The main content area is divided into three sections:

- 変数(V):** A list of variables including: 顧客数 [顧客数], 顧客ID [顧客ID], 年齢 [年齢], 教育レベル [教育], 現職の雇用期間(年) [雇用], 現住所の居住年月 [住居年数], 世帯の収入(千単位) [収入], 所得に対する負債の比率 (x1...), クレジットカードの負債(千...), その他の負債(千単位) [その...], 不履行経歴 [不履行], and 包含確率\_s2.
- ユニット(U):** A dropdown menu set to "包含確率".
- 値(A):** An empty text input field.
- ストラタに不等な値(S):** A radio button option with a "定義(D)..." button next to it.
- 変数から値を読み取る(R):** A radio button option with a dropdown menu showing "包含確率\_s1".

At the bottom, there are navigation buttons: "< 戻る(B)", "次へ >(N)", "完了", "キャンセル", and "ヘルプ".

このステップは、現在の段階の包含確率や母集団のサイズを指定するために使用されます。サイズは固定することも、ストラータごとに変えることもできます。サイズを指定するために、前の段階で指定されたクラスタを使用してストラータを定義できます。このステップは、等しい WOR が推定方法として選択されているときだけ必要です。

**単位。**母集団の正確なサイズ、または単位の抽出で使用される確率を指定できます。

- **値。**1 つの値がすべてのストラータに適用されます。単位の測定基準として [母集団のサイズ] が選択されている場合、負でない整数値を入力する必要があります。[包含確率] が選択されている場合、0 以上 1 以下の値を入力する必要があります。
- **ストラータに不等な値。**[等しくないサイズの定義] ダイアログ ボックスを使って、ストラータごとにサイズの値を入力できます。
- **変数から値を読み取る。**ストラータのサイズ値を含む数値型変数を選択できます。

## 等しくないサイズの定義

図 3-5  
[等しくないサイズの定義] ダイアログ ボックス



[等しくないサイズの定義] ダイアログ ボックスでは、サイズをストラータごとに入力できます。

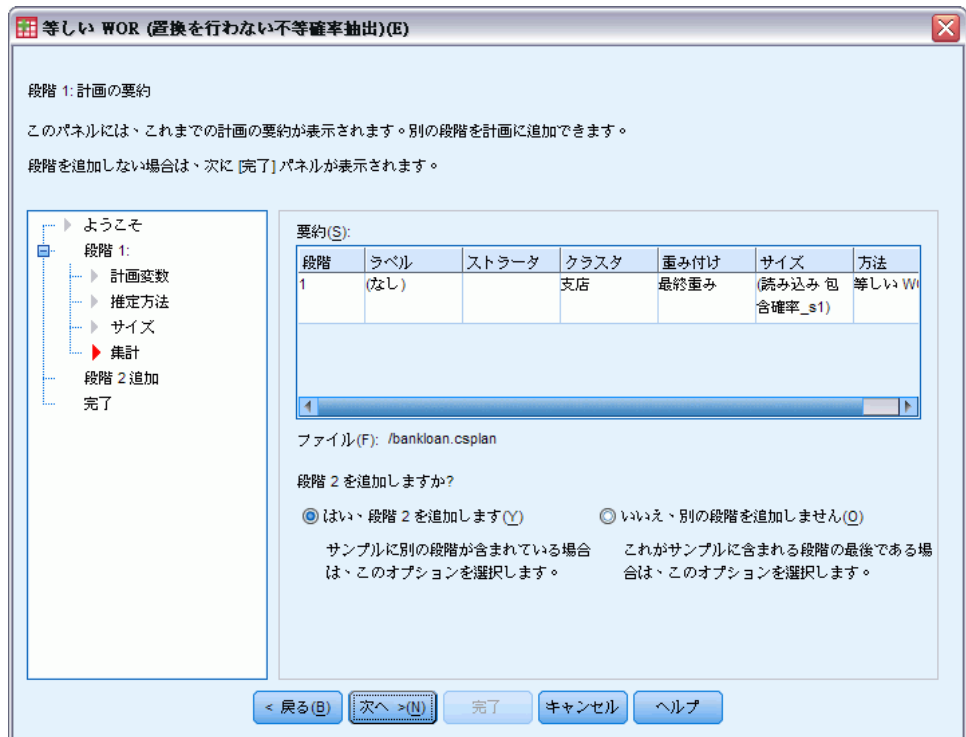
**[サイズ指定] グリッド。**このグリッドには、最大 5 個のストラータまたはクラスタ変数のクロス分類が表示されます。1 行に 1 つのストラータとクラスタの組み合わせが表示されます。適切なグリッド変数は、現在および以前の段階のすべての層化 (ストラータ) 変数と、以前の段階のすべてのクラスタ変数です。グリッド内部で変数を並べ替えたり、[除外] リストに移動したりできます。右端の列にサイズを入力します。[ラベル] または [値] をクリック

すると、グリッドのセル内の表示が値ラベルまたは層化変数およびクラス変数のデータ値に切り替わります。ラベルのない値を含むセルには常に値が表示されます。[ストラータを再表示] をクリックすると、グリッド内の変数に対するラベル付きデータ値の組み合わせがグリッドに再入力されます。

**除外。** ストラータとクラスタの組み合わせのサブセットにサイズを指定するには、1 つ以上の変数を [除外] リストに移動します。移動した変数はサンプルサイズの定義に使用されません。

## 分析準備ウィザード: 計画の要約

図 3-6  
分析準備ウィザードの [計画の要約] ステップ



これは各段階における最後のステップです。現在の段階全体における分析計画の指定の要約が得られます。ここから、次の段階（必要に応じて作成）に進むか、分析指定を保存します。

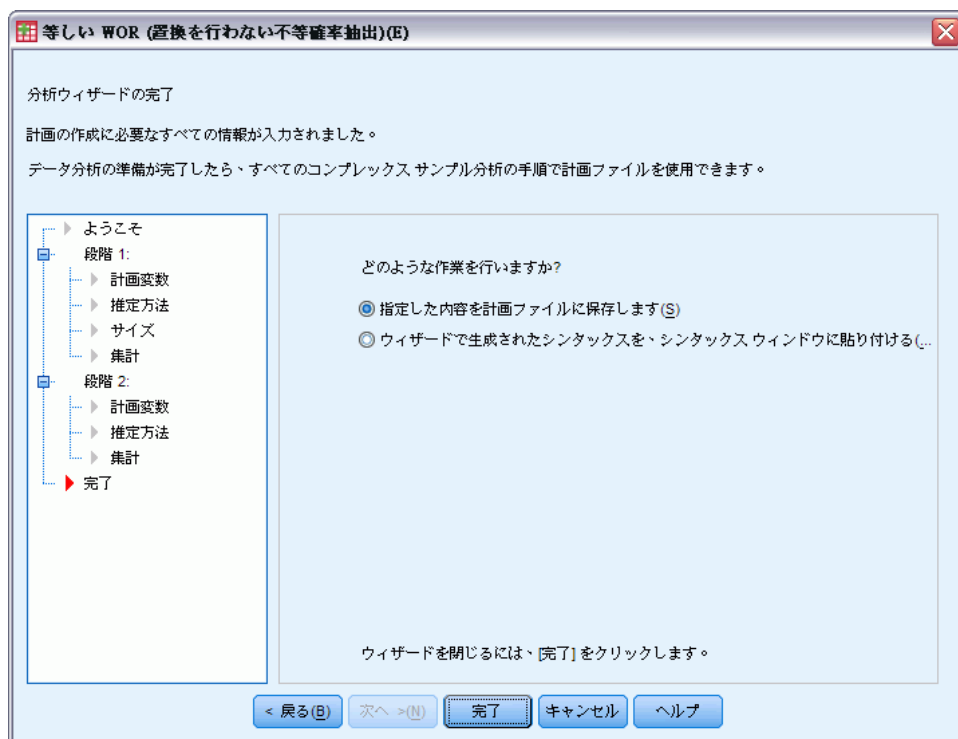
段階を追加できない場合、次の原因が考えられます。

- [計画変数] ステップでクラスタ変数が指定されていない。
- [推定方法] ステップで WR 推定を選択した。
- これが分析の第 3 段階である。ウィザードでは最大 3 つの段階がサポートされます。



## 分析準備ウィザード: 終了

図 3-7  
分析準備ウィザードの [完了] ステップ



これは最後のステップです。計画ファイルを保存するか、選択したものをシンタックス ウィンドウに貼り付けます。

既存の計画ファイルの段階に変更を加えた場合、編集した計画を新しいファイルに保存するか、既存のファイルに上書きできます。既存の段階に変更を加えずに段階を追加した場合は、既存の計画ファイルが自動的に上書きされます。計画を新しいファイルに保存したい場合は、[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] を選択し、シンタックス コマンドでファイル名を変更します。

## 既存の分析計画の修正

- ▶ メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > 分析の準備...
- ▶ [計画ファイルの編集] を選択し、分析計画を保存する計画ファイル名を選択します。
- ▶ [次へ] をクリックしてウィザードを進めます。

- ▶ [計画の要約] ステップで分析計画を確認した後、[次へ] をクリックします。以降のステップは、新規計画の場合とほとんど同じです。詳細については、各ステップのヘルプを参照してください。
- ▶ [終了] ステップを開き、編集した計画ファイルに新しい名前を指定するか、既存の計画ファイルを上書きします。  
オプションとして既存の段階を計画から削除することもできます。

## 分析準備ウィザード: 計画の要約

図 3-8  
分析準備ウィザードの [計画の要約] ステップ



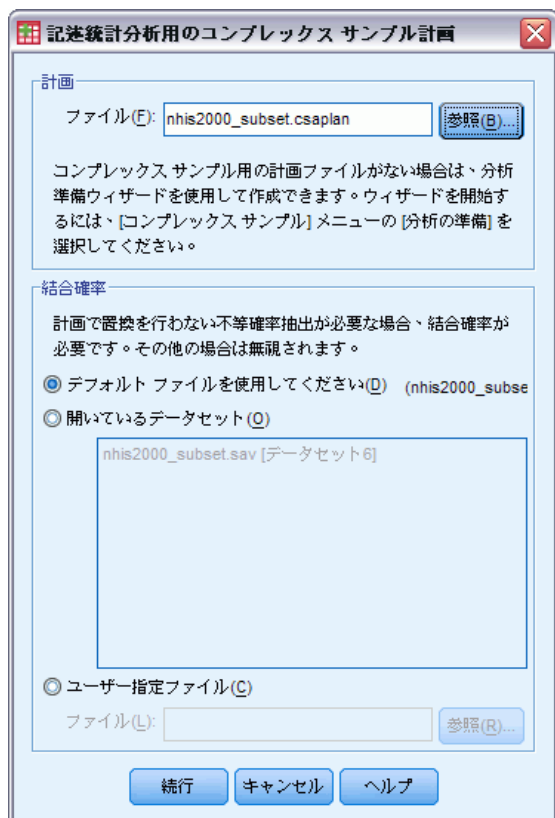
このステップでは、分析計画の確認と、計画からの段階の削除を行えます。

**段階の削除。**複数の段階を持つデザインから、段階 2 と 3 を削除できます。計画には少なくとも 1 つの段階 (ステージ) が必要なため、計画の第 1 段階については編集は可能ですが、削除はできません。

# コンプレックス サンプルの計画

[コンプレックス サンプル] 分析手続きでは、有効な結果を得るために分析またはサンプル計画ファイルから分析の指定を行うことが必要です。

図 4-1  
[コンプレックス サンプル計画] ダイアログ ボックス



**計画。**分析またはサンプル計画ファイルのパスを指定します。

**結合確率。**PPS WOR 推定を使用して抽出されたクラスターに [不等な WOR] 推定を使用するには、結合確率を含む独立したファイルまたは、開いたデータセットを指定する必要があります。このファイルまたはデータセットは、サンプリング ウィザードで抽出を行う際に作成されます。

# コンプレックス サンプル度数分布表

[コンプレックス サンプルの度数分布表] 手続きでは、選択した変数の度数分布表が作成され、1 変量の統計量が表示されます。1 つ以上のカテゴリ変数で定義したサブグループによって、統計を要求することもできます。

**例。**National Health Interview Survey (NHIS) の結果に基づき、[コンプレックス サンプルの度数分布表] 手続きを使用して、米国国民のビタミン使用に関する 1 変量の表統計量を作成できます。この場合、この一般使用データに対する適切な分析計画を使用します。

**統計量。**この手続きでは、セル母集団のサイズ、表パーセント、標準誤差、信頼区間、変動係数、デザイン効果、デザイン効果の平方根、累計値、各推定値の重み付けのない度数の推定値が生成されます。さらに、セルの比率が同じかどうか検定するために、カイ 2 乗統計量および尤度比統計量が計算されます。

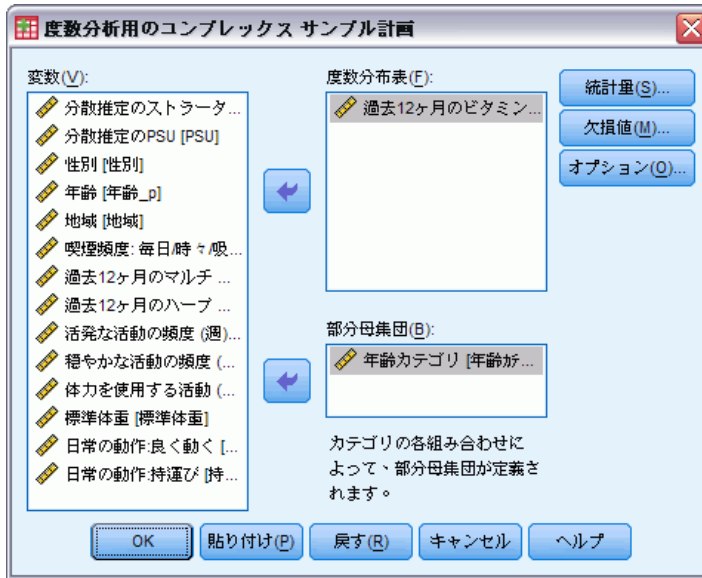
**データ。**度数分布表を作成する変数はカテゴリ変数である必要があります。部分母集団の変数は文字列にも数値にも設定できますが、カテゴリ変数である必要があります。

**仮定。**データ ファイルのケースは、[コンプレックス サンプル計画] [ダイアログ ボックス](#)で選択されたファイルの指定に従って分析されるコンプレックス デザインからのサンプルを表します。

## コンプレックス サンプル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 > コンプレックス サンプル > 度数分布表...
- ▶ 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルも選択できます。
- ▶ [続行] をクリックします。

図 5-1  
[度数分布表] ダイアログ ボックス

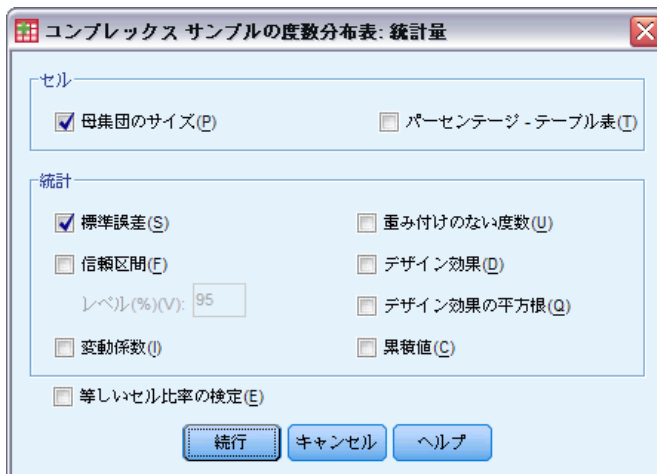


- ▶ 最低 1 つの度数変数を選択します。

オプションとして、部分母集団を定義する変数を指定できます。各部分母集団について、別々に統計が計算されます。

## コンプレックス サンプルの度数分布表の統計量

図 5-2  
[度数分布表: 統計] ダイアログ ボックス



**セル。**このグループで、セル母集団のサイズ、および表パーセントの推定値を要求します。

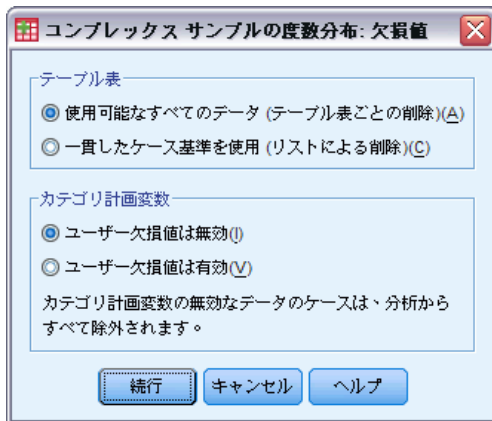
**統計量。**このグループで、母集団のサイズまたは表パーセントに関連する統計量を作成します。

- **標準誤差。**推定値の標準誤差。
- **信頼区間。**指定されたレベルを使用する、推定値の信頼区間。
- **変動係数。**推定値に対する推定値の標準誤差の比率。
- **重み付けのない度数。**推定値の計算に使用されるユニット数。
- **デザイン効果。**サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比。これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。
- **デザイン効果の平方根。**これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。
- **累積値。**変数のそれぞれの値を通した累積推定値。

**等しいセル比率の検定。**これを選択すると、変数のカテゴリの度数が等しいという仮説のカイ 2 乗検定と尤度比検定が作成されます。独立性の検定は、各変数ごとに実行されます。

## コンプレックス サンプルの欠損値

図 5-3  
[欠損値] ダイアログ ボックス



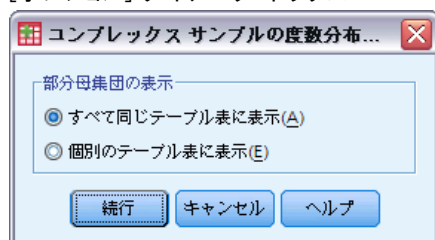
**テーブル。**このグループで、分析に使用されるケースを決定します。

- **使用可能なすべてのデータを使用。**欠損値はテーブルごとに決定されます。したがって、統計量の計算に使用されるケースは度数分布表またはクロス集計表に応じて変わる場合があります。
- **一貫したケース基準を使用。**欠損値は変数全体で決定されます。したがって、統計量の計算に使用されるケースはすべてのテーブルで同じになります。

**カテゴリ計画変数。**このグループで、ユーザー欠損値の有効または無効を決定します。

## コンプレックス サンプルのオプション

図 5-4  
[オプション] ダイアログ ボックス



**部分母集団の表示。**部分母集団を同じテーブルに表示するか、別のテーブルに表示するかを選択できます。

# コンプレックス サンプル記述統計

[コンプレックス サンプルの記述統計] 手続きでは、複数の変数に関する 1 変量の要約統計量が表示されます。1 つ以上のカテゴリ変数で定義したサブグループによって、統計を要求することもできます。

**例。**National Health Interview Survey (NHIS) の結果に基づき、[コンプレックス サンプルのクロス集計表] 手続きを使用して、米国国民の活動水準に関する 1 変量の記述統計表を作成できます。この場合、この一般使用データについての適切な分析計画を使用します。

**統計量。**この手続きでは、平均値と合計、t 検定、標準誤差、信頼区間、変動係数、重み付けのない度数、母集団のサイズ、デザイン効果、および各推定値のデザイン効果の平方根が得られます。

**データ。**測定方法はスケール変数である必要があります。部分母集団の変数は文字列にも数値にも設定できますが、カテゴリ変数である必要があります。

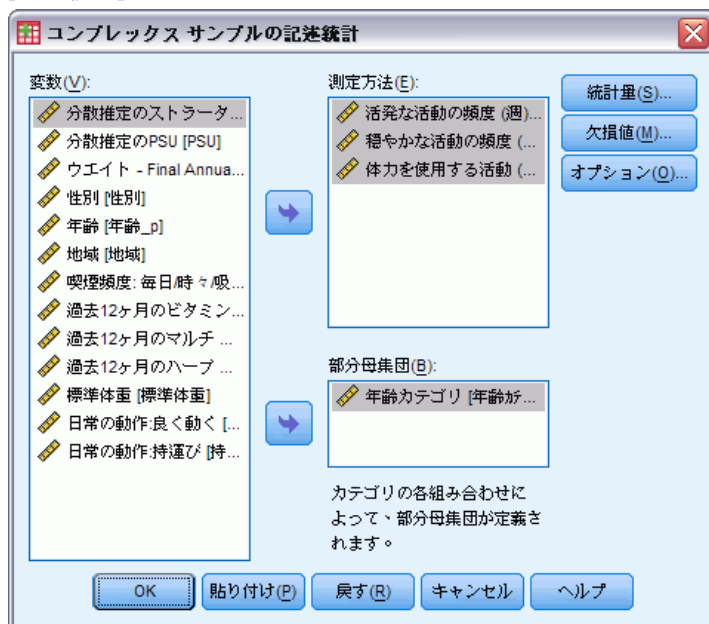
**仮定。**データ ファイルのケースは、[コンプレックス サンプル計画] [ダイアログ ボックス](#) で選択されたファイルの指定に従って分析されるコンプレックス デザインからのサンプルを表します。

## コンプレックス サンプルの記述統計を作成するには

- ▶ メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > 記述統計...
- ▶ 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルも選択できます。
- ▶ [続行] をクリックします。



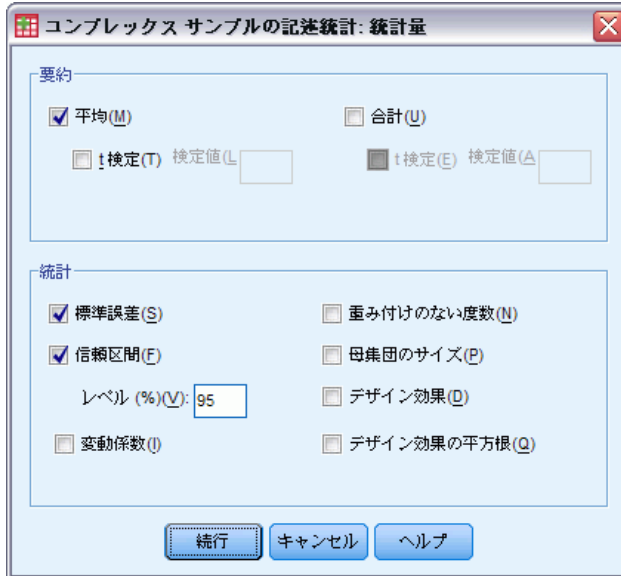
図 6-1  
[記述統計] ダイアログ ボックス



- ▶ 最低 1 つの測定変数を選択します。  
オプションとして、部分母集団を定義する変数を指定できます。各部分母集団について、別々に統計が計算されます。

## コンプレックス サンプルの記述統計の統計量

図 6-2  
[記述統計] ダイアログ ボックス



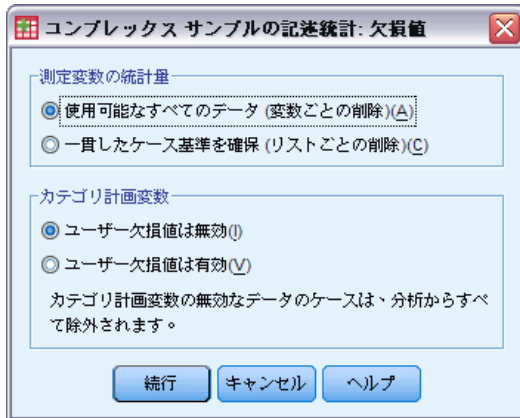
**要約。**このグループで、測定変数の平均と合計の推定値を要求できます。さらに、指定した値に対する t 検定も要求できます。

**統計量。**このグループでは、平均値または合計値に関連する統計量が作成されます。

- **標準誤差。**推定値の標準誤差。
- **信頼区間。**指定されたレベルを使用する、推定値の信頼区間。
- **変動係数。**推定値に対する推定値の標準誤差の比率。
- **重み付けのない度数。**推定値の計算に使用されるユニット数。
- **母集団のサイズ。**母集団の推定単位数。
- **デザイン効果。**サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比。これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。
- **デザイン効果の平方根。**これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。

## コンプレックス サンプルの記述統計の欠損値

図 6-3  
[記述統計: 欠損値] ダイアログ ボックス



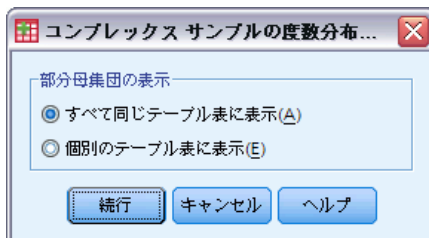
**測定変数の統計量。**このグループで、分析に使用されるケースを決定します。

- **使用可能なすべてのデータを使用。**欠損値は変数ごとに決定されます。したがって、統計量の計算に使用されるケースは測定変数に応じて変わる場合があります。
- **一貫したケース基準を確保。**欠損値は変数全体で決定されます。したがって、統計量の計算に使用されるケースはすべて同じになります。

**カテゴリ計画変数。**このグループで、ユーザー欠損値の有効または無効を決定します。

## コンプレックス サンプルのオプション

図 6-4  
[オプション] ダイアログ ボックス



**部分母集団の表示。**部分母集団を同じテーブルに表示するか、別のテーブルに表示するかを選択できます。

# コンプレックス サンプルのクロス集計表

[コンプレックス サンプルのクロス集計] 手続きでは、選択した変数組のクロス集計表が生成され、2次元統計量が表示されます。1つ以上のカテゴリ変数で定義したサブグループによって、統計を要求することもできます。

**例。**National Health Interview Survey (NHIS) の結果に基づき、[コンプレックス サンプルのクロス集計表] 手続きを使用して、米国国民のビタミン使用による喫煙頻度のクロス分類統計量を作成できます。この場合、この一般使用データについての適切な分析計画を使用します。

**統計量。**この手続きでは、セル母集団のサイズ、行、列、表パーセント、標準誤差、信頼区間、変動係数、予測値、デザイン効果、デザイン効果の平方根、残差、調整済み残差、および各推定値の重み付けのない度数の推定値が作成されます。オッズ比、相対リスク、およびリスク差は、 $2 \times 2$  の表に対して計算されます。さらに行と列の変数の独立性検定のために Pearson および尤度比統計量が計算されます。

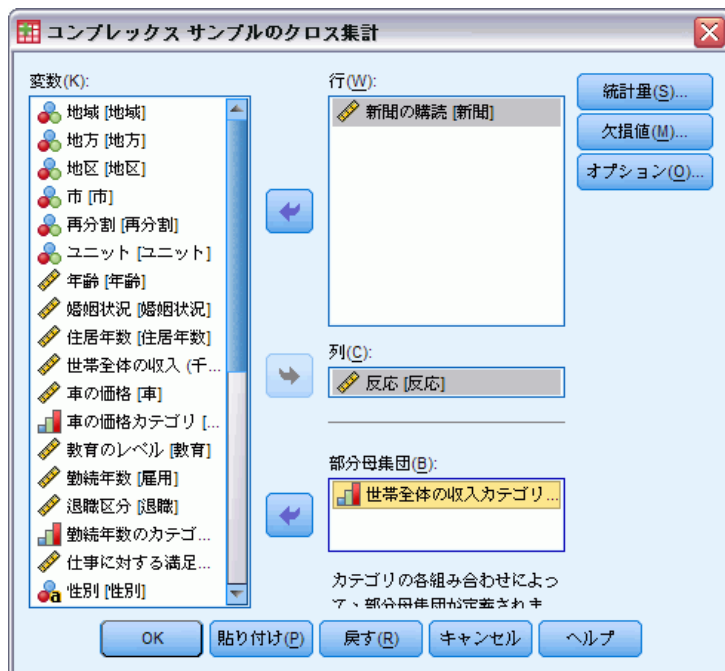
**データ。**行と列の変数は、カテゴリ変数である必要があります。部分母集団の変数は文字列にも数値にも設定できますが、カテゴリ変数である必要があります。

**仮定。**データ ファイルのケースは、[コンプレックス サンプル計画] **ダイアログ ボックス**で選択されたファイルの指定に従って分析されるコンプレックス デザインからのサンプルを表します。

## コンプレックス サンプル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 > コンプレックス サンプル > クロス集計表...
- ▶ 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルも選択できます。
- ▶ [続行] をクリックします。

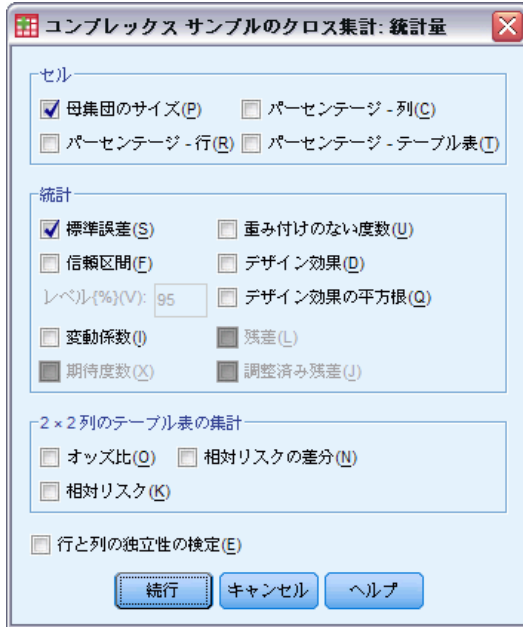
図 7-1  
[クロス集計表] ダイアログ ボックス



- ▶ 行変数と列変数をそれぞれ 1 つ以上選択します。  
オプションとして、部分母集団を定義する変数を指定できます。各部分母集団について、別々に統計が計算されます。

## コンプレックス サンプルのクロス集計表の統計量

図 7-2  
[クロス集計表: 統計量の指定] ダイアログ ボックス



**セル。**このグループで、セル母集団のサイズ、および行、列、表パーセントの推定値を要求します。

**統計量。**このグループで、母集団のサイズおよび行、列、表パーセントに関連する統計量を作成します。

- **標準誤差。**推定値の標準誤差。
- **信頼区間。**指定されたレベルを使用する、推定値の信頼区間。
- **変動係数。**推定値に対する推定値の標準誤差の比率。
- **期待度数。**行と列の変数が独立しているという仮説に基づいた、推定値の期待値。
- **重み付けのない度数。**推定値の計算に使用されるユニット数。
- **デザイン効果。**サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比。これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。
- **デザイン効果の平方根。**これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。

- **残差。**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。
- **調整済み残差。**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。

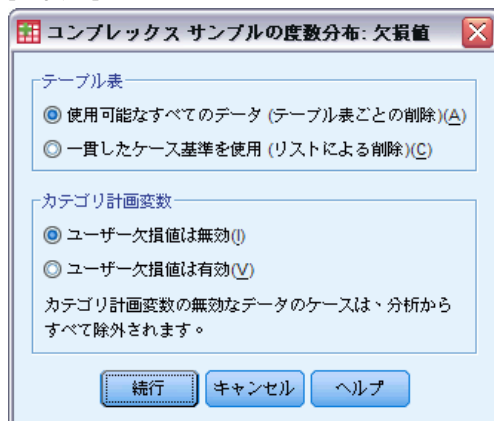
**2 × 2 列のテーブル表の集計。**このグループでは、行と列の変数にそれぞれ2つのカテゴリが含まれるテーブルの統計量が作成されます。それぞれが、ある因子の存在とあるイベントの発生の関連の強さの測定値になります。

- **オッズ比。**因子の発生がまれなときには、オッズ比を相対リスクの推定値として使うことができます。
- **相対リスク。**因子が発生しないリスクに対する、因子が発生するリスクの比。
- **相対リスクの差分。**因子が発生するリスクと、因子が発生しないリスクとの差。

**行と列の独立性の検定。**この手続きでは、行と列の変数が独立しているという仮説のカイ2乗検定と尤度比検定が作成されます。独立性の検定は、変数のペアごとに実行されます。

## コンプレックス サンプルの欠損値

図 7-3  
[欠損値] ダイアログ ボックス



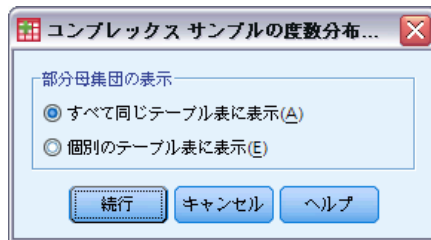
**テーブル。**このグループで、分析に使用されるケースを決定します。

- **使用可能なすべてのデータを使用。**欠損値はテーブルごとに決定されます。したがって、統計量の計算に使用されるケースは度数分布表またはクロス集計表に応じて変わる場合があります。
- **一貫したケース基準を使用。**欠損値は変数全体で決定されます。したがって、統計量の計算に使用されるケースはすべてのテーブルで同じになります。

**カテゴリ計画変数。**このグループで、ユーザー欠損値の有効または無効を決定します。

## コンプレックス サンプルのオプション

図 7-4  
[オプション] ダイアログ ボックス



**部分母集団の表示。**部分母集団を同じテーブルに表示するか、別のテーブルに表示するかを選択できます。



# コンプレックス サンプル比率

[コンプレックス サンプルの比率分析] 手続きでは、変数の比率に関する 1 変量の要約統計量が表示されます。1 つ以上のカテゴリ変数で定義したサブグループによって、統計を要求することもできます。

**例。**コンプレックス デザインに従って、データに適した分析計画を使用して実行された州全体の調査結果に基づき、[コンプレックス サンプルの比率分析] 手続きを使用することで、前回査定された資産額と現在の資産額との比を表す記述統計量を作成できます。

**統計量。**この手続きでは、比率の推定値、t 検定、標準誤差、信頼区間、変動係数、重み付けのない度数、母集団のサイズ、デザイン効果、およびデザイン効果の平方根が得られます。

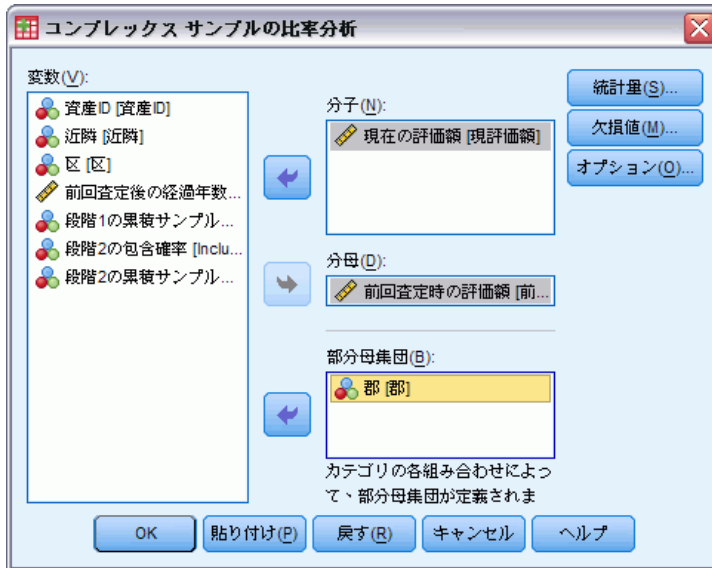
**データ。**分子と分母は正のスケール変数である必要があります。部分母集団の変数は文字列にも数値にも設定できますが、カテゴリ変数である必要があります。

**仮定。**データ ファイルのケースは、[コンプレックス サンプル計画] **ダイアログ ボックス**で選択されたファイルの指定に従って分析されるコンプレックス デザインからのサンプルを表します。

## コンプレックス サンプルの比率分析を算出するには

- ▶ メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > 比率...
- ▶ 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルも選択できます。
- ▶ [続行] をクリックします。

図 8-1  
[比率分析] ダイアログ ボックス

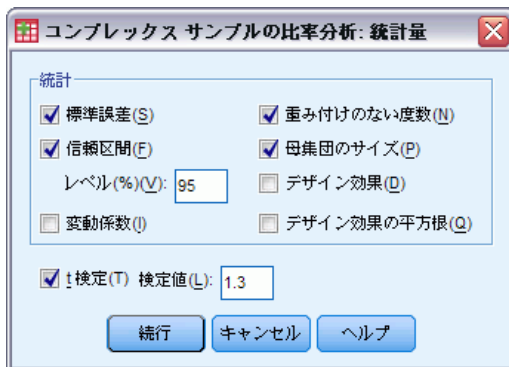


- ▶ 1 つ以上の分子変数と分母変数を選択します。

オプションとして、変数を指定して、生成される統計量のサブグループを定義できます。

## コンプレックス サンプルの比率分析の統計量

図 8-2  
[比率分析: 統計量] ダイアログ ボックス



**統計量。**このグループで、比率の推定値に関連する統計量を作成します。

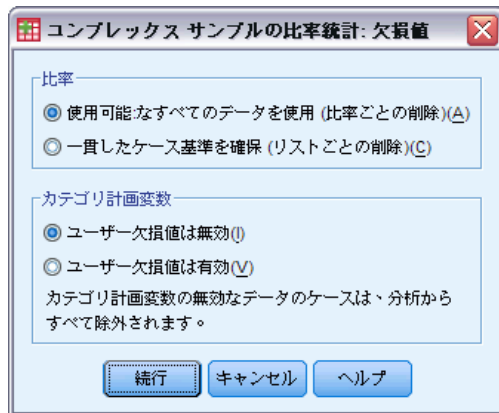
- **標準誤差。**推定値の標準誤差。
- **信頼区間。**指定されたレベルを使用する、推定値の信頼区間。
- **変動係数。**推定値に対する推定値の標準誤差の比率。

- **重み付けのない度数。**推定値の計算に使用されるユニット数。
- **母集団のサイズ。**母集団の推定単位数。
- **デザイン効果。**サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比。これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。
- **デザイン効果の平方根。**これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。

T **検定。**指定した値に対する推定値の t 検定を要求できます。

## コンプレックス サンプルの比率分析の欠損値

図 8-3  
[比率分析: 欠損値] ダイアログ ボックス



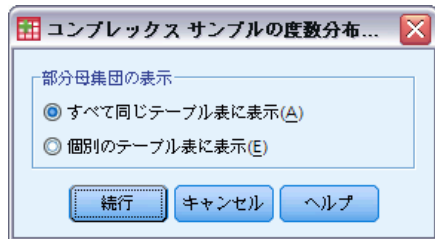
**比率。**このグループで、分析に使用されるケースを決定します。

- **使用可能なすべてのデータを使用。**欠損値は比率ごとに決定されます。したがって、統計量の計算に使用されるケースは分子と分母のペアごとに変わる場合があります。
- **一貫したケース基準を確保。**欠損値は変数全体で決定されます。したがって、統計量の計算に使用されるケースはすべて同じになります。

**カテゴリ計画変数。**このグループで、ユーザー欠損値の有効または無効を決定します。

## コンプレックス サンプルのオプション

図 8-4  
[オプション] ダイアログ ボックス



**部分母集団の表示。**部分母集団を同じテーブルに表示するか、別のテーブルに表示するかを選択できます。

# コンプレックス サンプルの一般線型モデル

コンプレックス サンプルの一般線型モデル (CSGLM) 手続きでは、コンプレックス サンプル法によって抽出したサンプルに対して、分散分析や共分散分析だけではなく、線型回帰分析を実行します。オプションとして、部分母集団の分析も実行できます。

**例:**ある食料雑貨店チェーンが、特定の顧客グループを対象に、購買習慣についてコンプレックス デザインに従って調査しました。その食料雑貨店は、調査結果と各顧客の前月の購買額を前提に、顧客の性別による分類や抽出計画を取り入れながら、専門店が顧客の 1 か月間の購買額にどの程度関わっているかを調査します。

**統計量。**この手続きでは、パラメータ推定値間の相関と共分散の他に、標準誤差、信頼区間、t 検定、デザイン効果、およびモデル パラメータに対するデザイン効果の平方根が推定されます。モデルの適合度、および従属変数と独立変数の記述統計も使用できます。また、モデルの因子と交互作用のレベルについての推定周辺平均値を求めることができます。

**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。共変量は、従属変数に関連する量的変数です。部分母集団の変数は文字列にも数値にも設定できますが、カテゴリ変数である必要があります。

**仮定** データ ファイルのケースは、[\[コンプレックス サンプル計画\] ダイアログ ボックス](#)で選択されたファイルの指定に従って分析されるコンプレックス デザインからのサンプルを表します。

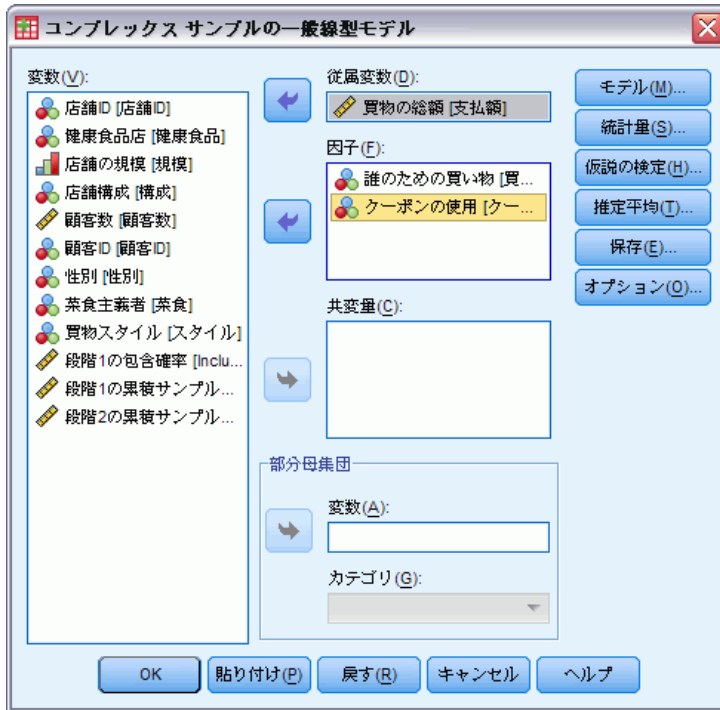
## コンプレックス サンプルの一般線型モデルを求める

メニューから次の項目を選択します。

分析(A) > コンプレックス サンプル > 一般線型モデル...

- ▶ 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルも選択できます。
- ▶ [\[続行\]](#) をクリックします。

図 9-1  
[一般線型モデル] ダイアログ ボックス

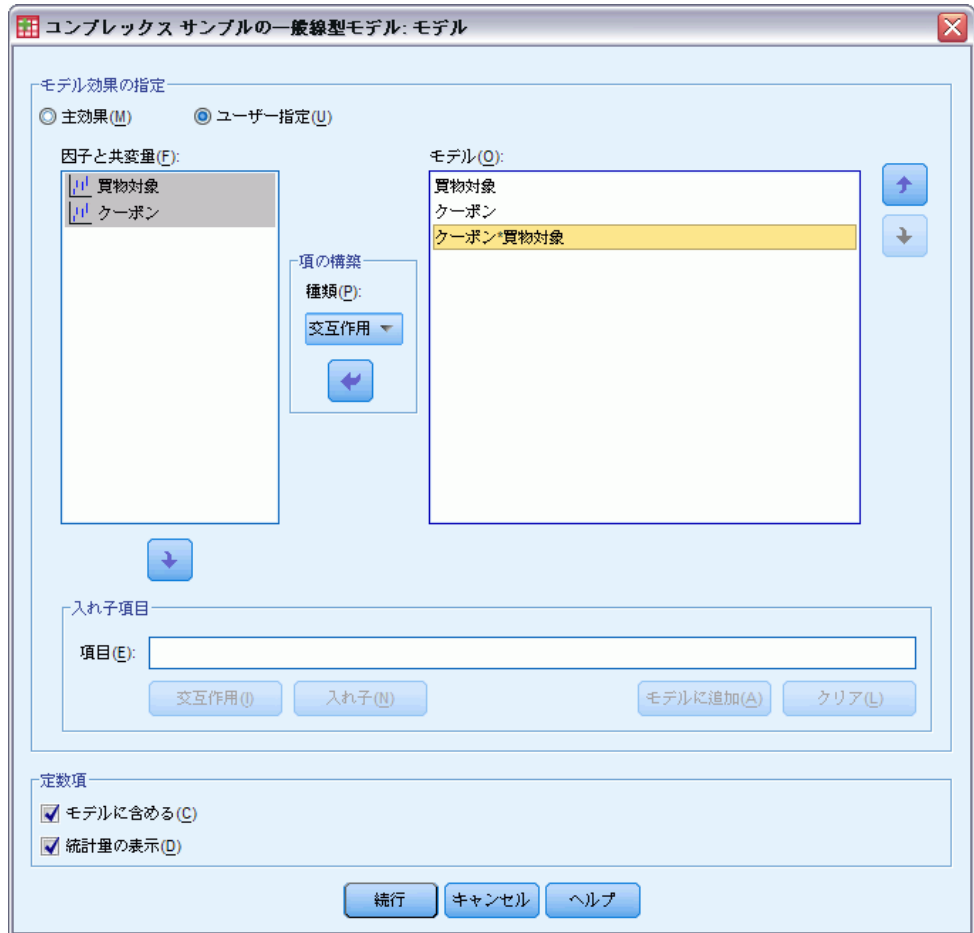


- ▶ [従属変数] ボックスに従属変数を選択します。

オプションとして、次の選択が可能です。

- 因子および共変量の変数をデータの内容に合わせて選びます。
- 変数を1つ指定して部分母集団を定義します。この分析は、部分母集団変数のカテゴリのうち、選択したカテゴリに対してのみ行われます。

図 9-2  
[モデル] ダイアログ ボックス



**モデル効果を指定。**デフォルトでは、この手続きで、メイン ダイアログ ボックスで指定した因子と共変量を使用して主効果モデルが作成されます。または、交互作用効果と入れ子項目を含む、ユーザーの指定によるモデルを作成できます。

### 非入れ子項目

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。

**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。

### 入れ子項目

この手続きでは、モデルに入れ子の項を構築できます。入れ子項目を構築すると、別の因子のレベルと相互作用しない因子または共変量の効果をモデリングできます。たとえば、ある食料雑貨店チェーンが、いくつかの店舗で顧客の消費傾向を調査すると仮定します。1 人の顧客が頻繁に通うのはこれらの店舗の 1 つだけであるため、「顧客」の効果は、「店舗の場所」の効果の**入れ子**になっていると言えます。

また、入れ子の項には、同一の共変量に関する多項式項などの交互作用効果を含めたり、複数レベルで入れ子になった項を追加したりできます。

**制限。**入れ子項目には、次の制限があります。

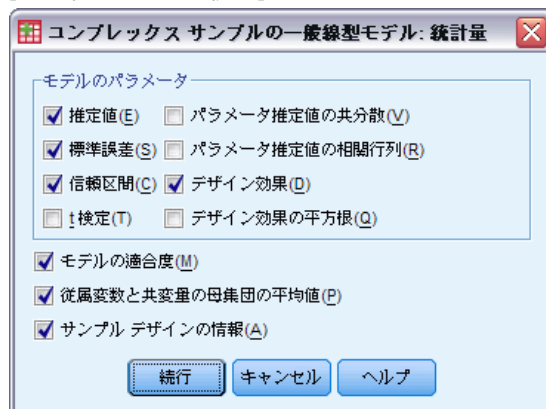
- 1 つの交互作用内のすべての因子は一意でなければなりません。したがって、A が因子である場合、A\*A の指定は無効です。
- 1 つの入れ子効果内のすべての因子は一意でなければなりません。したがって、A が因子である場合、A(A) の指定は無効です。
- 効果は共変量の入れ子にできません。したがって、A が因子で X が共変量である場合、A(X) の指定は無効です。

**定数項。**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。モデル内に定数項を含める場合でも、そのモデルに関連する統計量の抑制を選択できます。

## コンプレックス サンプルの一般線型モデルの統計量

図 9-3

[一般線型モデル: 統計] ダイアログ ボックス





**モデルのパラメータ。**このグループでは、モデルのパラメータに関連する統計量の表示を制御できます。

- **推定値。**係数の推定値を表示します。
- **標準誤差。**係数の推定値ごとの標準誤差を表示します。
- **信頼区間。**係数の推定値ごとの信頼区間を表示します。各区間の信頼水準は、[オプション] ダイアログ ボックスで設定します。
- **T 検定。**係数の推定値ごとの t 検定を表示します。検定ごとの帰無仮説は、係数の値が 0 の場合です。
- **パラメータ推定値の共分散。**モデル係数の共分散行列の推定値を表示します。
- **パラメータ推定値の相関。**モデル係数の相関行列の推定値を表示します。
- **デザイン効果。**サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比。これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。
- **デザイン効果の平方根。**これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。

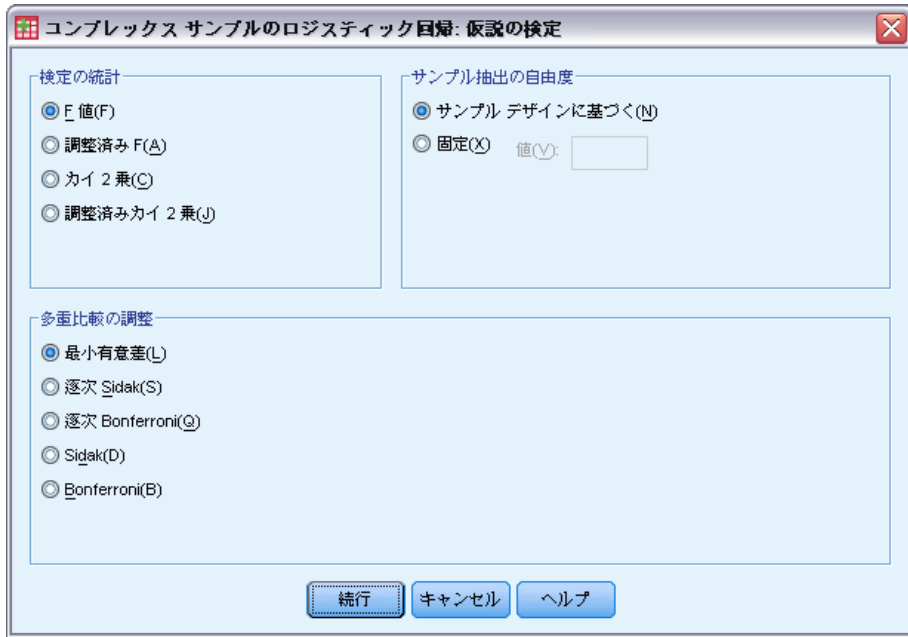
**モデルの適合度。** $R^2$  と平均平方誤差の統計量を表示します。

**従属変数と共変量の母集団の平均値。**従属変数、共変量、および因子に関する要約情報を表示します。

**サンプル デザインの情報。**重み付けのない度数、母集団のサイズを含め、サンプルに関する要約情報を表示します。

## コンプレックス サンプルの仮説の検定

図 9-4  
[仮説の検定] ダイアログ ボックス



**検定の統計。**このグループでは、仮説の検定に使用する統計量の種類を選択できます。[F 値]、[調整済み F]、[カイ 2 乗]、[調整済みカイ 2 乗]の中から選択します。

**サンプル抽出の自由度。**このグループでは、すべての検定統計量の p 値の計算に使用する抽出計画の自由度を制御できます。抽出計画に基づく場合は、主要な抽出単位数と抽出の第 1 段階のストラータ数とで値が異なります。また、正の整数を指定することにより自由度を設定することもできます。

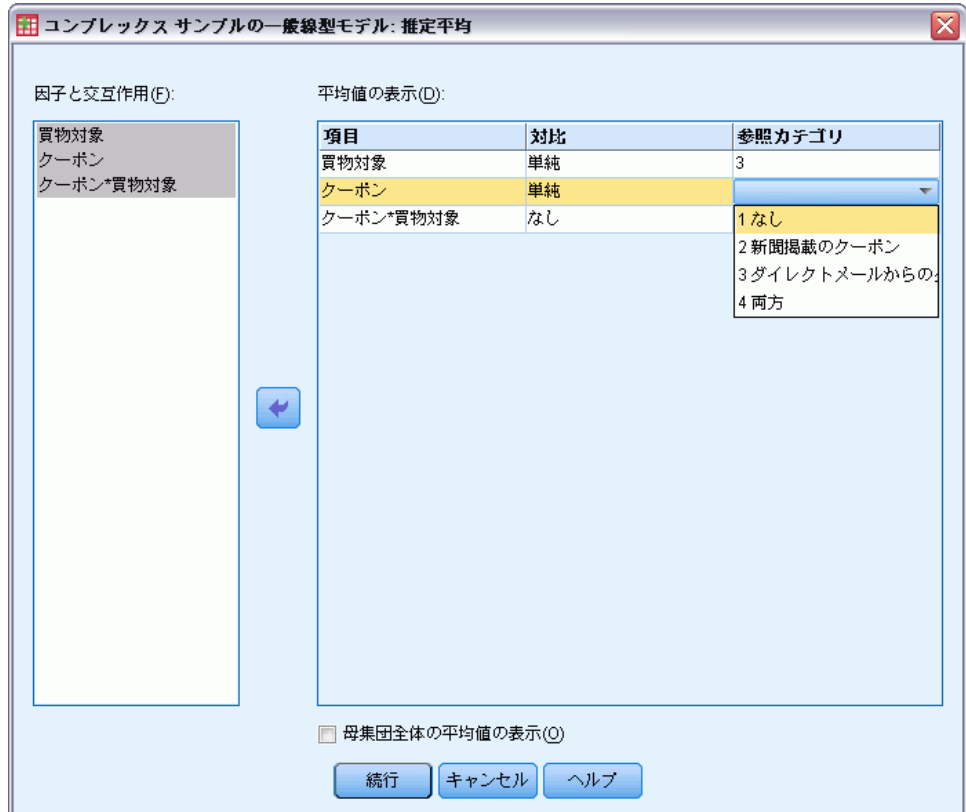
**多重比較の調整。**多重対比を用いた仮説検定を行う場合、各対比に関する有意水準を基に、全体の有意水準を調整できます。このグループでは、調整方法を選択できます。

- **最小有異差。**この方法では、帰無仮説の値と異なる線型の対比が存在するという仮説を棄却する確率の全体的な制御を行いません。
- **逐次 Sidak.**これは、すべて同じ有意水準を維持したままで個々の仮説が棄却される保守的でない逐次ステップダウン相対 Sidak プロシジャです。
- **逐次 Bonferroni.**個々の仮説を棄却しますが、すべて同じ有意水準を維持することでは、それほど控えめでもない逐次ステップダウン棄却 Bonferroni 手続きです。

- **Sidak.** この方法は Bonferroni の方法よりもっと狭い限界を提供します。
- **Bonferroni の方法.** この方法は、多重比較の検定がなされるときに、有意確率を修正します。

## コンプレックス サンプルの一般線型モデルの推定平均

図 9-5  
[一般線型モデル: 推定平均] ダイアログ ボックス



[推定平均] ダイアログ ボックスでは、[モデル] サブダイアログ ボックスで指定した因子と交互作用のレベルについて、モデルで推定される周辺平均を表示できます。母集団全体の平均も表示できます。

**項目。** 推定平均は、選択した因子と交互作用に対して計算されます。

**対比。** 対比により、どのように仮説の検定を設定して推定平均を計算するのかを決定します。

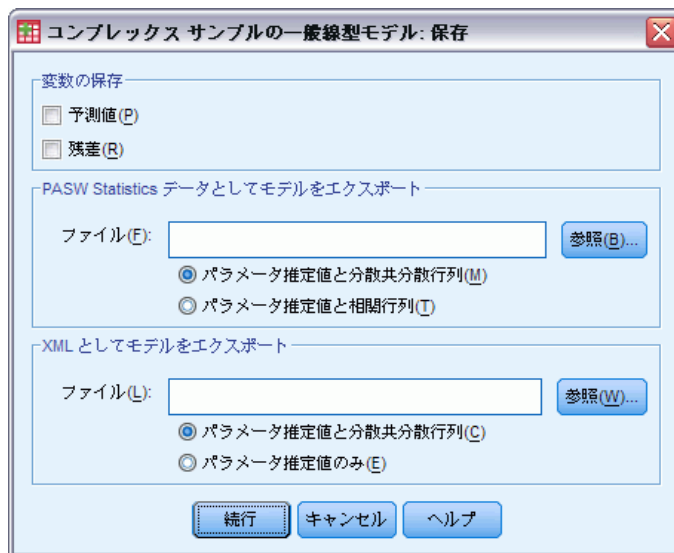
- **単純.** 各水準の平均と指定された水準の平均を比較します。この対比の型はコントロール群があるときに有効です。
- **偏差.** 各水準の平均 (参照カテゴリは除きます) をすべての水準 (全平均) の平均と比較します。因子の水準は任意の順序になります。

- **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。

**参照カテゴリ.** 単純対比と全平均対比では、参照カテゴリ、つまり、比較対象とする因子レベルを指定する必要があります。

## コンプレックス サンプルの一般線型モデルの保存

図 9-6  
[一般線型モデル: 保存] ダイアログ ボックス



**変数を保存。** このグループで、モデルで予測された値と残差を作業ファイルの新変数として保存できます。

**モデルを SPSS Statistics データとしてエクスポート。** パラメータ相関行列またはパラメータ共分散行列から構成される IBM® SPSS® Statistics データセットを出力します。これらの行列には、パラメータ推定値、標準誤差、有意確率、および自由度が含まれています。行列ファイルの変数の順序は次のとおりです。

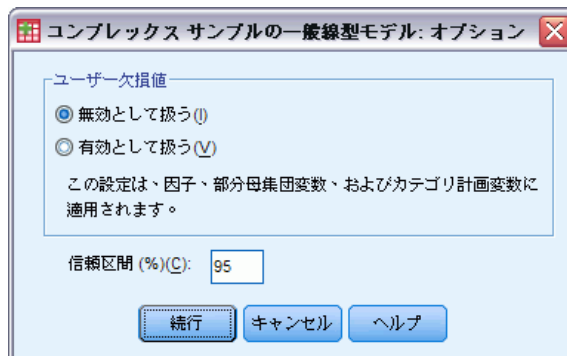
- **rowtype\_**。COV (共分散)、CORR (相関)、EST (パラメータ推定値)、SE (標準誤差)、SIG (有意水準)、および DF (抽出計画の自由度) の値 (および値ラベル) を取ります。各モデル パラメータに行型 COV (または CORR) を持つ個別のケースだけでなく、各モデル パラメータにその他の行型を持つ個別のケースがあります。
- **varname\_**。すべてのモデル パラメータの順序づけられたリストに応じて、行型 COV または CORR に対し値 P1、P2、... を取り、パラメータ推定値テーブルに表示されたパラメータ文字列に従って値ラベルを付けます。その他の行型については、セルが空白になります。
- **P1、P2、...** これらの変数は、すべてのモデル パラメータの順序づけられたリストに対応しており、パラメータ推定テーブルに表示されたパラメータ文字列に従った変数ラベルを持ち、行型に従って値を取ります。冗長パラメータの場合、すべての共分散は 0 に設定され、相関はシステム欠損値に設定されます。また、すべてのパラメータ推定値は 0 に設定され、すべての標準誤差、有意水準、残差自由度はシステム欠損値に設定されます。

注: 行列ファイルを読み込むその他の手続きでは、ここでエクスポートされたすべての行型が承認されない場合、このファイルをすぐに使用してより詳細な分析を行うことはできません。

**XML としてモデルをエクスポート。** 選択した場合、パラメータ推定値とパラメータ分散共分散行列を XML (PMML) 形式で保存します。このモデル ファイルを使用して、得点付けのために他のデータ ファイルにモデル情報を適用できます。

## コンプレックス サンプルの一般線型モデルのオプション

図 9-7  
[一般線型モデル: オプション] ダイアログ ボックス



**ユーザー欠損値。** 従属変数や共変量だけではなく、すべての計画変数でデータが有効である必要があります。これらの変数のデータが無効なケースは分析から除外されます。この項目を使用して、ユーザー欠損値をストラー

タ変数、クラスタ変数、部分母集団変数、および因子変数で有効な値として扱うかどうかを決定できます。

**信頼区間。**これは、係数の推定値と推定周辺平均値の信頼区間のレベルです。50 以上、100 未満の値を指定します。

## CSGLM コマンドの追加機能

コマンド シンタックス言語を使用して、次のことも実行できます。

- 線型に組み合わせた効果や値に対し、ユーザー指定による効果の検定を指定 (CUSTOM サブコマンドを使用)。
- 推定周辺平均の計算時に、平均値ではなく値で共変量を固定 (EMMEANS サブコマンドを使用)。
- 多項式対比の計量を指定 (EMMEANS サブコマンドを使用)。
- 特異性の点検で使用する許容値を指定 (CRITERIA サブコマンドを使用)。
- 保存した変数のユーザー指定の名前を作成 (SAVE サブコマンドを使用)。
- 一般推定可能関数を作成 (PRINT サブコマンドを使用)。

複雑なシンタックス情報については、「コマンド シンタックス リファレンス」を参照してください。

# コンプレックス サンプルのロジスティック回帰

[コンプレックス サンプルのロジスティック回帰] 手続きでは、コンプレックス サンプルング法によって抽出されたサンプルに対して、2 値または多項の従属変数のロジスティック回帰分析を実行します。オプションとして、部分母集団の分析も実行できます。

**例** ある融資担当者がコンプレックス デザインに従って、複数の支店で融資を受けた顧客の過去の記録を収集しました。担当者は、顧客の債務不履行が年齢、職歴、クレジットカードの負債額に関連している可能性があるかどうかを調べ、抽出計画を取り入れようとしています。

**統計量**。この手続きでは、パラメータ推定値間の相関と共分散の他に、指数化された推定値、標準誤差、信頼区間、t 検定、デザイン効果、およびモデルパラメータに対するデザイン効果の平方根が推定されます。擬 R<sup>2</sup> 統計量、分類テーブル、および従属変数と独立変数の記述統計も使用できます。

**データ**。従属変数はカテゴリ変数です。因子も同じくカテゴリ型です。共変量は、従属変数に関連する量的変数です。部分母集団の変数は文字列にも数値にも設定できますが、カテゴリ変数である必要があります。

**仮定** データ ファイルのケースは、[コンプレックス サンプル計画] **ダイアログ ボックス** で選択されたファイルの指定に従って分析されるコンプレックス デザインからのサンプルを表します。

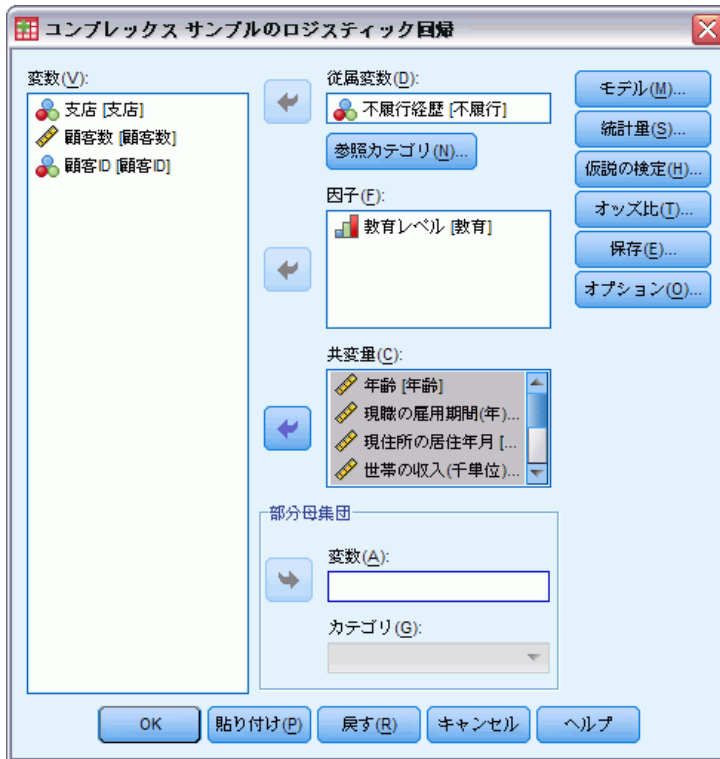
## コンプレックス サンプルのロジスティック回帰を取得する

メニューから次の項目を選択します。

分析(A) > コンプレックス サンプル > ロジスティック回帰...

- ▶ 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルも選択できます。
- ▶ [続行] をクリックします。

図 10-1  
[ロジスティック回帰] ダイアログ ボックス

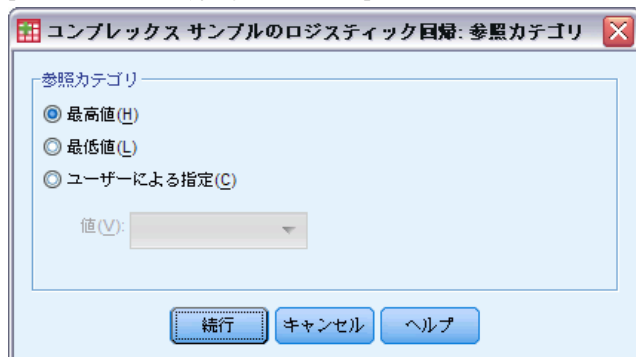


- ▶ [従属変数] ボックスに従属変数を選択します。  
オプションとして、次の選択が可能です。
- 因子および共変量の変数をデータの内容に合わせて選びます。
- 変数を1つ指定して部分母集団を定義します。この分析は、部分母集団変数のカテゴリのうち、選択したカテゴリに対してのみ行われます。



## コンプレックス サンプルのロジスティック回帰の参照カテゴリ

図 10-2  
[ロジスティック回帰: 参照カテゴリ] ダイアログ ボックス



デフォルトでは、ロジスティック回帰手続きにより、最高値カテゴリが参照カテゴリになります。このダイアログ ボックスでは参照カテゴリとして、最高値カテゴリ、最低値カテゴリ、またはユーザーの指定によるカテゴリを指定できます。

## コンプレックス サンプルのロジスティック回帰のモデル

図 10-3  
[ロジスティック回帰: モデル] ダイアログ ボックス



**モデル効果を指定。**デフォルトでは、この手続きで、メイン ダイアログ ボックスで指定した因子と共変量を使用して主効果モデルが作成されます。または、交互作用効果と入れ子項目を含む、ユーザーの指定によるモデルを作成できます。

### 非入れ子項目

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。

**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。

### 入れ子項目

この手続きでは、モデルに入れ子の項を構築できます。入れ子項目を構築すると、別の因子のレベルと相互作用しない因子または共変量の効果をモデリングできます。たとえば、ある食料雑貨店チェーンが、いくつかの店舗で顧客の消費傾向を調査すると仮定します。1 人の顧客が頻繁に通うのはこれらの店舗の 1 つだけであるため、「顧客」の効果は、「店舗の場所」の効果の入れ子になっていると言えます。

また、入れ子の項には、同一の共変量に関する多項式項などの交互作用効果を含めたり、複数レベルで入れ子になった項を追加したりできます。

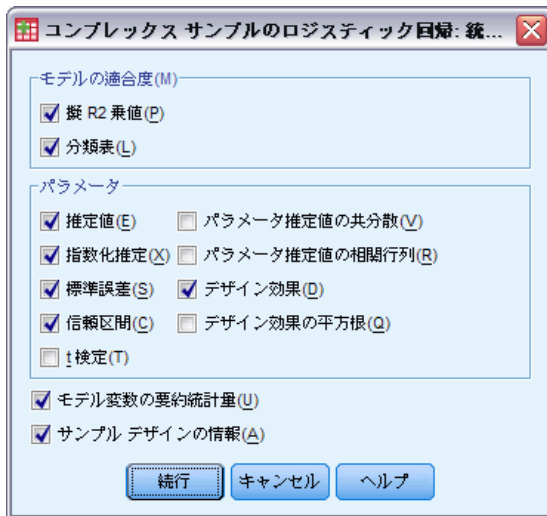
**制限。**入れ子項目には、次の制限があります。

- 1 つの交互作用内のすべての因子は一意でなければなりません。したがって、A が因子である場合、 $A*A$  の指定は無効です。
- 1 つの入れ子効果内のすべての因子は一意でなければなりません。したがって、A が因子である場合、 $A(A)$  の指定は無効です。
- 効果は共変量の入れ子にできません。したがって、A が因子で X が共変量である場合、 $A(X)$  の指定は無効です。

**定数項。**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。モデル内に定数項を含める場合でも、そのモデルに関連する統計量の抑制を選択できます。

## コンプレックス サンプルのロジスティック回帰の統計量

図 10-4  
[ロジスティック回帰: 統計] ダイアログ ボックス



**モデルの適合度。**モデル全体のパフォーマンスを測定する統計量の表示を制御します。

- **擬 R2 乗値。**ロジスティック回帰モデルには、線型回帰の  $R^2$  統計量に対応するものはありません。代わりに、 $R^2$  統計量のプロパティを模倣する測定値が複数存在します。
- **分類表。**従属変数について、モデルで予測されるカテゴリによる観測されるカテゴリのクロス分類表を表示します。

**パラメータ。**このグループでは、モデルのパラメータに関連する統計量の表示を制御できます。

- **推定値。**係数の推定値を表示します。
- **指数化推定。**係数の推定値でべき乗した自然対数の底を表示します。推定値のプロパティが統計上の検定に適している場合、指数化推定、つまり  $\exp(B)$  の解釈は比較的簡単です。
- **標準誤差。**係数の推定値ごとの標準誤差を表示します。
- **信頼区間。**係数の推定値ごとの信頼区間を表示します。各区間の信頼水準は、[オプション] ダイアログ ボックスで設定します。
- **T 検定。**係数の推定値ごとの t 検定を表示します。検定ごとの帰無仮説は、係数の値が 0 の場合です。
- **パラメータ推定値の共分散。**モデル係数の共分散行列の推定値を表示します。
- **パラメータ推定値の相関。**モデル係数の相関行列の推定値を表示します。

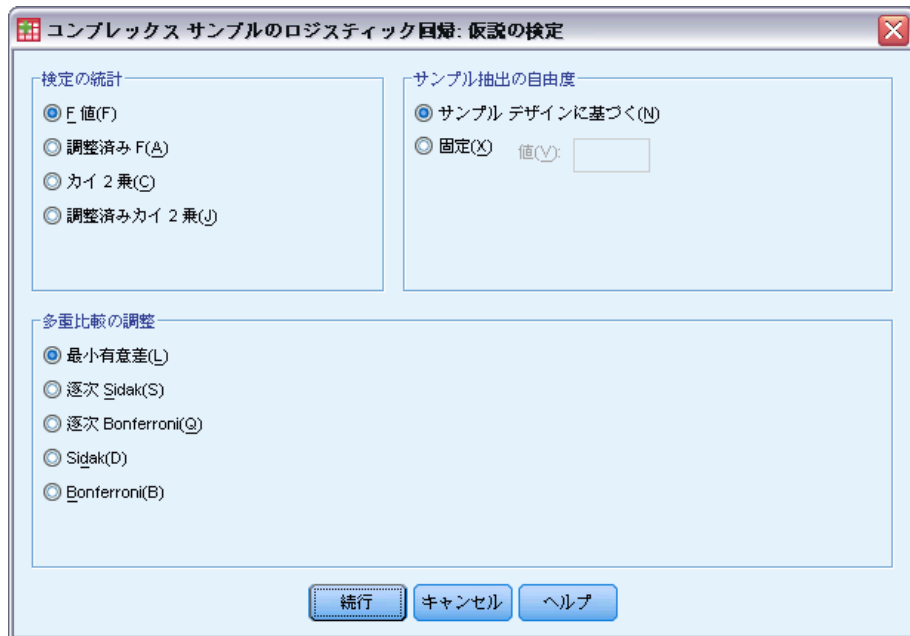
- **デザイン効果。** サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比。これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。
- **デザイン効果の平方根。** これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。

**モデル変数の要約統計量。** 従属変数、共変量、および因子に関する要約情報を表示します。

**サンプル デザインの情報。** 重み付けのない度数、母集団のサイズを含め、サンプルに関する要約情報を表示します。

## コンプレックス サンプルの仮説の検定

図 10-5  
[仮説の検定] ダイアログ ボックス



**検定の統計。** このグループでは、仮説の検定に使用する統計量の種類を選択できます。[F 値]、[調整済み F]、[カイ 2 乗]、[調整済みカイ 2 乗] の中から選択します。

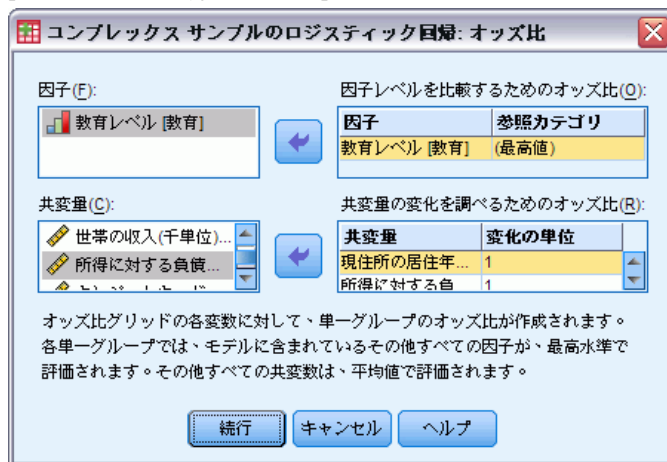
**サンプル抽出の自由度。** このグループでは、すべての検定統計量の p 値の計算に使用する抽出計画の自由度を制御できます。抽出計画に基づく場合は、主要な抽出単位数と抽出の第 1 段階のストラータ数とで値が異なります。また、正の整数を指定することにより自由度を設定することもできます。

**多重比較の調整。**多重対比を用いた仮説検定を行う場合、各対比に関する有意水準を基に、全体の有意水準を調整できます。このグループでは、調整方法を選択できます。

- **最小有異差。**この方法では、帰無仮説の値と異なる線型の対比が存在するという仮説を棄却する確率の全体的な制御を行いません。
- **逐次 Sidak.**これは、すべて同じ有意水準を維持したままで個々の仮説が棄却される保守的でない逐次ステップダウン相対 Sidak プロシジャです。
- **逐次 Bonferroni.**個々の仮説を棄却しますが、すべて同じ有意水準を維持することでは、それほど控えめでもない逐次ステップダウン棄却 Bonferroni 手続きです。
- **Sidak.**この方法は Bonferroni の方法よりもっと狭い限界を提供します。
- **Bonferroni の方法.**この方法は、多重比較の検定がなされるときに、有意確率を修正します。

## コンプレックス サンプルのロジスティック回帰のオッズ比

図 10-6  
[ロジスティック回帰: オッズ比] ダイアログ ボックス



[オッズ比] ダイアログ ボックスでは、指定した因子と共変量について、モデルで推定されるオッズ比を表示できます。個々のオッズ比が、参照カテゴリを除く従属変数のカテゴリごとに計算されます。

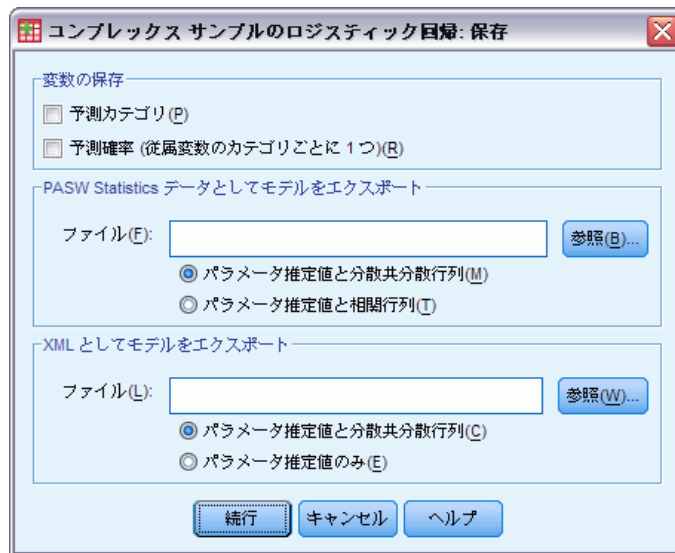
**因子。**選択した因子ごとに、その因子の各カテゴリのオッズと指定した参照カテゴリのオッズの比率を表示します。

**共変量。**選択した共変量ごとに、(変量の平均値 + 指定した変化の単位)でのオッズと平均値のオッズの比率を表示します。

因子または共変量のオッズ比の計算時には、その他すべての因子が最高水準で、また、その他すべての共変量が平均値で固定されます。因子または共変量とモデルの他の予測値との間に交互作用がある場合、オッズ比は指定した変数の変化だけではなく交互作用が見られる変数の値にも依存します。指定した共変量とモデル内の共変量自体との間に交互作用がある場合（たとえば age\*age）、オッズ比は共変量の変化と共変量の値の両方に依存します。

## コンプレックス サンプルのロジスティック回帰の保存

図 10-7  
[ロジスティック回帰: 保存] ダイアログ ボックス



**変数を保存。**このグループでは、モデルで予測されたカテゴリと予測確率をアクティブなデータセットの新変数として保存できます。

**モデルを SPSS Statistics データとしてエクスポート。**パラメータ相関行列またはパラメータ共分散行列から構成される IBM® SPSS® Statistics データセットを出力します。これらの行列には、パラメータ推定値、標準誤差、有意確率、および自由度が含まれています。行列ファイルの変数の順序は次のとおりです。

- **rowtype\_。**COV (共分散)、CORR (相関)、EST (パラメータ推定値)、SE (標準誤差)、SIG (有意水準)、および DF (抽出計画の自由度) の値 (および値ラベル) を取ります。各モデル パラメータに行型 COV (または CORR) を持つ個別のケースだけでなく、各モデル パラメータにその他の行型を持つ個別のケースがあります。

- **varname\_.** すべてのモデルパラメータの順序づけられたリストに応じて、行型 COV または CORR に対し値 P1、P2、... を取り、パラメータ推定値テーブルに表示されたパラメータ文字列に従って値ラベルを付けます。その他の行型については、セルが空白になります。
- **P1、P2、...** これらの変数は、すべてのモデルパラメータの順序づけられたリストに対応しており、パラメータ推定テーブルに表示されたパラメータ文字列に従った変数ラベルを持ち、行型に従って値を取ります。冗長パラメータの場合、すべての共分散は 0 に設定され、相関はシステム欠損値に設定されます。また、すべてのパラメータ推定値は 0 に設定され、すべての標準誤差、有意水準、残差自由度はシステム欠損値に設定されます。

注: 行列ファイルを読み込むその他の手続きでは、ここでエクスポートされたすべての行型が承認されない場合、このファイルをすぐに使用してより詳細な分析を行うことはできません。

**XML としてモデルをエクスポート。** 選択した場合、パラメータ推定値とパラメータ分散共分散行列を XML (PMML) 形式で保存します。このモデルファイルを使用して、得点付けのために他のデータファイルにモデル情報を適用できます。

## コンプレックス サンプルのロジスティック回帰のオプション

図 10-8  
[ロジスティック回帰:オプション] ダイアログ ボックス

コンプレックス サンプルのロジスティック回帰: オプション

**推定**

最大反復回数 (M): 100

最大段階 2 分 (S): 5

パラメータ推定値の変化に基づいて反復回数を制限 (L)

最小変更 (H): 0.000001 種類 (Y): 相対値

対数尤度の変化に基づいて反復回数を制限 (T)

最小変更 (G): 種類 (P): 相対値

データポイントの完全な区切りを確認 (K)

開始している反復 (R): 20

反復の記述を表示 (Q)

増分 (N): 1

**ユーザー欠損値**

無効として扱う (I)

有効として扱う (V)

この設定は、カテゴリ計画変数とモデル変数に適用されます。

信頼区間 (%) (C): 95

続行 キャンセル ヘルプ



**推定。**このグループでは、モデルの推定で使用するさまざまな基準を制御します。

- **最大反復回数。**アルゴリズムで実行される反復の最大回数です。負でない整数を指定してください。
- **最大段階 2 分。**対数尤度が増加するか、最大段階 2 分に達するまで、反復ごとにステップ サイズが 0.5 の因子ずつ減少します。正の整数を指定します。
- **パラメータ推定値の変化に基づいて反復回数を制限。**選択すると、パラメータ推定値の絶対変化または相対変化が指定した値よりも小さい反復の後にアルゴリズムが停止します。指定する値は負でない必要があります。
- **対数尤度の変化に基づいて反復回数を制限。**選択した場合、対数尤度関数の絶対変化または相対変化が指定した値よりも小さい反復の後にアルゴリズムが停止します。
- **データポイントの完全な区切りを確認。**選択した場合、パラメータ推定値が一意的な値になるようにアルゴリズムによって検定が行われます。区切りは、この手続きによって各ケースが正しく分類されるモデルを作成できる場合に存在します。
- **反復の記述を表示。**パラメータ推定値、および 0 回目（初期推定値）から始まる反復の n 回ごとの統計量を表示します。反復の記述の出力を選択した場合、n の値にかかわらず、最後の反復は常に出力されます。

**ユーザー欠損値。**従属変数や共変量だけではなく、すべての計画変数でデータが有効である必要があります。これらの変数のデータが無効なケースは分析から除外されます。この項目を使用して、ユーザー欠損値をストラータ変数、クラスタ変数、部分母集団変数、および因子変数で有効な値として扱うかどうかを決定できます。

**信頼区間。**これは、係数の推定値、指数化された係数の推定値、およびオッズ比の信頼区間のレベルです。50 以上、100 未満の値を指定します。

## CSLOGISTIC コマンドの追加機能

コマンド シンタックス言語を使用して、次のことも実行できます。

- 線型に組み合わせた効果や値に対し、ユーザー指定による効果の検定を指定 (CUSTOM サブコマンドを使用)。
- 因子と共変量のオッズ比の計算時におけるその他のモデル変数の値の固定 (ODDSRATIOS サブコマンドを使用)。
- 特異性の点検で使用する許容値を指定 (CRITERIA サブコマンドを使用)。
- 保存した変数のユーザー指定の名前を作成 (SAVE サブコマンドを使用)。
- 一般推定可能関数を作成 (PRINT サブコマンドを使用)。

複雑なシンタックス情報については、「コマンド シンタックス リファレンス」を参照してください。

# コンプレックス サンプルの順序 回帰

[コンプレックス サンプルの順序回帰] 手続きでは、コンプレックス サンプルング法によって抽出されたサンプルに対して、2 値または順序の従属変数の順序回帰分析を実行します。オプションとして、部分母集団の分析も実行できます。

**例:** 議会開会前に法案の提出を検討している議員は、市民が法案を支持しているかどうか、法案への支持がどの程度有権者の人口統計に関連しているかに関心を持っています。そこで、世論調査会社が複合抽出計画に基づいて対面式の調査を実施することになりました。コンプレックス サンプルの順序回帰を使用すると、有権者の人口統計に基づいて法案の支持水準のモデルを当てはめることができます。

**データ。** 従属変数は順序変数です。因子も同じくカテゴリ型です。共変量は、従属変数に関連する量的変数です。部分母集団の変数は文字列にも数値にも設定できますが、カテゴリ変数である必要があります。

**仮定。** データ ファイルのケースは、[コンプレックス サンプル計画] **ダイアログ ボックス** で選択されたファイルの指定に従って分析されるコンプレックス デザインからのサンプルを表します。

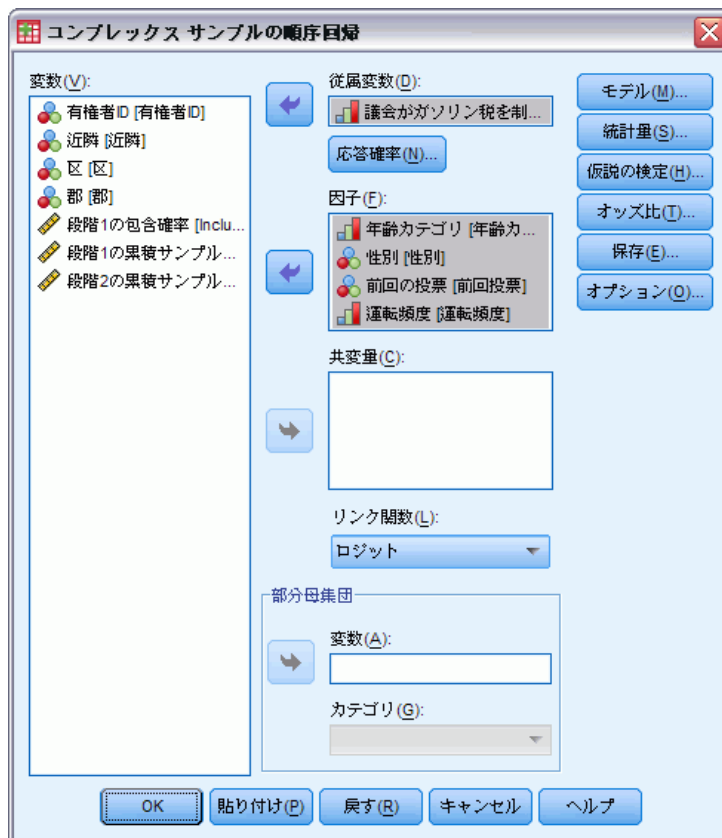
## コンプレックス サンプルの順序回帰を求めるには

メニューから次の項目を選択します。

分析(A) > コンプレックス サンプル > 順序回帰...

- ▶ 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルも選択できます。
- ▶ [続行] をクリックします。

図 11-1  
[順序回帰] 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。

オプションとして、次の選択が可能です。

- 因子および共変量の変数をデータの内容に合わせて選びます。
- 変数を 1 つ指定して部分母集団を定義します。この分析は、分散がデータセット全体に基づいて適正に推定される場合でも、部分母集団変数のカテゴリのうち、選択したカテゴリに対してのみ行われます。
- リンク関数を選択します。

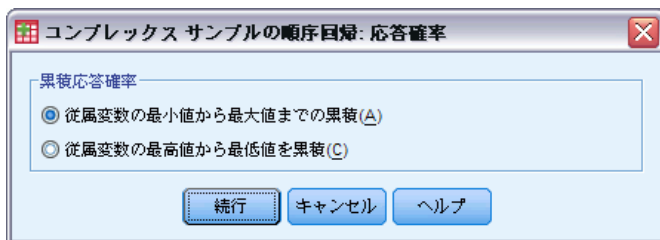
**リンク関数。**リンク関数とは、モデルの推定を行えるように累積確率を変換するためのものです。利用できるリンク関数は次の表に示す 5 種類です。

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

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
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)	$\tan(\pi(\xi - 0.5))$	潜在変数に多数の外れ値が存在する

## コンプレックス サンプルの順序回帰の応答確率

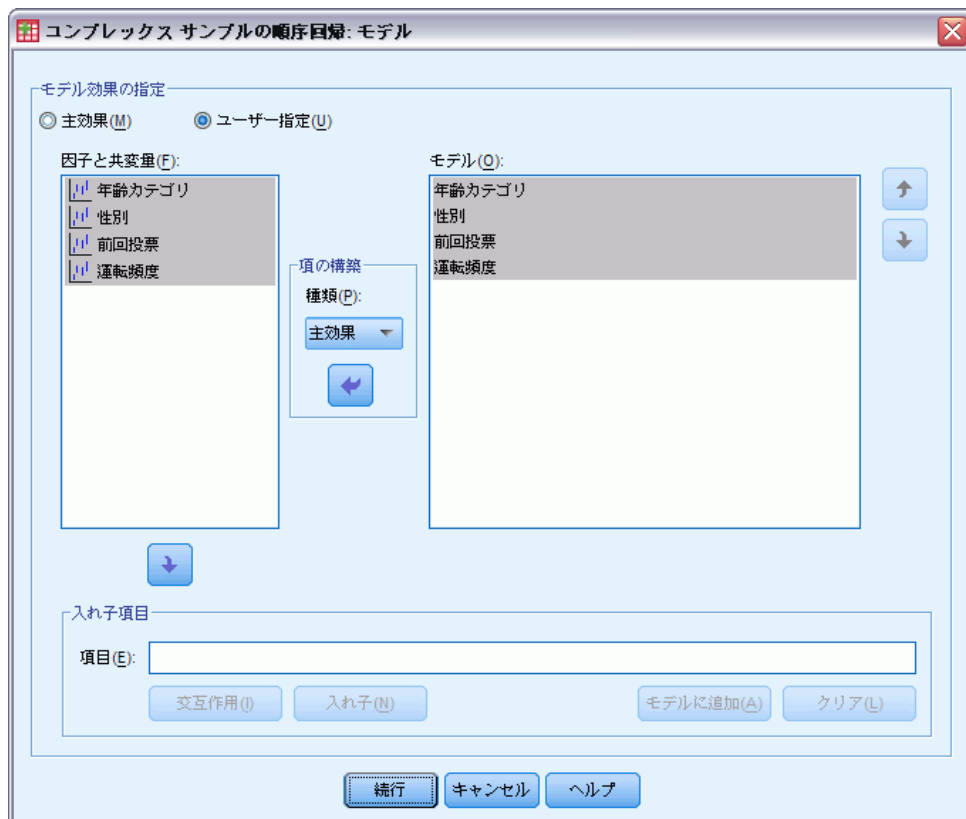
図 11-2  
[順序回帰: 応答確率] ダイアログ ボックス



[応答確率] ダイアログ ボックスでは、応答の累積確率（従属変数の特定のカテゴリに属す、あるいはそれを含む確率）を、従属変数の値の増加に応じて増加するのか、あるいは従属変数の値の減少に応じて増加するのかを指定できます。

## コンプレックス サンプルの順序回帰のモデル

図 11-3  
[順序回帰モデル] ダイアログ ボックス



**モデル効果を指定。**デフォルトでは、この手続きで、メインダイアログボックスで指定した因子と共変量を使用して主効果モデルが作成されます。または、交互作用効果と入れ子項目を含む、ユーザーの指定によるモデルを作成できます。

### 非入れ子項目

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。

**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。

### 入れ子項目

この手続きでは、モデルに入れ子の項を構築できます。入れ子項目を構築すると、別の因子のレベルと相互作用しない因子または共変量の効果をモデリングできます。たとえば、ある食料雑貨店チェーンが、いくつかの店舗で顧客の消費傾向を調査すると仮定します。1 人の顧客が頻繁に通うのはこれらの店舗の 1 つだけであるため、「顧客」の効果は、「店舗の場所」の効果の**入れ子**になっていると言えます。

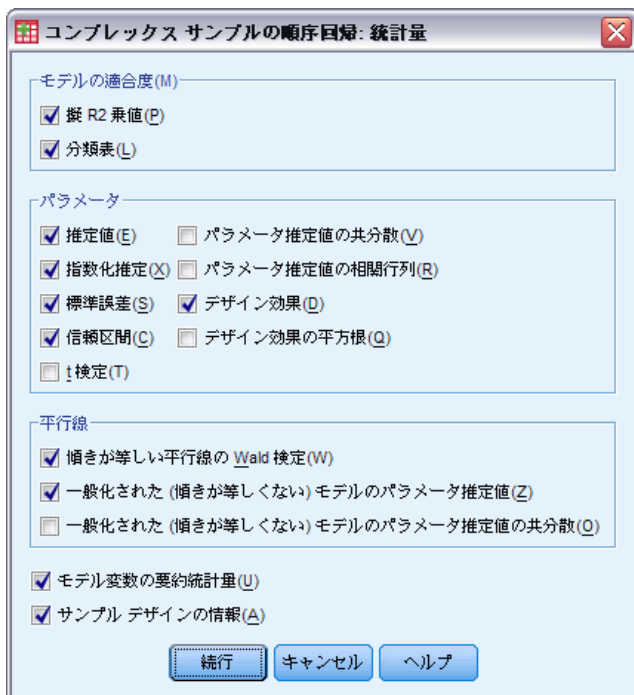
また、入れ子の項には、同一の共変量に関する多項式項などの交互作用効果を含めたり、複数レベルで入れ子になった項を追加したりできます。

**制限。**入れ子項目には、次の制限があります。

- 1 つの交互作用内のすべての因子は一意でなければなりません。したがって、A が因子である場合、 $A*A$  の指定は無効です。
- 1 つの入れ子効果内のすべての因子は一意でなければなりません。したがって、A が因子である場合、 $A(A)$  の指定は無効です。
- 効果は共変量の入れ子にできません。したがって、A が因子で X が共変量である場合、 $A(X)$  の指定は無効です。

## コンプレックス サンプルの順序回帰の統計量

図 11-4  
[順序回帰: 統計量] ダイアログ ボックス



**モデルの適合度。**モデル全体のパフォーマンスを測定する統計量の表示を制御します。

- **擬 R2 乗値。**順序回帰モデルには、線型回帰の  $R^2$  統計量に対応するものがありません。代わりに、 $R^2$  統計量のプロパティを模倣する測定値が複数存在します。
- **分類表。**従属変数について、モデルで予測されるカテゴリによる観測されるカテゴリのクロス分類表を表示します。

**パラメータ。**このグループでは、モデルのパラメータに関連する統計量の表示を制御できます。

- **推定値。**係数の推定値を表示します。
- **指数化推定。**係数の推定値でべき乗した自然対数の底を表示します。推定値のプロパティが統計上の検定に適している場合、指数化推定、つまり  $\exp(B)$  の解釈は比較的簡単です。
- **標準誤差。**係数の推定値ごとの標準誤差を表示します。
- **信頼区間。**係数の推定値ごとの信頼区間を表示します。各区間の信頼水準は、[オプション] ダイアログ ボックスで設定します。



- **T 検定。**係数の推定値ごとの t 検定を表示します。検定ごとの帰無仮説は、係数の値が 0 の場合です。
- **パラメータ推定値の共分散。**モデル係数の共分散行列の推定値を表示します。
- **パラメータ推定値の相関。**モデル係数の相関行列の推定値を表示します。
- **デザイン効果。**サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比。これはコンプレックスデザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。
- **デザイン効果の平方根。**コンプレックス デザインを指定する効果の測度。標準誤差と比較可能な単位で表され、値が 1 から遠ざかるほど効果が大きいことを意味します。

**平行線。**このグループでは、各応答カテゴリ（最後のものを除く）に個別の回帰直線が当てはまる、平行線以外のモデルに関する統計量を要求できます。

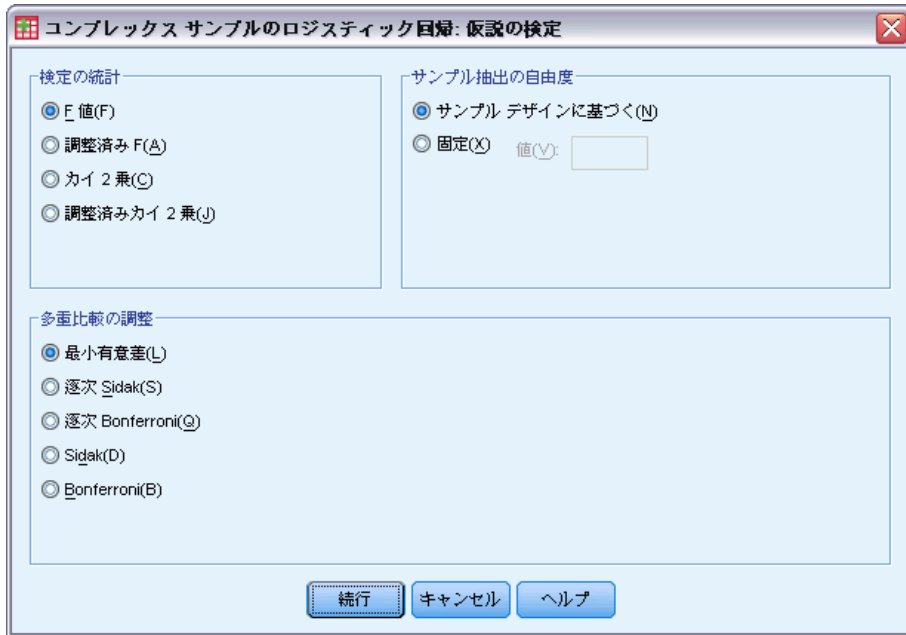
- **Wald 検定。**すべての累積応答について回帰パラメータが等しいという帰無仮説の検定を行うことができます。平行線以外のモデルが推定され、等しいパラメータの Wald 検定が適用されます。
- **パラメータ推定値。**平行線以外のモデルの係数および標準誤差の推定値を表示します。
- **パラメータ推定値の共分散。**平行線以外のモデル係数の共分散行列の推定値を表示します。

**モデル変数の要約統計量。**従属変数、共変量、および因子に関する要約情報を表示します。

**サンプル デザインの情報。**重み付けのない度数、母集団のサイズを含め、サンプルに関する要約情報を表示します。

## コンプレックス サンプルの仮説の検定

図 11-5  
[仮説の検定] ダイアログ ボックス



**検定の統計。**このグループでは、仮説の検定に使用する統計量の種類を選択できます。[F 値]、[調整済み F]、[カイ 2 乗]、[調整済みカイ 2 乗]の中から選択します。

**サンプル抽出の自由度。**このグループでは、すべての検定統計量の p 値の計算に使用する抽出計画の自由度を制御できます。抽出計画に基づく場合は、主要な抽出単位数と抽出の第 1 段階のストラータ数とで値が異なります。また、正の整数を指定することにより自由度を設定することもできます。

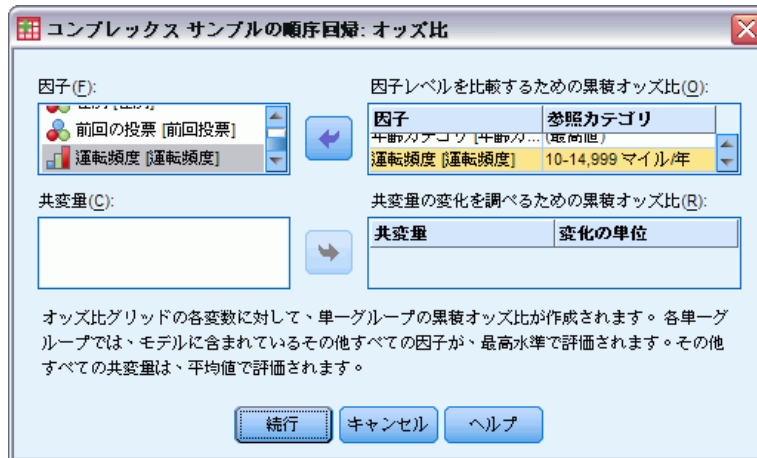
**多重比較の調整。**多重対比を用いた仮説検定を行う場合、各対比に関する有意水準を基に、全体の有意水準を調整できます。このグループでは、調整方法を選択できます。

- **最小有意差。**この方法では、帰無仮説の値と異なる線型の対比が存在するという仮説を棄却する確率の全体的な制御を行いません。
- **逐次 Sidak。**これは、すべて同じ有意水準を維持したままで個々の仮説が棄却される保守的でない逐次ステップダウン相対 Sidak プロシジャです。
- **逐次 Bonferroni。**個々の仮説を棄却しますが、すべて同じ有意水準を維持することでは、それほど控えめでもない逐次ステップダウン棄却 Bonferroni 手続きです。

- **Sidak.** この方法は Bonferroni の方法よりもっと狭い限界を提供します。
- **Bonferroni の方法.** この方法は、多重比較の検定がなされるときに、有意確率を修正します。

## コンプレックス サンプルの順序回帰のオッズ比

図 11-6  
[順序回帰: オッズ比] ダイアログ ボックス



[オッズ比] ダイアログ ボックスでは、指定した因子と共変量について、モデルで推定される累積オッズ比を表示できます。この機能は、ロジットリンク関数を使用したモデルでのみ使用できます。最後のカテゴリを除く従属変数の全カテゴリに対して 1 つの累積オッズ比が計算されます。比例オッズモデルでは、オッズ比はすべて等しいと仮定します。

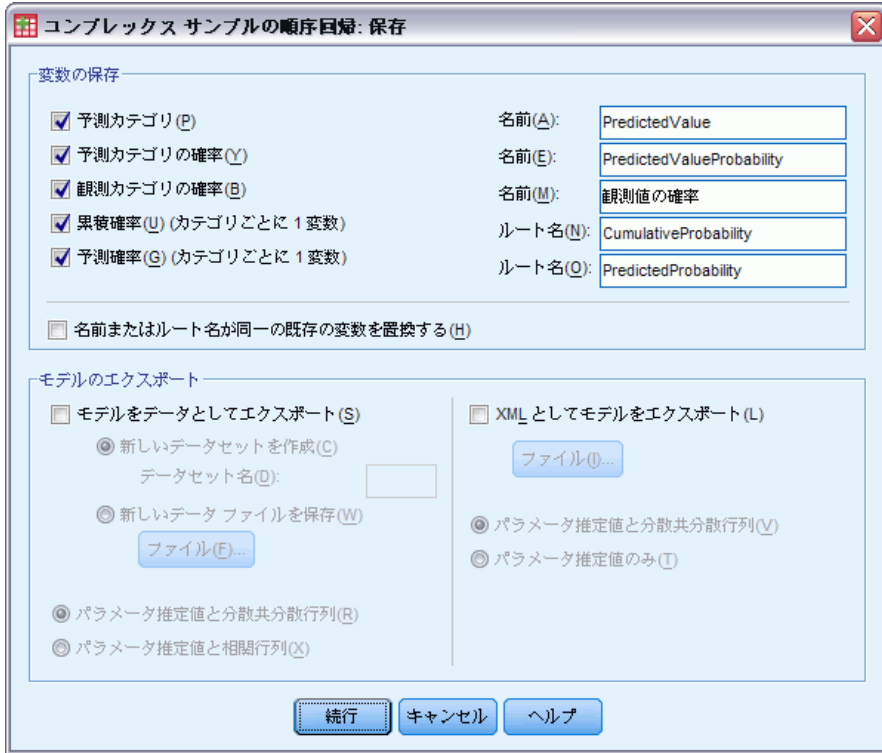
**因子。** 選択した因子ごとに、その因子の各カテゴリのオッズと指定した参照カテゴリの累積オッズの比率を表示します。

**共変量。** 選択した共変量ごとに、(変量の平均値 + 指定した変化の単位) での累積オッズと平均値のオッズの比率を表示します。

因子または共変量のオッズ比の計算時には、その他すべての因子が最高水準で、また、その他すべての共変量が平均値で固定されます。因子または共変量とモデルの他の予測値との間に交互作用がある場合、オッズ比は指定した変数の変化だけでなく交互作用が見られる変数の値にも依存します。指定した共変量とモデル内の共変量自体との間に交互作用がある場合 (たとえば age\*age)、オッズ比は共変量の変化と共変量の値の両方に依存します。

## コンプレックス サンプルの順序回帰の保存

図 11-7  
[順序回帰保存] ダイアログ ボックス



**変数を保存。**このグループでは、モデルで予測されたカテゴリ、予測カテゴリの確率、観測カテゴリの確率、累積確率、および予測確率を作業データファイルの新変数として保存できます。

**モデルを SPSS Statistics データとしてエクスポート。**パラメータ相関行列またはパラメータ共分散行列から構成される IBM® SPSS® Statistics データセットを出力します。これらの行列には、パラメータ推定値、標準誤差、有意確率、および自由度が含まれています。行列ファイルの変数の順序は次のとおりです。

- **rowtype\_。**COV (共分散)、CORR (相関)、EST (パラメータ推定値)、SE (標準誤差)、SIG (有意水準)、および DF (抽出計画の自由度) の値 (および値ラベル) を取ります。各モデル パラメータに行型 COV (または CORR) を持つ個別のケースだけでなく、各モデル パラメータにその他の行型を持つ個別のケースがあります。

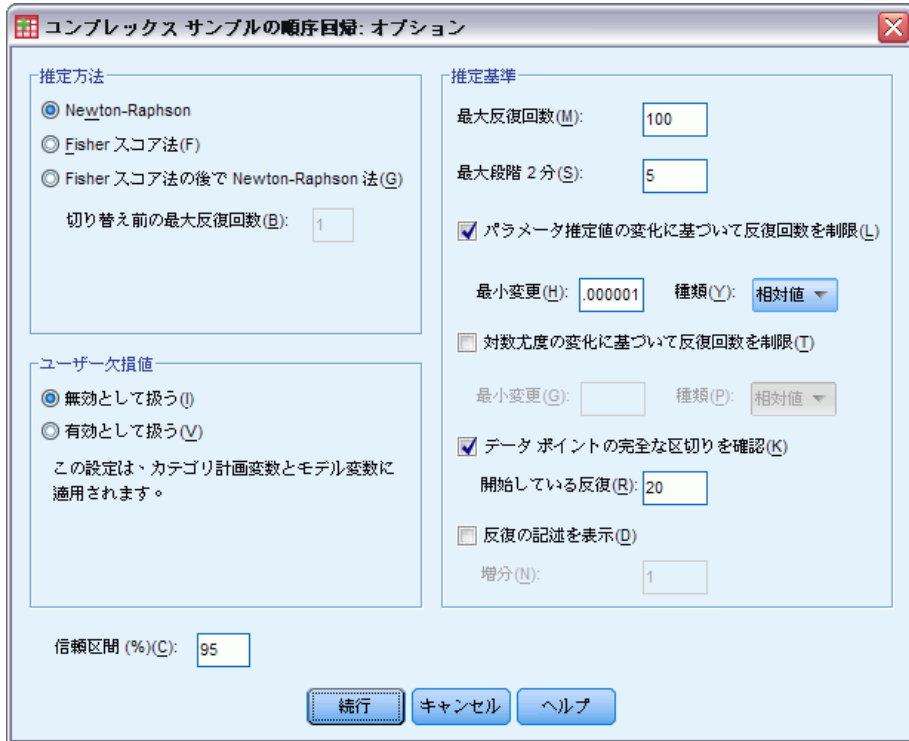
- **varname\_.** すべてのモデル パラメータの順序づけられたリストに応じて、行型 COV または CORR に対し値 P1、P2、... を取り、パラメータ推定値テーブルに表示されたパラメータ文字列に従って値ラベルを付けます。その他の行型については、セルが空白になります。
- **P1、P2、...** これらの変数は、すべてのモデル パラメータの順序づけられたリストに対応しており、パラメータ推定テーブルに表示されたパラメータ文字列に従った変数ラベルを持ち、行型に従って値を取ります。冗長パラメータの場合、すべての共分散は 0 に設定され、相関はシステム欠損値に設定されます。また、すべてのパラメータ推定値は 0 に設定され、すべての標準誤差、有意水準、残差自由度はシステム欠損値に設定されます。

注: 行列ファイルを読み込むその他の手続きでは、ここでエクスポートされたすべての行型が承認されない場合、このファイルをすぐに使用してより詳細な分析を行うことはできません。

**モデルを XML としてエクスポート。** 選択した場合、パラメータ推定値とパラメータ分散共分散行列を XML (PMML) 形式で保存します。このモデルファイルを使用して、得点付けのために他のデータ ファイルにモデル情報を適用できます。

## コンプレックス サンプルの順序回帰のオプション

図 11-8  
[順序回帰: オプション] ダイアログ ボックス



**推定方法。** パラメータ推定の方法は選択することができます。Newton-Raphson 法、Fisher スコア法、および Fisher スコア法を何回か反復実行した後 Newton-Raphson 法に切り替える HYBRID 法の中からいずれかを選択できます。HYBRID 法における Fisher スコア法を実行中、その最大反復回数に到達する前に収束した場合も、Newton-Raphson 法のアルゴリズムは続行されます。

**推定。** このグループでは、モデルの推定で使用するさまざまな基準を制御します。

- **最大反復回数。** アルゴリズムで実行される反復の最大回数です。負でない整数を指定してください。
- **最大段階 2 分。** 対数尤度が増加するか、最大段階 2 分に達するまで、反復ごとにステップ サイズが 0.5 の因子ずつ減少します。正の整数を指定します。
- **パラメータ推定値の変化に基づいて反復回数を制限。** 選択すると、パラメータ推定値の絶対変化または相対変化が指定した値よりも小さい反復の後にアルゴリズムが停止します。指定する値は負でない必要があります。

- **対数尤度の変化に基づいて反復回数を制限。**選択した場合、対数尤度関数の絶対変化または相対変化が指定した値よりも小さい反復の後にアルゴリズムが停止します。
- **データポイントの完全な区切りを確認。**選択した場合、パラメータ推定値が一意的な値になるようにアルゴリズムによって検定が行われます。区切りは、この手続きによって各ケースが正しく分類されるモデルを作成できる場合に存在します。
- **反復の記述を表示。**パラメータ推定値、および 0 回目（初期推定値）から始まる反復の n 回ごとの統計量を表示します。反復の記述の出力を選択した場合、n の値にかかわらず、最後の反復は常に出力されます。

**ユーザー欠損値。**従属変数や共変量だけではなく、計画変数でデータが有効である必要があります。これらの変数のデータが無効なケースは分析から除外されます。この項目を使用して、ユーザー欠損値をストラータ変数、クラスタ変数、部分母集団変数、および因子変数で有効な値として扱うかどうかを決定できます。

**信頼区間。**これは、係数の推定値、指数化された係数の推定値、およびオッズ比の信頼区間のレベルです。50 以上、100 未満の値を指定します。

## CSORDINAL コマンドの追加機能

コマンド シンタックス言語を使用して、次のことも実行できます。

- 線型に組み合わせた効果や値に対し、ユーザー指定による効果の検定を指定 (CUSTOM サブコマンドを使用)。
- 因子と共変量の累積オッズ比の計算時に、平均値以外の値における他のモデル変数の値を固定 (ODDSRATIOS サブコマンドを使用)。
- オッズ比が必要な場合に、因子のユーザー指定の参照カテゴリとしてラベルのない値を使用 (ODDSRATIOS サブコマンドを使用)。
- 特異性の点検で使用する許容値を指定 (CRITERIA サブコマンドを使用)。
- 一般推定可能関数を作成 (PRINT サブコマンドを使用)。
- 25 個以上の確率変数を保存 (SAVE サブコマンドを使用)。

複雑なシンタックス情報については、「コマンド シンタックス リファレンス」を参照してください。

# コンプレックス サンプル Cox 回帰

コンプレックス サンプルの Cox 回帰手続きでは、コンプレックス サンプル法によって抽出されたサンプルに対して、生存分析を実行します。オプションとして、部分母集団の分析も実行できます。

**例。**政府の法執行機関は、管轄地域での累犯率に注意を払っています。累犯率の測定方法の 1 つは、犯罪者の 2 回目の逮捕までの時間です。法執行機関は、Cox 回帰を使用して再逮捕までの時間をモデル化したいと考えていますが、年齢カテゴリ全体でハザードの比例仮定が適用できないことを心配しています。

医療研究者は、虚血性脳卒中のリハビリ プログラムを終えた患者の生存時間を調査しています。死亡以外の重大な事象の発生とこれらの事象の時間が記録されるにつれて患者の病歴が変わるため、被験者ごとに複数のケースがある可能性があります。また、虚血性脳卒中の発生時がリスクの起点となる一方で、リハビリ プログラム後生存している患者のみがサンプルに含まれるため、観測される生存時間はリハビリの長さによって「膨張している」という意味で、サンプルは左側が切り捨てられます。

**生存時間。**この手続きでは、Cox 回帰を生存時間、すなわちイベントの出現の前の時間の長さの分析に適用します。生存時間を指定する方法は、区間の開始時間に応じて次の 2 通りあります。

- **時間 =0。**一般的に、被験者ごとに区間の開始時について完全な情報があり、終了時間を含む変数があるだけです（または、日時の変数から終了時間の単一の変数を作成します。下の説明を参照してください）。
- **被験者により変化。**これは、左側の切り捨て（遅延した投入とも呼ばれます）を行う場合に適合します。たとえば、脳卒中後のリハビリ プログラムを終えた患者の生存時間を分析する場合、脳卒中の発生時をリスクの起点と見なします。ただし、リハビリ プログラムで生存した患者のみを含むサンプルの場合、観測される生存時間はリハビリの長さによって「膨張している」という意味で、サンプルは左側が切り捨てられます。調査に投入した時間としてリハビリを終えた時間を指定することでこれを説明できます。



**日時の変数。**日時の変数を使用して区間の開始と終了を直接定義することはできません。日時の変数がある場合、日時の変数を使用して生存時間を含む変数を作成します。左側の切り捨てがない場合、調査に投入された日付と観測の日付の差に基づいて、終了時間を含む変数を作成するだけです。左側の切り捨てがある場合、調査の開始日付と投入の日付の差に基づいて開始時間を含む変数を作成し、調査の開始日付と観測の日付の差に基づいて終了時間を含む変数を作成します。

**事象の状態。**被験者が区間内に関心のある事象を経験したかどうかを記録する変数が必要です。関心のある事象が発生していない被験者は、右側が打ち切られます。

**被験者の識別子。**複数のケースから単一の被験者に対する観測を分割することで、区分的に一定な時間依存の予測変数を容易に組み込むことができます。たとえば、脳卒中後の患者の生存時間を分析する場合、病歴を表す変数は予測変数として有用です。時間の経過とともに、患者は病歴が変わる重大な事象を経験する可能性があります。次の表は、そのようなデータセットの構成方法を示します。患者 ID は被験者の識別子です。終了時間は観測された区間を定義します。状態は重大な事象を記録します。心臓発作の前の病歴と 出血の前の病歴は、区分的に一定な時間依存の予測変数です。

患者 ID	終了時間	状態	心臓発作の前の病歴	出血の前の病歴
1	5	心臓発作	No	No
1	7	出血	Yes	No
1	8	死亡	Yes	Yes
2	24	死亡	No	No
3	8	心臓発作	No	No
3	15	死亡	Yes	No

**仮定。**データ ファイルのケースは、[コンプレックス サンプル計画] [ダイアログ ボックス](#) で選択されたファイルの指定に従って分析されるコンプレックス デザインからのサンプルを表します。

通常、Cox 回帰分析モデルではハザードの比例を仮定します。つまり、あるケースと他のケースとのハザードの比率は時間に対して変化してはならないということです。この仮定が適用されない場合、モデルに時間依存の予測変数を追加する必要がある場合があります。

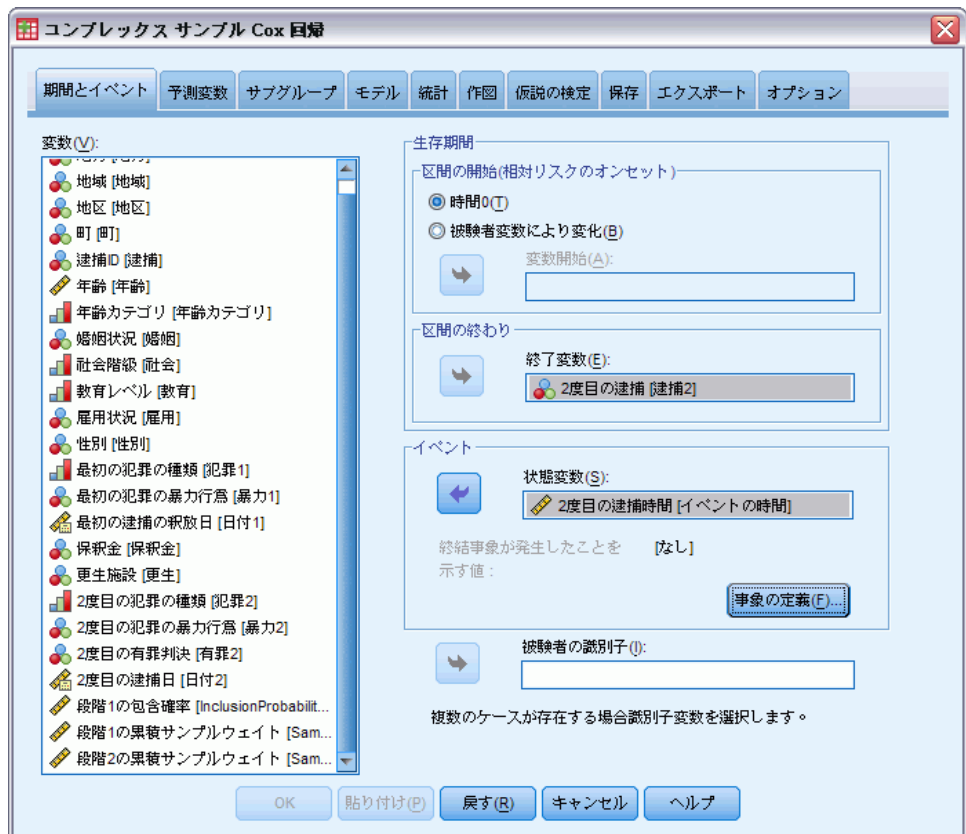
**Kaplan-Meier 分析。**予測変数を選択せず（または、選択した予測変数をモデルに投入しない）、[オプション] タブでベースラインの生存曲線の計算に積の上限の方法を選択する場合、この手続きでは Kaplan-Meier 型の生存分析が実行されます。

## コンプレックス サンプルの Cox 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(A) > コンプレックス サンプル > Cox 回帰...
- ▶ 計画ファイルを選択します。オプションとして、ユーザー指定の結合確率ファイルも選択できます。
- ▶ [続行] をクリックします。

図 12-1

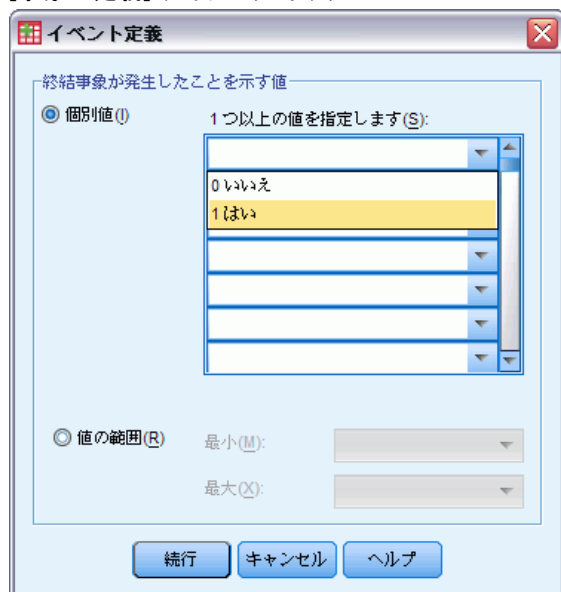
[Cox 回帰分析] ダイアログ ボックスの [時刻およびイベント] タブ



- ▶ 調査から投入時間および終了時間を選択して、生存時間を指定します。
- ▶ 事象の状態の変数を選択します。
- ▶ [事象の定義] をクリックし、少なくとも 1 つの事象の値を定義します。  
オプションとして、被験者の識別子を 1 つ選択できます。

## 事象の定義

図 12-2  
[事象の定義] ダイアログ ボックス

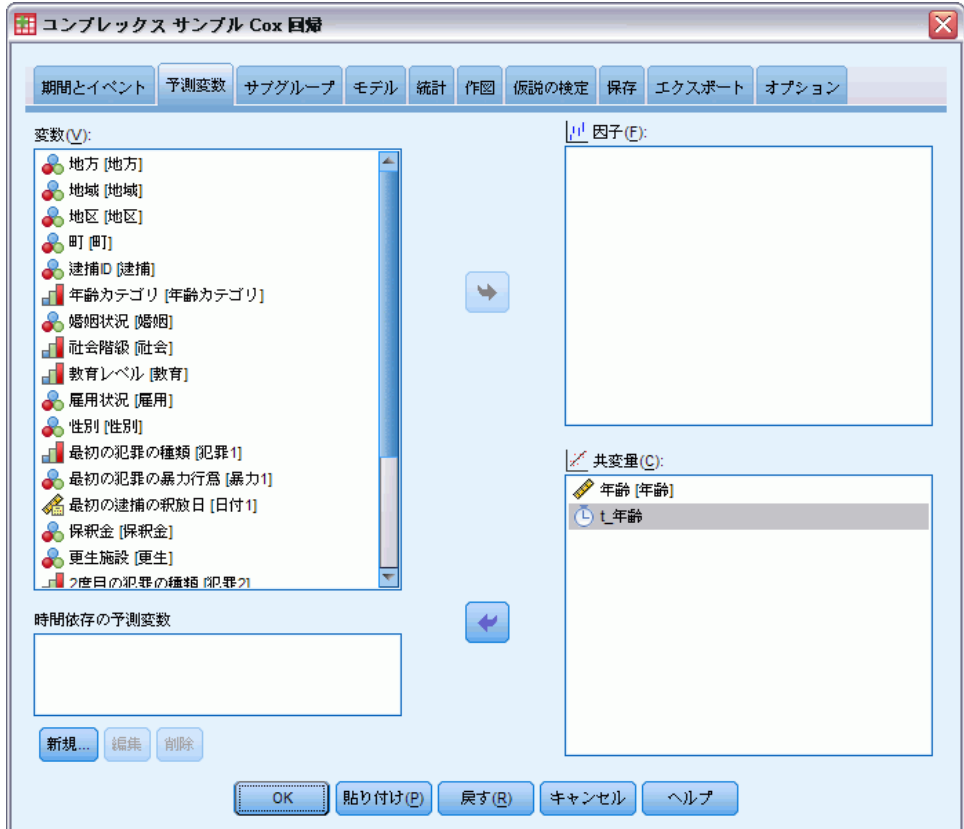


終結事象が発生したことを示す値を指定します。

- **個別値。** グリッドに入力するか、定義済みの値ラベルの値のリストから選択して、1つ以上の値を指定します。
- **値の範囲。** 最小値と最大値を入力するか、定義済みの値ラベルのリストから値を選択して、値の範囲を指定します。

## 予測変数

図 12-3  
[Cox 回帰分析] ダイアログ ボックスの [予測変数] タブ



[予測変数] タブでは、モデル効果の構成に使用する因子および共変量を指定できます。

**因子。** 因子はカテゴリ予測変数です。因子には数値か文字を指定できます。

**共変量。** 共変量は尺度予測変数で、数値を指定することが必要です。

**時間依存の予測変数。** ハザードの比例仮定が適用できない状況があります。具体的には、ハザード率が時間によって変化し、(1つ以上の) 予測変数の値が時点によって異なるような場合です。そのような場合には、時間依存の予測変数を指定する必要があります。詳細は、p. 87 [時間依存の予測変数の定義](#) を参照してください。時間依存の予測変数は、因子または共変量として選択することができます。

## 時間依存の予測変数の定義

図 12-4  
[Cox サンプル回帰分析: 時間依存の予測変数を定義] ダイアログ ボックス



[時間依存の共変量を定義] ダイアログ ボックスでは、組み込み時間変数  $T_$  に依存する予測変数を作成できます。この変数を使うと、次のように、一般的な 2 通りの方法で時間依存の共変量を定義できます。

- 比例しないハザードを可能にする拡張 Cox 回帰分析モデルを推定したい場合は、時間変数  $T_$  と対象の共変量の関数として時間依存の予測変数を定義すれば、その作業を実行できます。時間変数と予測変数の単純な積などが一般的な例と言えますが、同様に、複雑な関数も指定できます。
- 変数の中には、時間によって値が異なりますが、系統だった時間との関連性がないものもあります。このような場合、**セグメント化した時間依存の予測変数**を定義する必要があります。これは、論理式を使って実行できます。この論理式では、真であれば 1、偽であれば 0 という値を取ります。一連の論理式を使うと、1 組の測定から時間依存の予測変数を作成できます。たとえば、研究のため、週に 1 度の血圧測定を 4 週間続ける場合は (BP1 から BP4 という名前で識別)、時間依存の予測変数を  $(T_ < 1) * BP1 + (T_ \geq 1 \ \& \ T_ < 2) * BP2 + (T_ \leq 2 \ \& \ T_ < 3) * BP3 +$

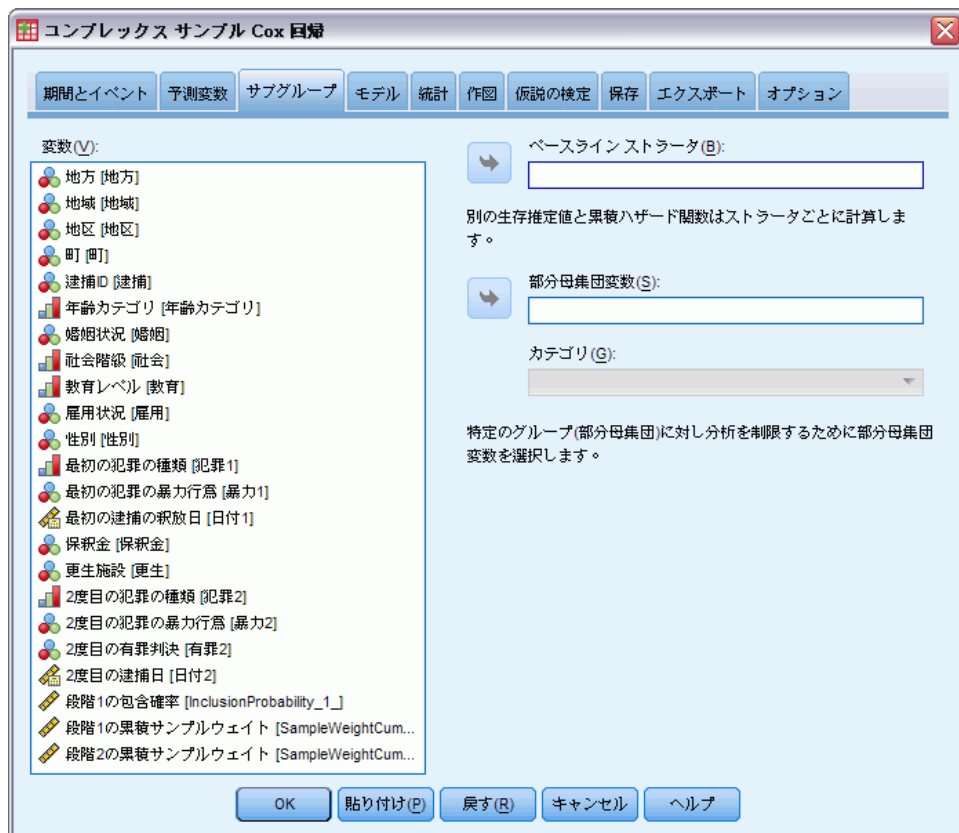
$(T_1 \geq 3 \ \& \ T_1 < 4) * BP4$  と定義できます。指定されたケースおよび残りはすべて 0 になるため、かっこ内の 1 つの項目は 1 となります。つまり、時間が 1 週間より短い場合、BP1 を使用し、1 週間より長く 2 週間より短い場合は BP2 となり、後に続きます。

注：セグメント化した時間依存の予測変数がセグメント内で一定の場合（上記の血圧の例など）、複数のケースに被験者を分割すると、区分的に一定な時間依存の予測変数を指定するのが簡単になる場合があります。詳細については、[p. 82 コンプレックス サンプル Cox 回帰](#)の被験者の識別子についての説明を参照してください。

[時間依存の予測変数を定義] ダイアログ ボックスで、関数作成用のコントロールを使うことにより、時間依存の共変量の式を作成できますが、[数式] テキスト領域に直接入力することもできます。文字型定数はダブルクォーテーションかアポストロフィで囲み、数値定数の場合は小数部を点で区切って米国方式で入力しなければならないで、注意してください。この結果得られる変数は、指定した名前が付けられ、[予測変数] タブに因子または共変量として含まれます。

## サブグループ

図 12-5  
[Cox 回帰分析] ダイアログ ボックスの [サブグループ] タブ



**ベースライン ストラータ。** 個別のベースラインのハザード関数と生存関数がこの変数の値ごとに計算されます。また、モデル係数の単一集合はストラータ間で推定されます。

**部分母集団変数。** 変数を 1 つ指定して部分母集団を定義します。この分析は、部分母集団変数のカテゴリのうち、選択したカテゴリに対してのみ行われます。

## モデル

図 12-6  
[Cox 回帰分析] ダイアログ ボックスの [モデル] タブ



**モデル効果を指定。**デフォルトでは、この手続きで、メイン ダイアログ ボックスで指定した因子と共変量を使用して主効果モデルが作成されます。または、交互作用効果と入れ子項目を含む、ユーザーの指定によるモデルを作成できます。

### 非入れ子項目

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。

**主効果。**選択した変数のそれぞれに主効果の項目を作成します。

**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。

### 入れ子項目

この手続きでは、モデルに入れ子の項を構築できます。入れ子項目を構築すると、別の因子のレベルと相互作用しない因子または共変量の効果をモデリングできます。たとえば、ある食料雑貨店チェーンが、いくつかの店舗で顧客の消費傾向を調査すると仮定します。1 人の顧客が頻繁に通うのはこれらの店舗の 1 つだけであるため、「顧客」の効果は、「店舗の場所」の効果の**入れ子**になっていると言えます。

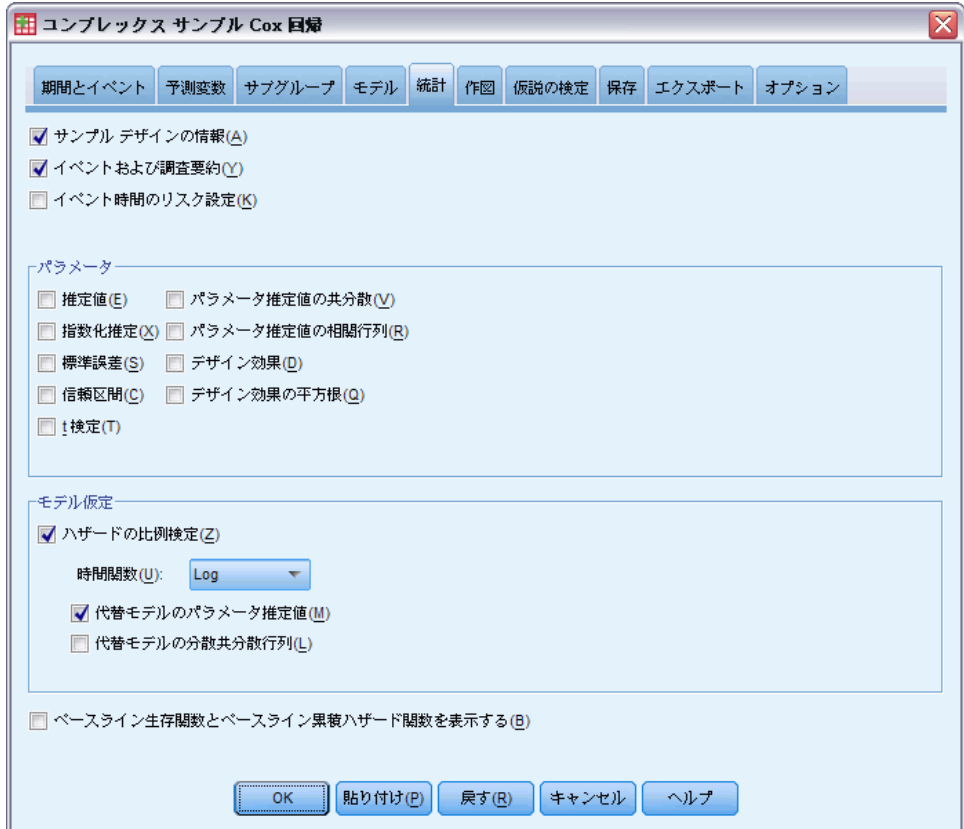
また、入れ子の項には、同一の共変量に関する多項式項などの交互作用効果を含めたり、複数レベルで入れ子になった項を追加したりできます。

**制限。** 入れ子項目には、次の制限があります。

- 1 つの交互作用内のすべての因子は一意でなければなりません。したがって、A が因子である場合、A\*A の指定は無効です。
- 1 つの入れ子効果内のすべての因子は一意でなければなりません。したがって、A が因子である場合、A(A) の指定は無効です。
- 効果は共変量の入れ子にできません。したがって、A が因子で X が共変量である場合、A(X) の指定は無効です。

## 統計量

図 12-7  
[Cox 回帰分析] ダイアログ ボックスの [統計] タブ



**サンプル デザインの情報。**重み付けのない度数、母集団のサイズを含め、サンプルに関する要約情報を表示します。

**事象と打ち切りの要約。**打ち切られたケースの数と割合についての集計情報が表示されます。

**事象の時点のリスクセット。**各ベースライン ストラータ内の事象の時間ごとに事象の数と危険の数を表示します。

**パラメータ。**このグループでは、モデルのパラメータに関連する統計量の表示を制御できます。

- **推定値。**係数の推定値を表示します。
- **指数化推定。**係数の推定値でべき乗した自然対数の底を表示します。推定値のプロパティが統計上の検定に適している場合、指数化推定、つまり  $\exp(B)$  の解釈は比較的簡単です。
- **標準誤差。**係数の推定値ごとの標準誤差を表示します。

- **信頼区間。**係数の推定値ごとの信頼区間を表示します。各区間の信頼水準は、[オプション] ダイアログ ボックスで設定します。
- **t 検定。**係数の推定値ごとの t 検定を表示します。検定ごとの帰無仮説は、係数の値が 0 の場合です。
- **パラメータ推定値の共分散。**モデル係数の共分散行列の推定値を表示します。
- **パラメータ推定値の相関。**モデル係数の相関行列の推定値を表示します。
- **デザイン効果。**サンプルが単純な無作為サンプルであると仮定することで得られる分散と、推定値の分散との比。これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。
- **デザイン効果の平方根。**これはコンプレックス デザインを指定する効果の程度です。値が 1 から遠ざかるほど大きな効果が現れます。

**モデル仮定。**このグループにより、ハザードの比例仮定の検定を生成できます。検定では、各予測変数  $x$  について、適合モデルが時間依存の予測変数  $x*_TF$  を含む代替モデルと比較されます。ここで、 $_TF$  は指定された時刻関数です。

- **時刻関数。**代替モデルについて  $_TF$  の形式を指定します。同一関数の場合、 $_TF=T_$  です。対数関数の場合、 $_TF=\log(T_)$  です。Kaplan-Meier の場合、 $_TF=1 -S_{KM}(T_)$  です。ここで、 $S_{KM}(\cdot)$  は生存関数の Kaplan-Meier 推定です。順位の場合、 $_TF$  は観測された終了時間の中の  $T_$  の順位です。
- **代替モデルのパラメータ推定値。**代替モデルのパラメータごとに、推定値、標準誤差、および信頼区間が表示されます。
- **代替モデルの分散共分散行列。**代替モデルのパラメータ間の推定共分散の行列が表示されます。

**ベースラインの生存関数および累積ハザード関数。**ベースラインの生存関数および累積ハザード関数を、標準誤差とともに表示します。

注：[予測変数] タブで定義された時間依存の予測変数がモデルに含まれる場合、このオプションは使用できません。

## 作図

図 12-8  
[Cox 回帰分析] ダイアログ ボックスの [作図] タブ



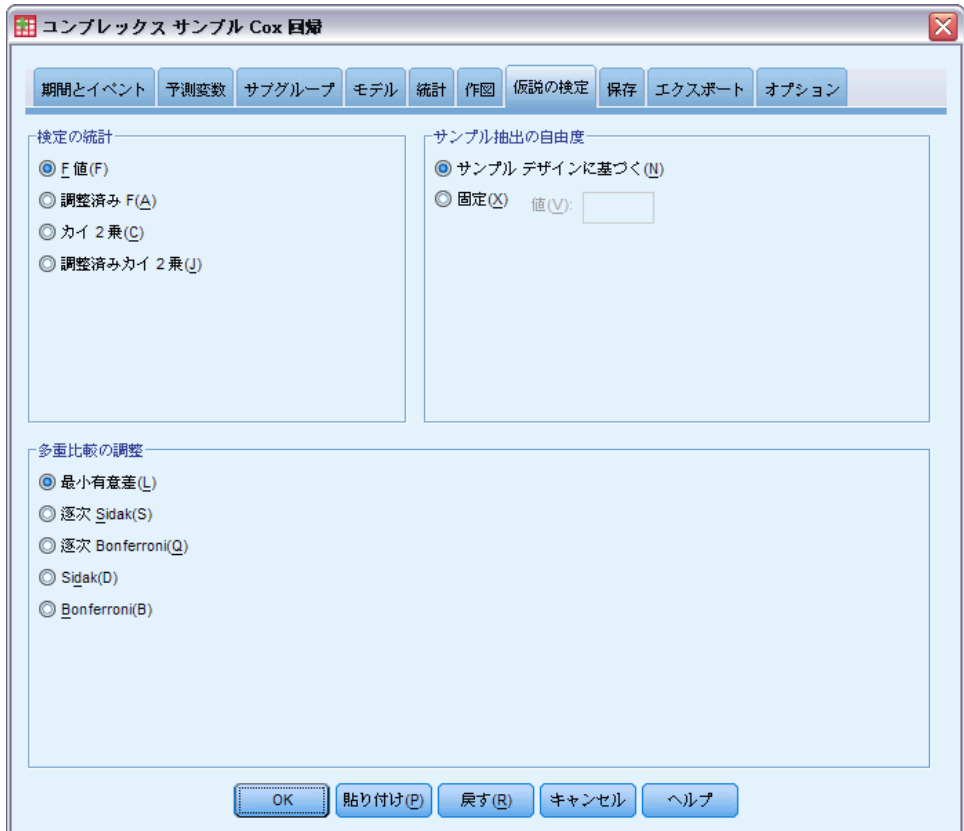
[作図] タブでは、ハザード関数、生存関数、生存関数のログマイナスログ、および1マイナス累積生存関数の作図を要求できます。また、指定した関数に沿って信頼区間を作図することもできます。信頼水準は [オプション] タブで設定します。

**予測変数のパターン。** 要求される作図および [エクスポート] タブでエクスポートされる生存ファイルに使用される予測変数の値のパターンを指定できます。[予測変数] タブで定義された時間依存の予測変数がモデルに含まれる場合、これらのオプションは使用できません。

- **因子の作図。** デフォルトでは、各因子は最大レベルで評価されます。必要に応じて、異なるレベルを入力または選択します。または、因子のチェック ボックスを選択することによって、単一の因子のレベルごとに個別の線を作図することもできます。
- **共変量の作図。** 各共変量は平均値で評価されます。必要に応じて、異なる値を入力または選択します。

## 仮説の検定

図 12-9  
[Cox 回帰分析] ダイアログ ボックスの [仮説の検定] タブ



**検定の統計。**このグループでは、仮説の検定に使用する統計量の種類を選択できます。[F 値]、[調整済み F]、[カイ 2 乗]、[調整済みカイ 2 乗]の中から選択します。

**サンプル抽出の自由度。**このグループでは、すべての検定統計量の p 値の計算に使用する抽出計画の自由度を制御できます。抽出計画に基づく場合は、主要な抽出単位数と抽出の第 1 段階のストラータ数とで値が異なります。また、正の整数を指定することにより自由度を設定することもできます。

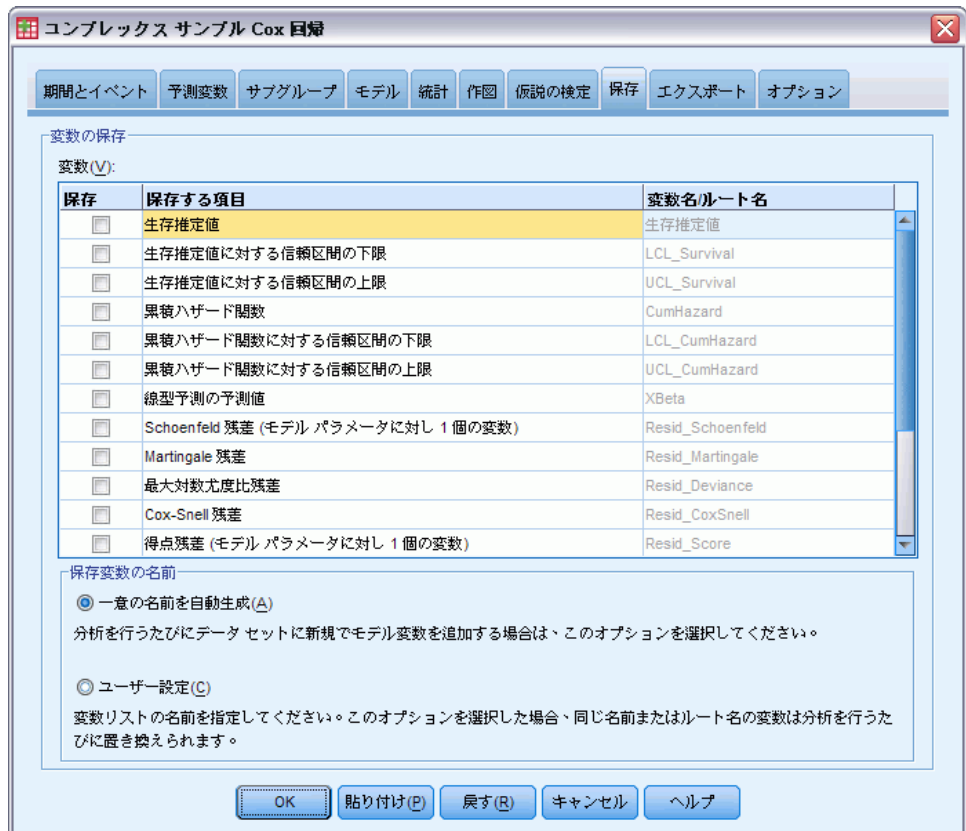
**多重比較の調整。**多重対比を用いた仮説検定を行う場合、各対比に関する有意水準を基に、全体の有意水準を調整できます。このグループでは、調整方法を選択できます。

- **最小有意差。**この方法では、帰無仮説の値と異なる線型の対比が存在するという仮説を棄却する確率の全体的な制御を行いません。

- **逐次 Sidak.** これは、すべて同じ有意水準を維持したままで個々の仮説が棄却される保守的でない逐次ステップダウン相対 Sidak プロシジャです。
- **逐次 Bonferroni.** 個々の仮説を棄却しますが、すべて同じ有意水準を維持することでは、それほど控えめでもない逐次ステップダウン棄却 Bonferroni 手続きです。
- **Sidak.** この方法は Bonferroni の方法よりもっと狭い限界を提供します。
- **Bonferroni の方法.** この方法は、多重比較の検定がなされるときに、有意確率を修正します。

## 保存

図 12-10  
[Cox 回帰分析] ダイアログ ボックスの [保存] タブ



**変数を保存。** このグループでは、モデル関連変数をアクティブなデータセットに保存して、結果の診断および報告で使用できます。時間依存の予測変数がモデルに含まれる場合、これらは使用できません。

- **生存関数。**ケースごとに、観測された時間および予測変数の値での生存の確率（生存関数の値）を保存します。
- **生存関数に対する信頼区間の下限。**ケースごとに、観測された時間および予測変数の値での生存関数に対する信頼区間の下限を保存します。
- **生存関数に対する信頼区間の上限。**ケースごとに、観測された時間および予測変数の値での生存関数に対する信頼区間の上限を保存します。
- **累積ハザード関数。**ケースごとに、観測された時間および予測変数の値での累積ハザード ( $-\ln(\text{生存})$ ) を保存します。
- **累積ハザード関数に対する信頼区間の下限。**ケースごとに、観測された時間および予測変数の値での累積ハザード関数に対する信頼区間の下限を保存します。
- **累積ハザード関数に対する信頼区間の上限。**ケースごとに、観測された時間および予測変数の値での累積ハザード関数に対する信頼区間の上限を保存します。
- **線型予測の予測値。**修正された予測変数に回帰係数を掛けた参照値の線型結合を保存します。線型予測変数は、ベースラインハザードに対するハザード関数の比率です。ハザードの比例モデルでは、この値はすべての時間で一定です。
- **Schoenfeld 残差。**モデル内の打ち切られていない各ケースおよび各非冗長パラメータについて、Schoenfeld 残差は、モデルパラメータに関連付けられた予測変数の観測値と、観測された事象の時間のリスクセット内のケースに対する予測変数の期待値の差です。Schoenfeld 残差は、ハザードの比例仮定の評価に使用できます。たとえば、予測変数  $x$  について、時間依存の予測変数  $x \cdot \ln(T)$  についての Schoenfeld 残差と時間を作図すると、ハザードの比例が適用される場合、0 の水平線が表示されます。モデル内の非冗長パラメータごとに個別の変数が保存されます。Schoenfeld 残差は、打ち切られていないケースについてのみ計算されます。
- **マーチンゲール残差。**各ケースについて、マーチンゲール残差は、観測された打ち切り（打ち切られた場合 0、それ以外の場合 1）と観測時間中の事象の期待値の差です。
- **最大対数尤度比残差。**最大対数尤度比残差は、「調整済み」マーチンゲール残差で、0 についてより対称になります。予測に対する最大対数尤度比残差の作図から、パターンはわかりません。
- **Cox-Snell 残差。**各ケースについて、Cox-Snell 残差は観測時間中の事象の期待値です。つまり、観測された打ち切りからマーチンゲール残差を引いた値です。
- **スコア残差。**モデル内の各ケースおよび各非冗長パラメータについて、スコア残差は擬似尤度の最初の微分係数に対するケースの寄与率です。モデル内の非冗長パラメータごとに個別の変数が保存されます。

- **DFBeta 残差**。モデル内の各ケースおよび各非冗長パラメータについて、DFBeta 残差は、ケースがモデルから除外された場合のパラメータ推定値の値の変化の近似です。比較的大きい DFBeta 残差があるケースでは、分析の影響力が取り消される場合があります。モデル内の非冗長パラメータごとに個別の変数が保存されます。
- **集計残差**。複数のケースが単一の被験者を表す場合、被験者の集計残差は、同じ被験者に属するすべてのケースの対応するケースの残差の単純な合計です。Schoenfeld の残差の場合、Schoenfeld の残差は打ち切られていないケースについてのみ定義されるため、集計された残差は、集計されていない残差と同じです。これらの残差は、[時刻およびイベント] タブで被験者の識別子が指定されている場合のみ使用できます。

**保存する変数の名前** 自動的な名前の生成によって、すべての作業を保存することができます。ユーザー指定の名前によって、Data Editor で保存された変数を最初に削除することなく、前回実行された結果を破棄または置き換えることができます。

## エクスポート

図 12-11  
[Cox 回帰分析] ダイアログ ボックスの [エクスポート] タブ





**モデルを SPSS Statistics データとしてエクスポート。** パラメータ相関行列またはパラメータ共分散行列から構成される IBM® SPSS® Statistics データセットを出力します。これらの行列には、パラメータ推定値、標準誤差、有意確率、および自由度が含まれています。行列ファイルの変数の順序は次のとおりです。

- **rowtype\_。** COV (共分散)、CORR (相関)、EST (パラメータ推定値)、SE (標準誤差)、SIG (有意水準)、および DF (抽出計画の自由度) の値 (および値ラベル) を取ります。各モデル パラメータに行型 COV (または CORR) を持つ個別のケースだけでなく、各モデル パラメータにその他の行型を持つ個別のケースがあります。
- **varname\_。** すべてのモデル パラメータの順序づけられたリストに応じて、行型 COV または CORR に対し値 P1、P2、... を取り、パラメータ推定値テーブルに表示されたパラメータ文字列に従って値ラベルを付けます。その他の行型については、セルが空白になります。
- **P1、P2、...** これらの変数は、すべてのモデル パラメータの順序づけられたリストに対応しており、パラメータ推定テーブルに表示されたパラメータ文字列に従った変数ラベルを持ち、行型に従って値を取ります。冗長パラメータの場合、すべての共分散は 0 に設定され、相関はシステム欠損値に設定されます。また、すべてのパラメータ推定値は 0 に設定され、すべての標準誤差、有意水準、残差自由度はシステム欠損値に設定されます。

注: 行列ファイルを読み込むその他の手続きでは、ここでエクスポートされたすべての行型が承認されない場合、このファイルをすぐに使用してより詳細な分析を行うことはできません。

**SPSS Statistics データとして生存関数をエクスポート。** 失敗または事象の時間ごとに、[作図] タブで指定されたベースラインおよび予測変数のパターンで評価された生存関数、生存関数の標準誤差、生存関数の信頼区間の上限と下限、および累積ハザード関数を含んだ SPSS Statistics 形式のデータセットを出力します。行列ファイルの変数の順序は次のとおりです。

- **ベースライン ストラータ変数。** ストラータ変数の値ごとに個別の生命表が作成されます。
- **生存時間変数。** 事象の時間。異なる事象の時間ごとに個別のケースが作成されます。
- **Sur\_0, LCL\_Sur\_0, UCL\_Sur\_0。** ベースラインの生存関数および生存関数の信頼区間の上限と下限。
- **Sur\_R, LCL\_Sur\_R, UCL\_Sur\_R。** 「参照」パターンで評価された生存関数 (出力のパターン値の表を参照) および生存関数の信頼区間の上限と下限。
- **Sur\_##, LCL\_Sur\_##, UCL\_Sur\_##。** [作図] タブで指定された各予測変数のパターンで評価された生存関数および生存関数の信頼区間の上限と下限。パターンを数 ## と一致させるには、出力のパターン値の表を参照してください。

- **Haz\_0, LCL\_Haz\_0, UCL\_Haz\_0。** ベースラインの累積ハザード関数および累積ハザード関数の信頼区間の上限と下限。
- **Haz\_R, LCL\_Haz\_R, UCL\_Haz\_R。** 「参照」パターンで評価された累積ハザード関数（出力のパターン値の表を参照）および累積ハザード関数の信頼区間の上限と下限。
- **Haz\_##, LCL\_Haz\_##, UCL\_Haz\_##。** [作図] タブで指定された各予測変数のパターンで評価された累積ハザード関数および累積ハザード関数の信頼区間の上限と下限。パターンを数 ## と一致させるには、出力のパターン値の表を参照してください。

**モデルを XML としてエクスポート。** パラメータ推定値およびベースラインの生存関数を含む、生存関数の予測に必要なすべての情報を XML (PMML) 形式で保存します。このモデルファイルを使用して、得点付けのために他のデータファイルにモデル情報を適用できます。

## オプション

図 12-12  
[Cox 回帰分析] ダイアログ ボックスの [オプション] タブ

The screenshot shows the 'オプション' (Options) tab of the 'コンプレックス サンプル Cox 回帰' dialog box. The dialog is divided into several sections:

- 推定 (Estimation):**
  - 最大反復回数 (M): 100
  - 最大段階 2分 (S): 5
  - パラメータ推定値の変化に基づいて反復回数を制限 (L)
  - 最小変更 (H): 0.000001 種類 (Y): 相対値
  - 対数尤度の変化に基づいて反復回数を制限 (I)
  - 最小変更 (G): 種類 (P): 相対値
  - 反復の記述を表示 (O)
  - 増分 (N): 1
  - パラメータ推定の均衡を破る方法:
    - Efron
    - Breslow
  - 信頼区間 (%) (E): 95
- 生存関数 (Survival Function):**
  - 生存関数ベースラインの推定方法:
    - Efron法
    - Breslow法
    - 積極限法 (C)
  - 生存関数の信頼区間:
    - 変換した生存関数に基づいて計算し、元の単位に戻す (K)
    - 変換 (A): Log
    - 生存関数の元の単位に基づいて計算する (U)
- ユーザー欠損値 (User Missing Values):**
  - 無効として扱う (I)
  - 有効として扱う (V)
  - この設定は、カテゴリ計画変数とモデル変数に適用されます。

At the bottom, there are buttons for 'OK', '貼り付け (P)', '戻す (R)', 'キャンセル', and 'ヘルプ'.

**推定。**これらのコントロールでは、回帰係数の推定条件を指定します。

- **最大反復回数。**アルゴリズムで実行される反復の最大回数です。負でない整数を指定してください。
- **最大段階 2 分。**対数尤度が増加するか、最大段階 2 分に達するまで、反復ごとにステップ サイズが 0.5 の因子ずつ減少します。正の整数を指定します。
- **パラメータ推定値の変化に基づいて反復回数を制限。**選択すると、パラメータ推定値の絶対変化または相対変化が指定した値よりも小さくなる反復が発生した後に、アルゴリズムが停止します。指定する値は正の数であることが必要です。
- **対数尤度の変化に基づいて反復回数を制限。**選択した場合、対数尤度関数の絶対変化または相対変化が指定した値よりも小さくなる反復が発生した後に、アルゴリズムが停止します。指定する値は正の数であることが必要です。
- **反復の記述を表示。**パラメータ推定値および擬対数尤度に関する反復の記述が表示されます。また、パラメータ推定値および擬対数尤度の変化の最新の評価が出力されます。[反復の記述] 表には、0 回目（初期推定値）から数えて n 回のすべての反復が表示されます。ただし n は増分を示す値です。反復の記述を指定した場合、n の値にかかわらず、最後の反復は常に出力されます。
- **パラメータ推定の同順位のブレイク方法。**同一の失敗の時間が観測された場合、いずれかの方法を使用して同順位をブレイクします。Efron の方法は演算により多くの時間がかかります。

**生存関数。**これらのコントロールでは、生存関数に関する計算の条件を指定します。

- **ベースラインの生存関数の推定方法。**Breslow（または Nelson-Aalan もしくは経験的）の方法では、観測された失敗の時間に階段がある非減少の階段関数によってベースラインの累積ハザードを推定して、 $\text{生存} = \exp(-\text{累積ハザード})$  という関係によってベースラインの生存を計算します。Efron の方法は演算により多くの時間がかかり、同順位がない場合は Breslow の方法になります。積の上限の方法では、非増加の右側連続関数によってベースラインの生存を推定します。モデルに予測変数がない場合、この方法は Kaplan-Meier の推定になります。
- **生存関数の信頼区間。**信頼区間は、元の単位を使用、対数変換、ログマイナスログ変換の 3 つの方法で計算できます。信頼区間の範囲が 0 から 1 の間になることが保証されるのはログマイナスログ変換のみですが、一般的に対数変換が「最適」な実行であるように見えます。

**ユーザー欠損値。**すべての変数は、分析の対象となるケースに対して有効な値を取る必要があります。この項目を使用して、ユーザー欠損値をカテゴリモデル（因子、事象、ストラタ、および部分母集団の変数を含む）および抽出計画変数で有効な値として扱うかどうかを決定できます。

**信頼区間 (%)**。これは、係数の推定値、指数化された係数の推定値、生存関数の推定値、および累積ハザード関数の推定値に使用される信頼区間のレベルです。0 以上、100 未満の値を指定します。

## CSCOXREG コマンドの追加機能

コマンド言語を使用すると、次の作業も実行できます。

- ユーザー指定の仮説の検定の実行 (CUSTOM サブコマンドおよび /PRINT LMATRIX を使用)。
- 許容度の指定 (/CRITERIA SINGULAR を使用)。
- 一般推定可能関数 (/PRINT GEF を使用)。
- 複数の予測変数のパターン (複数の PATTERN サブコマンドを使用)。
- ルート名指定時の保存された変数の最大数 (SAVE サブコマンドを使用)。ダイアログでは CSCOXREG のデフォルトの 25 変数が使用されます。

複雑なシンタックス情報については、「コマンド シンタックス リファレンス」を参照してください。

# パート II: 例

# コンプレックス サンプルのサンプリング ウィザード

サンプリング ウィザードでは、抽出計画ファイルの作成や修正、実行のステップを、順を追って行います。ウィザードを使用する前に、対象母集団を適切に定義しておき、抽出単位の一覧を準備し、適切なサンプル計画を検討しておく必要があります。

## 完全サンプリング枠からのサンプル抽出

州当局は、各郡の財産税が公正であることを確認する責任があります。税は財産の評価額に基づいているため、財産のサンプルを郡ごとに調査し、各郡のレコードが同等の最新状態であることを確認する必要があります。ただし、現在の評価を取得するためのリソースは限られているため、利用可能なリソースを幅広く活用することが重要です。州当局は、コンプレックス サンプリング法を使って財産のサンプルを選択することにしました。

財産の一覧は、property\_assess\_cs.sav に収集されています。詳細は、A 付録 p. 281 [サンプル ファイル](#) を参照してください。コンプレックス サンプルのサンプリング ウィザードを使用して、サンプルを選択します。

## ウィザードの使用方法

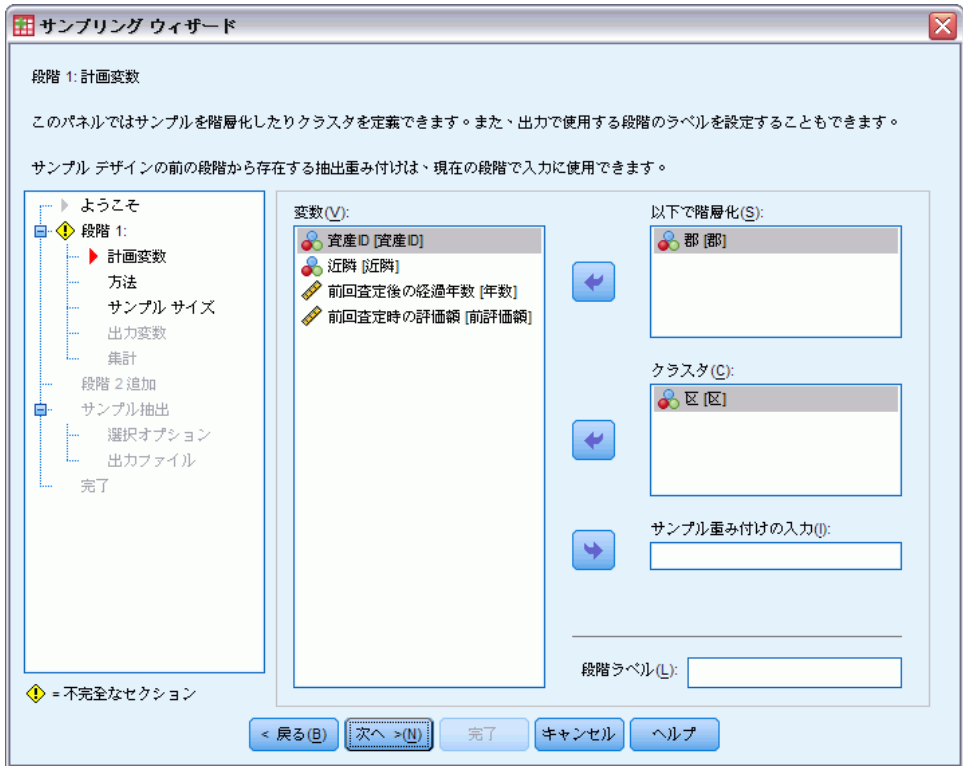
- ▶ コンプレックス サンプルのサンプリング ウィザードを実行するには、メニューから次の項目を選択します。  
分析(A) > コンプレックス サンプル > サンプルの選択...

図 13-1  
サンプリング ウィザードの [ようこそ] ステップ



- ▶ [サンプルを計画する] を選択し、ファイルの保存場所を参照して、計画ファイル名として「property\_assess.csplan」と入力します。
- ▶ [次へ] をクリックします。

図 13-2  
サンプリング ウィザードの [計画変数] ステップ (第 1 段階)

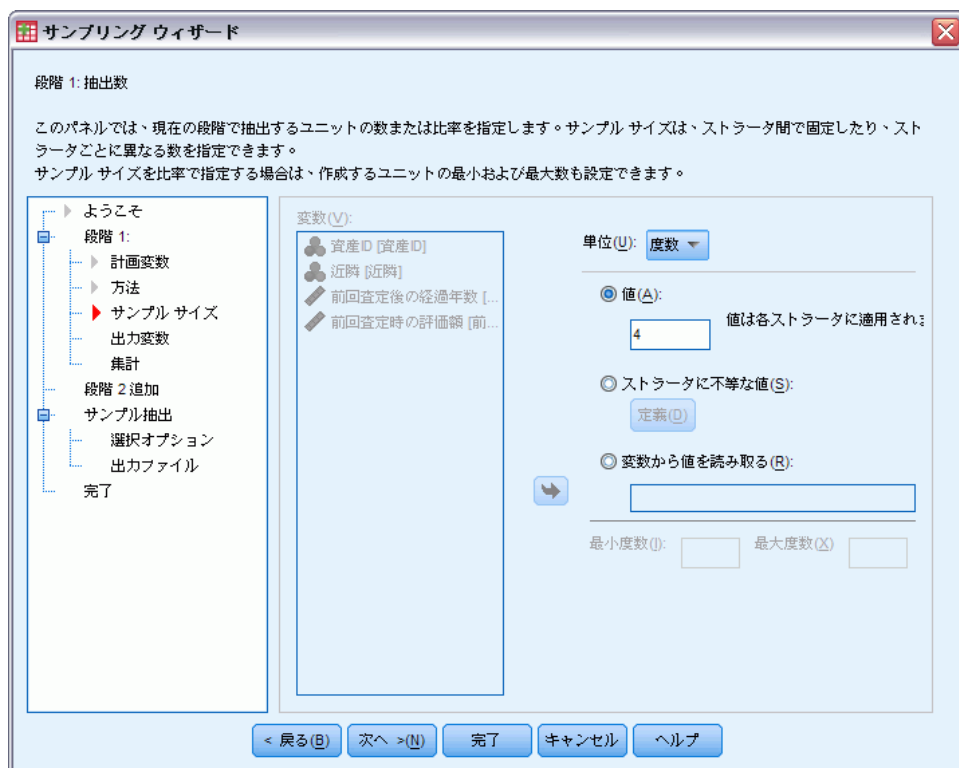


- ▶ 「郡」を層化（ストラータ）変数として選択します。
- ▶ 「町」をクラスタ変数として選択します。
- ▶ [次へ] をクリックし、[抽出方法] ステップで [次へ] をクリックします。

この計画構造は、各郡に対して独立サンプルが抽出されることを意味します。この段階（ステージ）では、町はデフォルトの [単純な無作為抽出] 方法により主要な抽出単位として抽出されます。

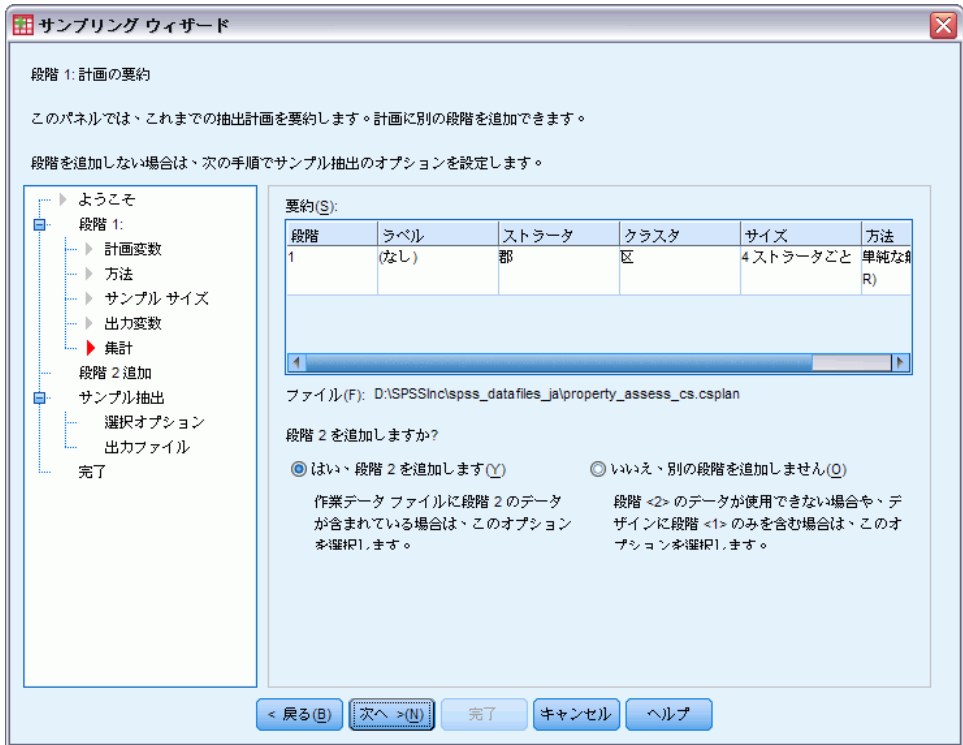


図 13-3  
サンプリング ウィザードの [サンプル サイズ] ステップ (第 1 段階)



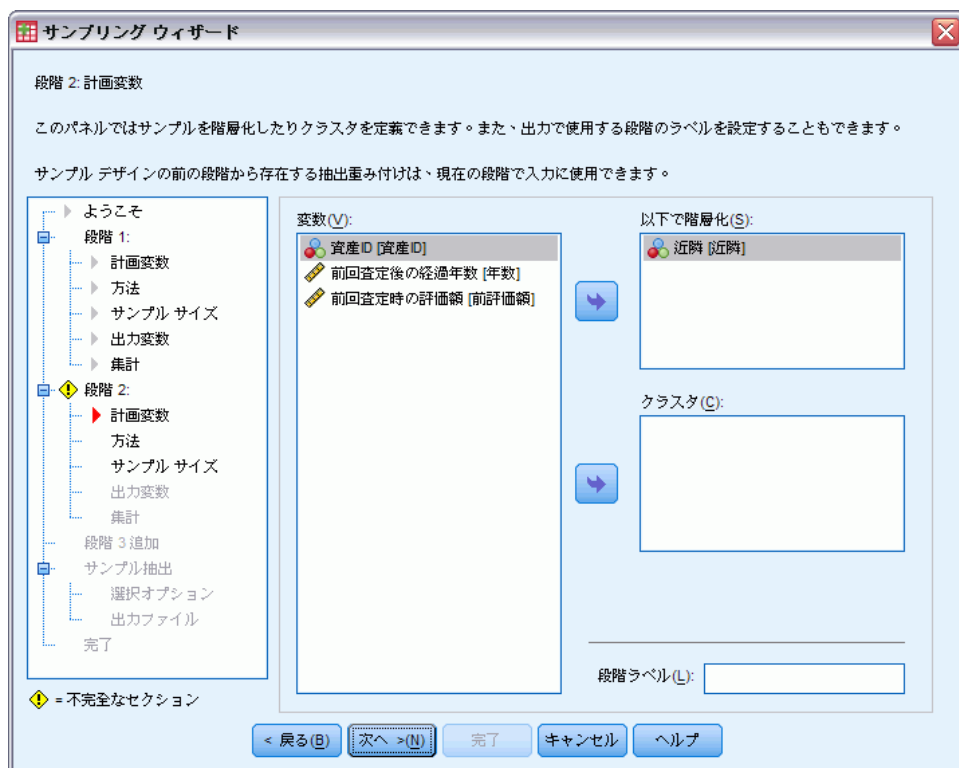
- ▶ [ユニット] ドロップダウン リストから、[度数] を選択します。
- ▶ この段階 (ステージ) で選択するユニットの数の値として、「4」と入力します。
- ▶ [次へ] をクリックし、[出力変数] ステップで [次へ] をクリックします。

図 13-4  
サンプリング ウィザードの [計画の要約] ステップ (第 1 段階)



- ▶ [はい、段階 2 を追加します] を選択します。
- ▶ [次へ] をクリックします。

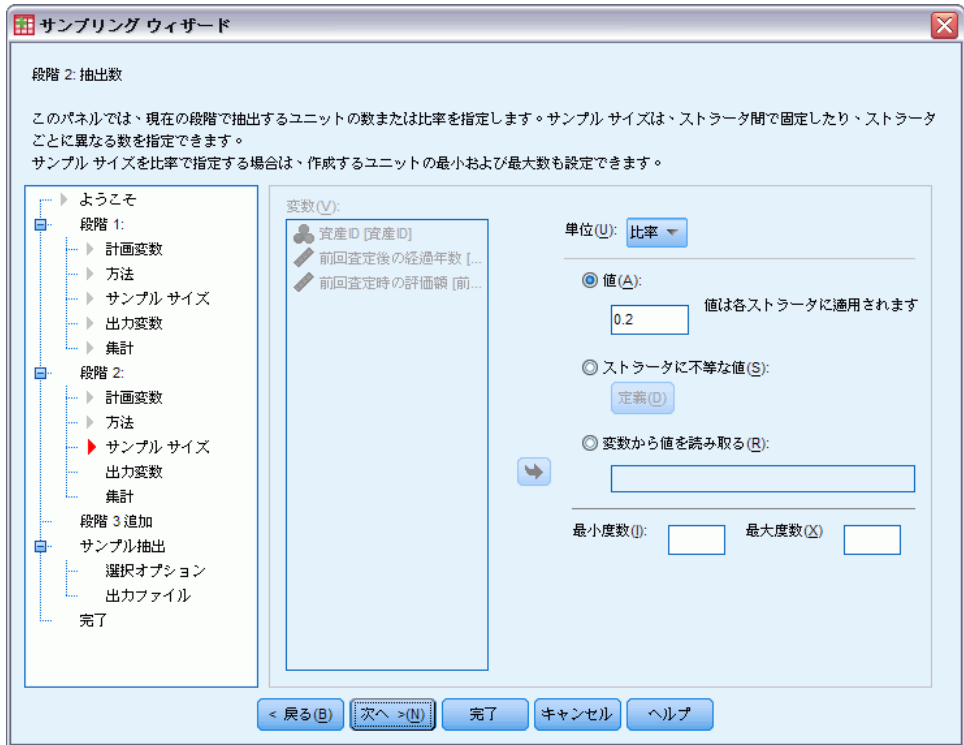
図 13-5  
サンプリング ウィザードの [計画変数] ステップ (第 2 段階)



- ▶ 「区域」を層化（ストラータ）変数として選択します。
- ▶ [次へ] をクリックし、[抽出方法] ステップで [次へ] をクリックします。

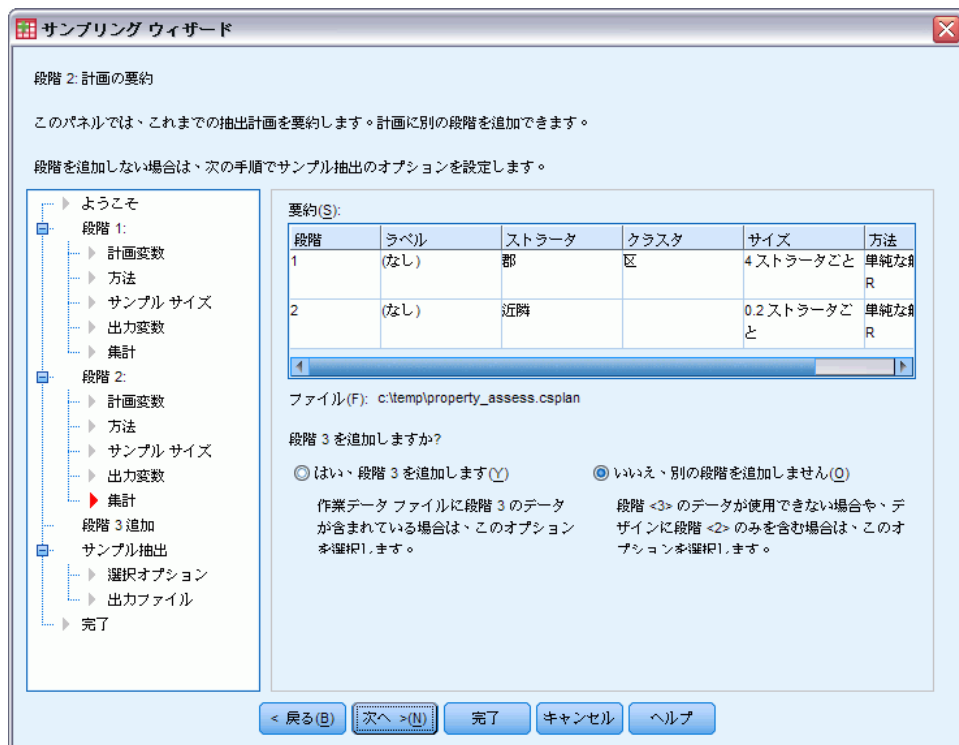
この計画構造は、第 1 段階（ステージ）で抽出された町の各区域に対して独立サンプルが抽出されることを意味します。この段階（ステージ）では、財産は [単純な無作為抽出] により主要な抽出ユニットとして抽出されます。

図 13-6  
サンプリング ウィザードの [サンプル サイズ] ステップ (第 2 段階)



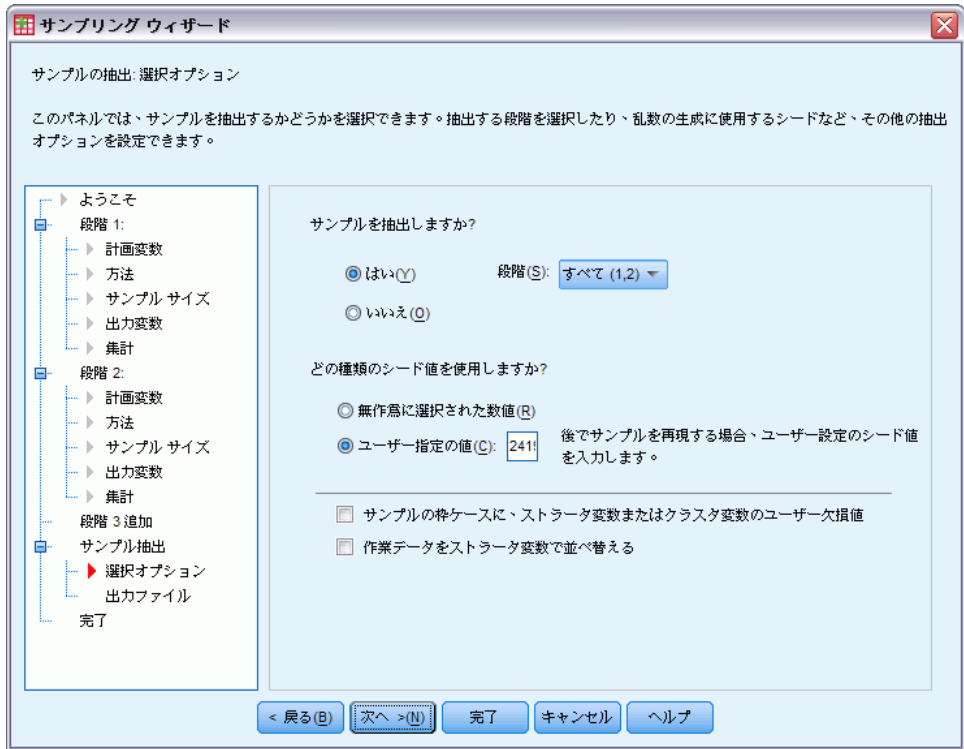
- ▶ [ユニット] ドロップダウン リストから、[比率] を選択します。
- ▶ 各ストラータからのサンプルに対するユニットの割合の値として、「0.2」と入力します。
- ▶ [次へ] をクリックし、[出力変数] ステップで [次へ] をクリックします。

図 13-7  
サンプリング ウィザードの [計画の要約] ステップ (第 2 段階)



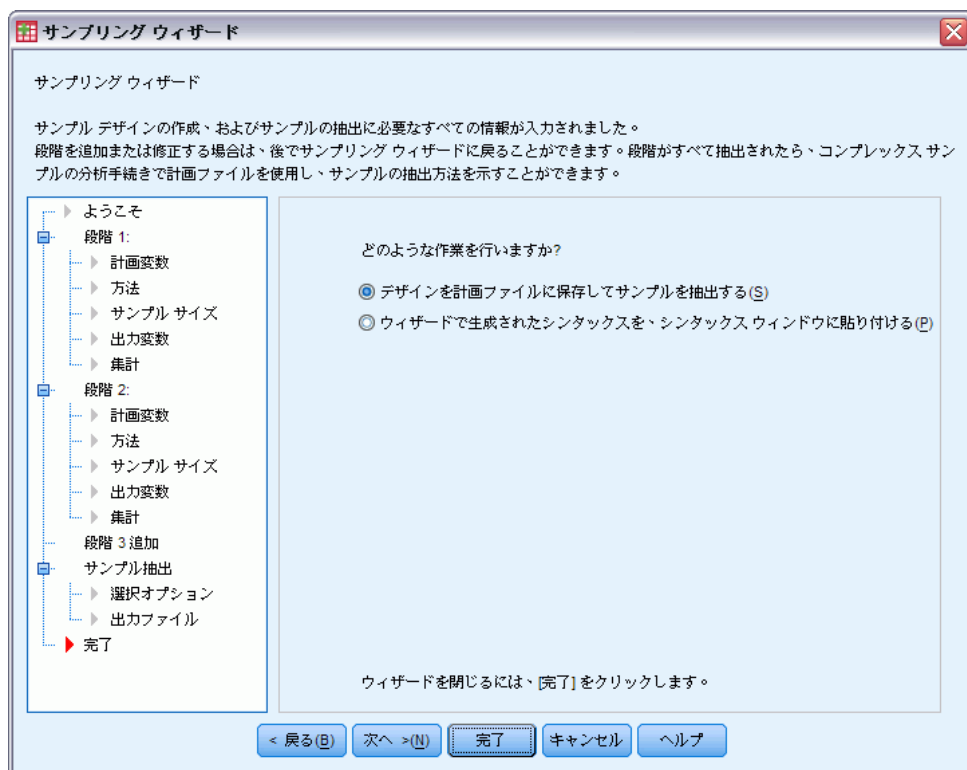
- ▶ 抽出計画を確認し、[次へ] をクリックします。

図 13-8  
サンプリング ウィザードの [サンプルの抽出: 選択オプション] ステップ



- ▶ 使用する乱数シードのタイプについて [ユーザー指定の値] を選択し、「241972」を値として入力します。  
カスタム値を使用すると、この例の結果をそのまま複製できます。
- ▶ [次へ] をクリックし、[サンプルの抽出: 出力ファイル] ステップで [次へ] をクリックします。

図 13-9  
サンプリング ウィザードの [完了] ステップ



- ▶ [完了] をクリックします。

以上の選択により、抽出計画ファイルの `property_assess.csplan` が作成され、計画に基づいてサンプルが抽出されます。

## 計画の要約

図 13-10  
計画の要約

			第 1 段階	第 2 段階
計画変数	ストラータ	1	郡	近所
	クラスタ	1	郡の下位行政区分	
サンプル情報	選択方法		置換を行わない単純無作為抽出	置換を行わない単純無作為抽出
	抽出されたユニットの数 作成または変更された変数	段階に関連する包含(選択)確率	4 Inclusion Probability _1_	Inclusion Probability _2_
		段階に関連する累積抽出重み付け	Sample Weight Cumulative e_1_	Sample Weight Cumulative e_2_
分析情報	抽出されたユニットの比率 推定量の仮定		置換を行わない等確率抽出 変数 Inclusion Probability _1_ から取得	置換を行わない等確率抽出 変数 Inclusion Probability _2_ から取得
	包含確率			.2

計画ファイル: C:\property\_assess.csplan  
重み付け変数: SampleWeight\_Final\_

要約表は抽出計画を確認し、計画が固有の目的を表しているかを調べるために便利です。

## 抽出の要約

図 13-11  
段階の集計

郡	抽出されたユニットの数		抽出されたユニットの比率	
	要求数	実際数	要求数	実際数
東部	4	4	44.4%	44.4%
中部	4	4	57.1%	57.1%
西部	4	4	25.0%	25.0%
北部	4	4	44.4%	44.4%
南部	4	4	50.0%	50.0%

この要約表は、抽出の最初の段階(ステージ)を確認します。抽出が計画どおりに行われたかを調べるために便利です。要求どおり、各郡から4つの町が抽出されました。



図 13-12  
段階の集計

郡	郡の低位行政区分	近所	抽出されたユニットの数		抽出されたユニットの比率	
			要求数	実際数	要求数	実際数
東部	2	8	4	4	20.0%	19.0%
		9	14	14	20.0%	20.6%
		10	7	7	20.0%	18.9%
		11	14	14	20.0%	20.0%
	6	36	13	13	20.0%	20.3%
		37	14	14	20.0%	20.6%
		38	13	13	20.0%	20.6%
	7	43	12	12	20.0%	20.7%
		44	11	11	20.0%	19.6%
		45	11	11	20.0%	20.8%
		46	13	13	20.0%	20.0%
	9	57	13	13	20.0%	20.6%
		58	5	5	20.0%	18.5%
		59	11	11	20.0%	19.3%
		60	13	13	20.0%	19.4%
中部	22	148	9	9	20.0%	19.6%
		149	8	8	20.0%	20.0%
		150	14	14	20.0%	20.0%
		151	11	11	20.0%	19.6%
	24	162	7	7	20.0%	20.6%
		163	7	7	20.0%	20.6%
		164	5	5	20.0%	21.7%
	25	169	14	14	20.0%	20.3%
		170	16	16	20.0%	20.3%
		171	15	15	20.0%	20.0%
		172	8	8	20.0%	19.0%
	27	183	14	14	20.0%	20.3%
		184	15	15	20.0%	19.7%
		185	15	15	20.0%	20.0%
		186	15	15	20.0%	20.5%
187		6	6	20.0%	20.7%	

この要約表（ここで示される最上位の部分）は、抽出の第2段階（ステージ）を確認します。これは、抽出が計画どおりに実行されたかどうかを調べる場合にも便利です。要求に応じて、第1段階（ステージ）で抽出された各町の各区域から財産の約20%が抽出されました。

## 抽出結果

図 13-13  
データ エディタの抽出結果

	資産ID	近隣	区	郡	年数	前評価額	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	SampleWeight_Final_
279	642.0	9	2	1	6	236.70	.	.	.	.	.
280	643.0	9	2	1	6	150.40	.	.	.	.	.
281	644.0	9	2	1	8	204.80	.	.	.	.	.
282	645.0	9	2	1	6	225.40	.	.	.	.	.
283	646.0	9	2	1	7	180.80	.	.	.	.	.
284	647.0	9	2	1	5	176.90	.	.	.	.	.
285	648.0	9	2	1	6	216.20	0.44	2.25	0.21	10.93	10.93
286	649.0	9	2	1	5	203.50	.	.	.	.	.
287	650.0	9	2	1	6	223.30	.	.	.	.	.
288	651.0	9	2	1	7	215.90	0.44	2.25	0.21	10.93	10.93
289	652.0	9	2	1	6	207.60	0.44	2.25	0.21	10.93	10.93
290	653.0	9	2	1	6	208.10	.	.	.	.	.

データ エディタで抽出結果を参照できます。新しい5つの変数が作業ファイルに保存され、各段階（ステージ）の包含確率や累積抽出重み付け、および最終的な抽出重み付けを示しています。

- これらの変数の値を持つケースがサンプルに選択されました。
- 変数に対してシステム欠損値を持つケースは選択されませんでした。

州当局は、州のリソースを使用して、サンプルで選択された財産の現在の評価を収集します。これらの評価が利用可能になったら、サンプルを [コンプレックス サンプル] 分析手続きで処理し、抽出計画の property\_assess.csplan を使用して抽出仕様を指定できます。

## 部分サンプリング枠からのサンプル抽出

ある企業が、高品質の調査情報データベースのコンパイルや販売に興味を持っています。調査サンプルは代表でなければなりません、効率的に実行する必要があるため、コンプレックス サンプリング法が使用されません。完全抽出計画には以下の構造が必要です。

段階	ストラータ	クラスター (U)
1	地域	地方

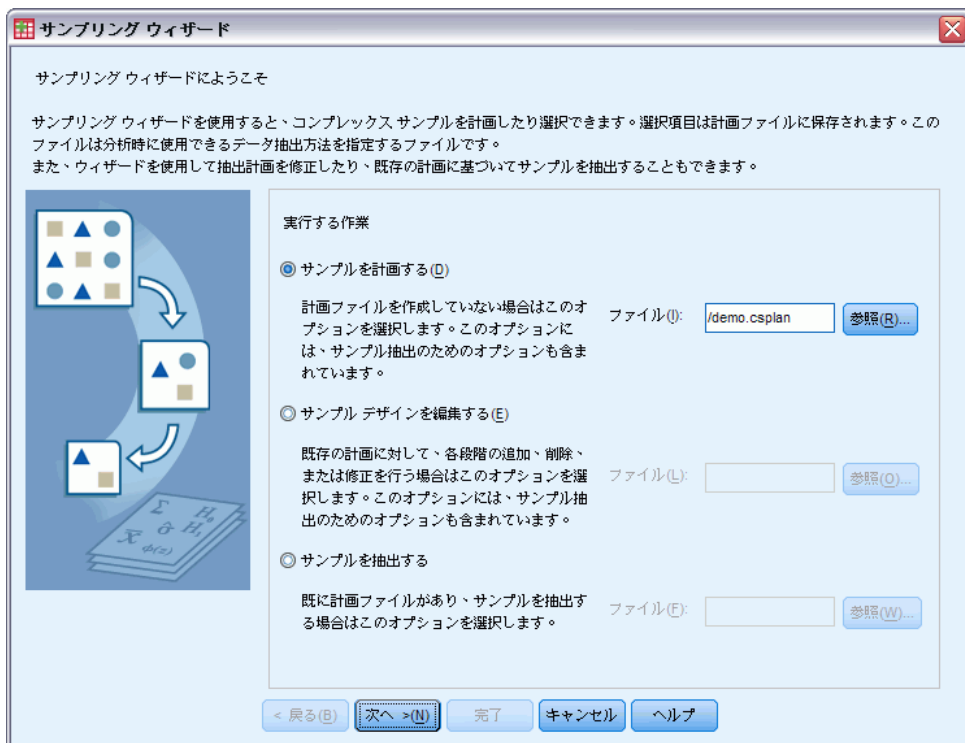
- 2 地区 市
- 3 再分割

第3段階（ステージ）では、世帯が主要な抽出ユニットであり、選択された世帯が調査されます。ただし、情報を簡単に利用できるのは市レベルでのみなので、企業は、まず計画の最初の2つの段階（ステージ）を実行してから、抽出した市の区画および世帯の数に関する情報を収集することになりました。市レベルで利用可能な情報は、demo\_cs\_1.sav に収集されています。詳細は、A 付録 p. 281 [サンプル ファイル](#) を参照してください。このファイルには、すべて 1 が指定された変数再分割が含まれています。これは、「真」の変数のプレースホルダです。この変数の値は、計画の最初の2つの段階（ステージ）を実行した後に収集されます。これにより、完全な3段階の抽出計画を指定できます。コンプレックス サンプルのサンプリング ウィザードを使用して、完全な複合抽出計画を指定し、最初の2つの段階（ステージ）を抽出します。

## ウィザードによる最初の部分枠からの抽出

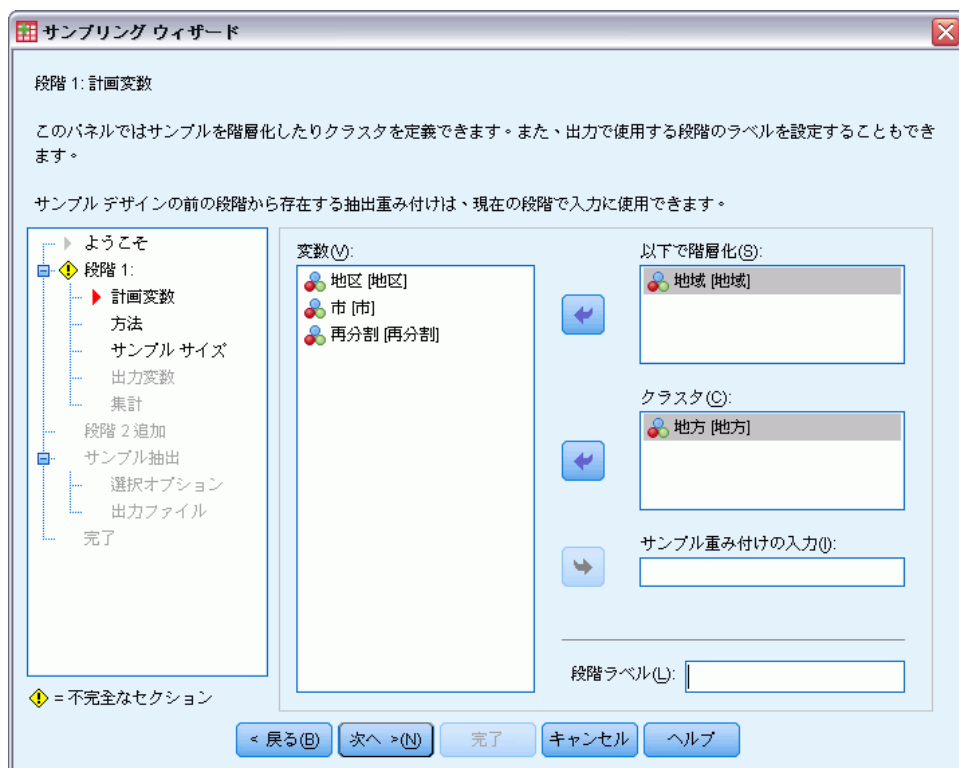
- ▶ コンプレックス サンプルのサンプリング ウィザードを実行するには、メニューから次の項目を選択します。  
分析(A) > コンプレックス サンプル > サンプルの選択...

図 13-14  
サンプリング ウィザードの [ようこそ] ステップ



- ▶ [サンプルを計画する] を選択し、ファイルの保存場所を参照して、計画ファイル名として「demo.csplan」と入力します。
- ▶ [次へ] をクリックします。

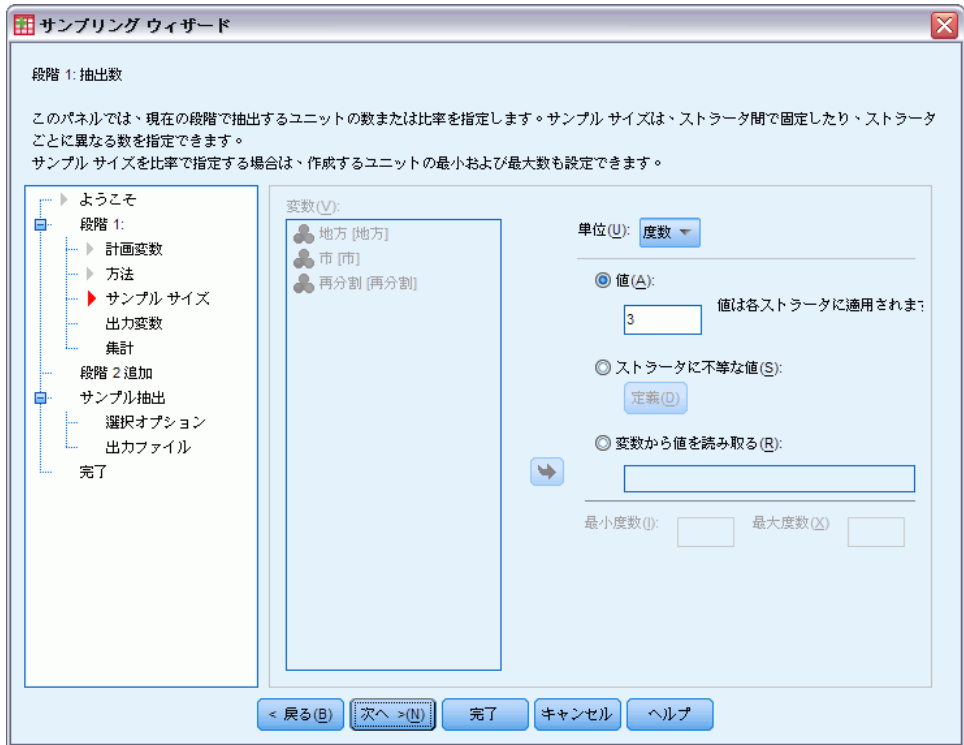
図 13-15  
サンプリング ウィザードの [計画変数] ステップ (第 1 段階)



- ▶ 「地域」を層化 (ストラータ) 変数として選択します。
- ▶ 「地方」をクラスタ変数として選択します。
- ▶ [次へ] をクリックし、[抽出方法] ステップで [次へ] をクリックします。

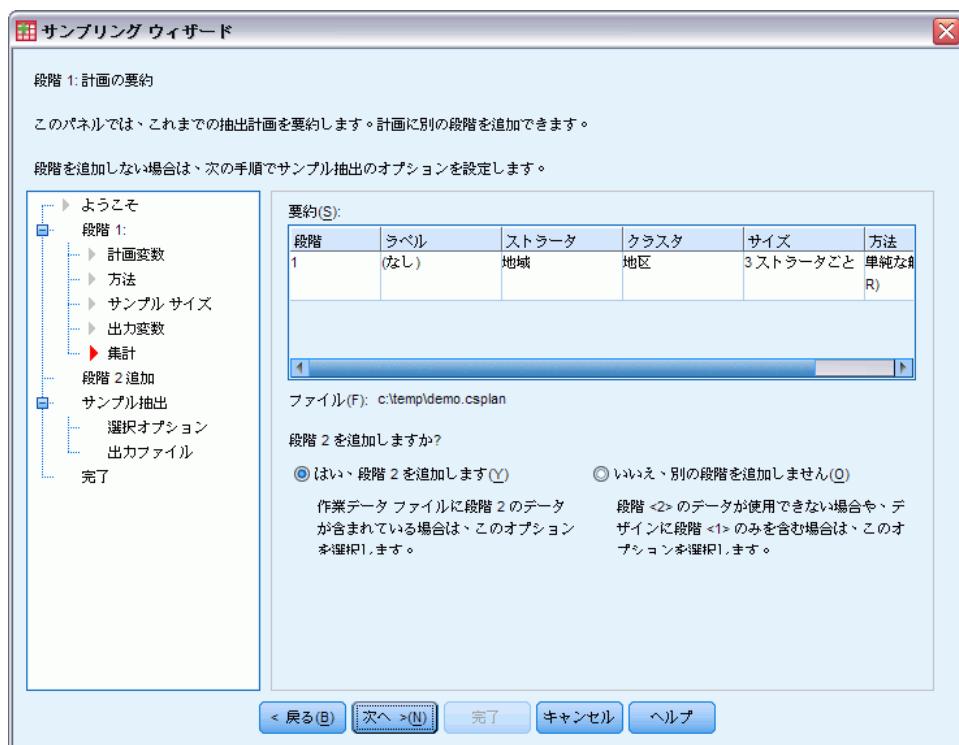
この計画構造は、各地域に対して独立サンプルが抽出されることを意味します。この段階 (ステージ) では、地方はデフォルトの [単純な無作為抽出] 方法により主要な抽出単位として抽出されます。

図 13-16  
サンプリング ウィザードの [サンプル サイズ] ステップ (第 1 段階)



- ▶ [ユニット] ドロップダウン リストから、[度数] を選択します。
- ▶ この段階 (ステージ) で選択するユニットの数の値として、「3」と入力します。
- ▶ [次へ] をクリックし、[出力変数] ステップで [次へ] をクリックします。

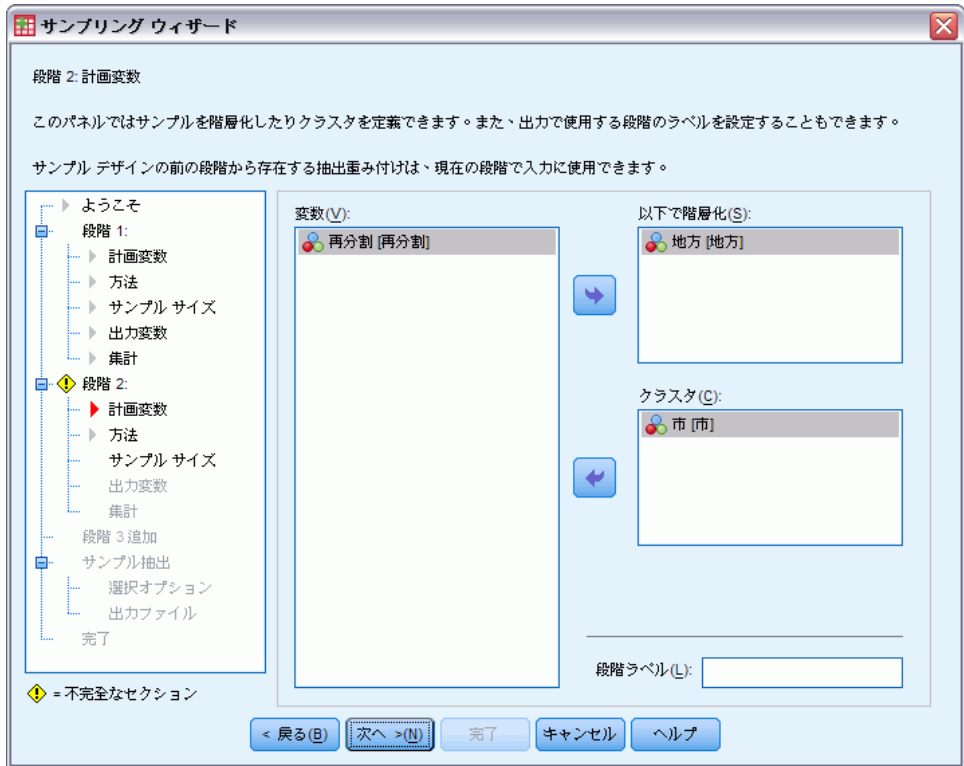
図 13-17  
サンプリング ウィザードの [計画の要約] ステップ (第 1 段階)



▶ [はい。段階 2 を追加します] を選択します。

▶ [次へ] をクリックします。

図 13-18  
サンプリング ウィザードの [計画変数] ステップ (第 2 段階)

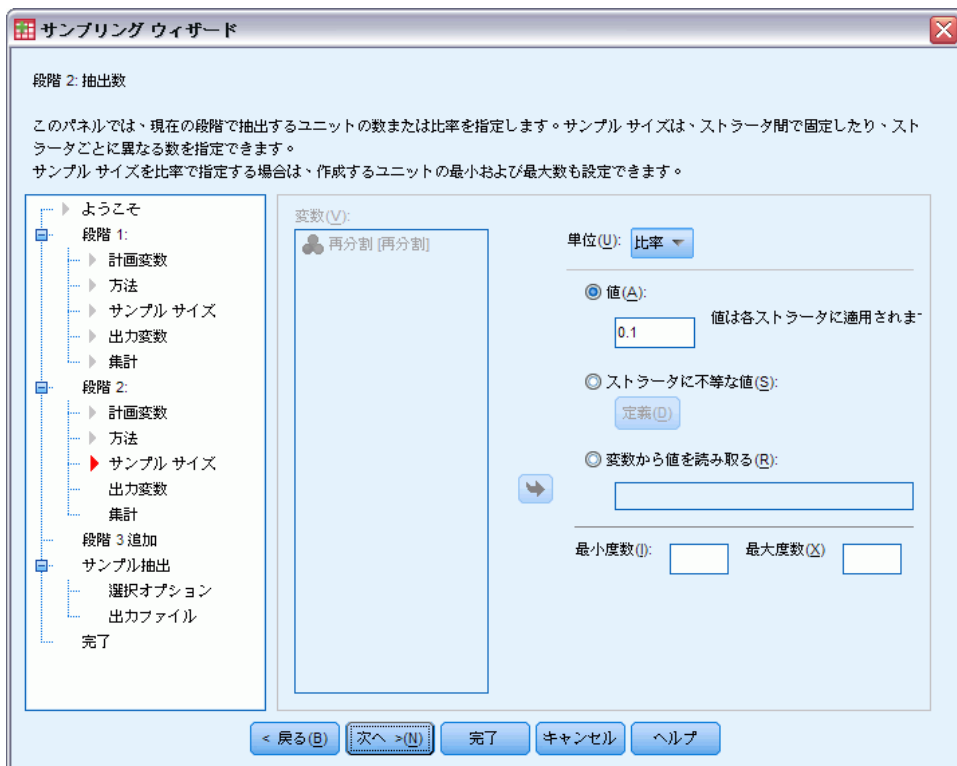


- ▶ 「地区」を層化（ストラータ）変数として選択します。
- ▶ 「市」をクラスタ変数として選択します。
- ▶ [次へ] をクリックし、[抽出方法] ステップで [次へ] をクリックします。

この計画構造は、各地区に対して独立サンプルが抽出されることを意味します。この段階（ステージ）では、都市はデフォルトの [単純な無作為抽出] 方法により主要な抽出単位として抽出されます。

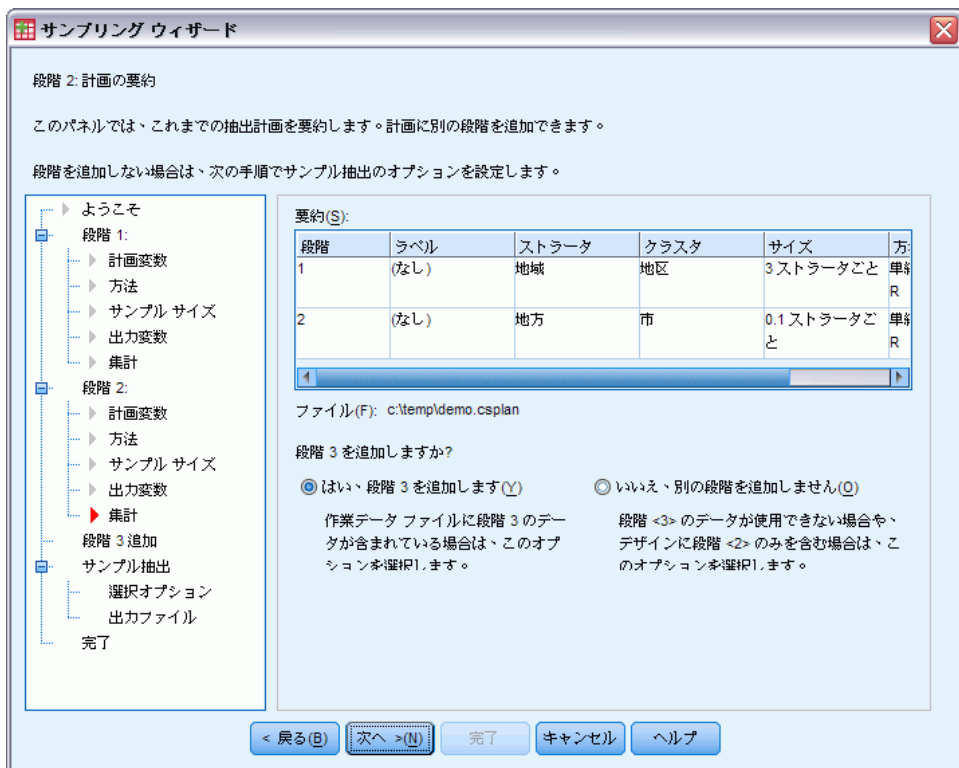


図 13-19  
サンプリング ウィザードの [サンプル サイズ] ステップ (第 2 段階)



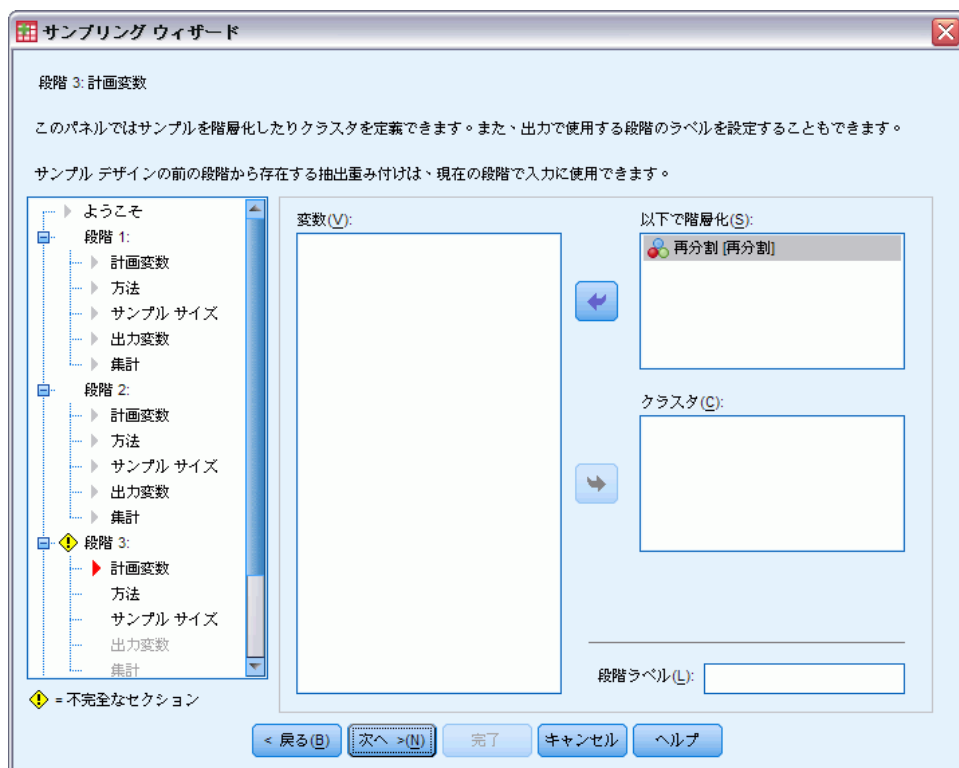
- ▶ [ユニット] ドロップダウン リストから、[比率] を選択します。
- ▶ 各ストラータからのサンプルに対するユニットの割合の値として、「0.1」と入力します。
- ▶ [次へ] をクリックし、[出力変数] ステップで [次へ] をクリックします。

図 13-20  
サンプリング ウィザードの [計画の要約] ステップ (第 2 段階)



- ▶ [はい。段階 3 を追加します] を選択します。
- ▶ [次へ] をクリックします。

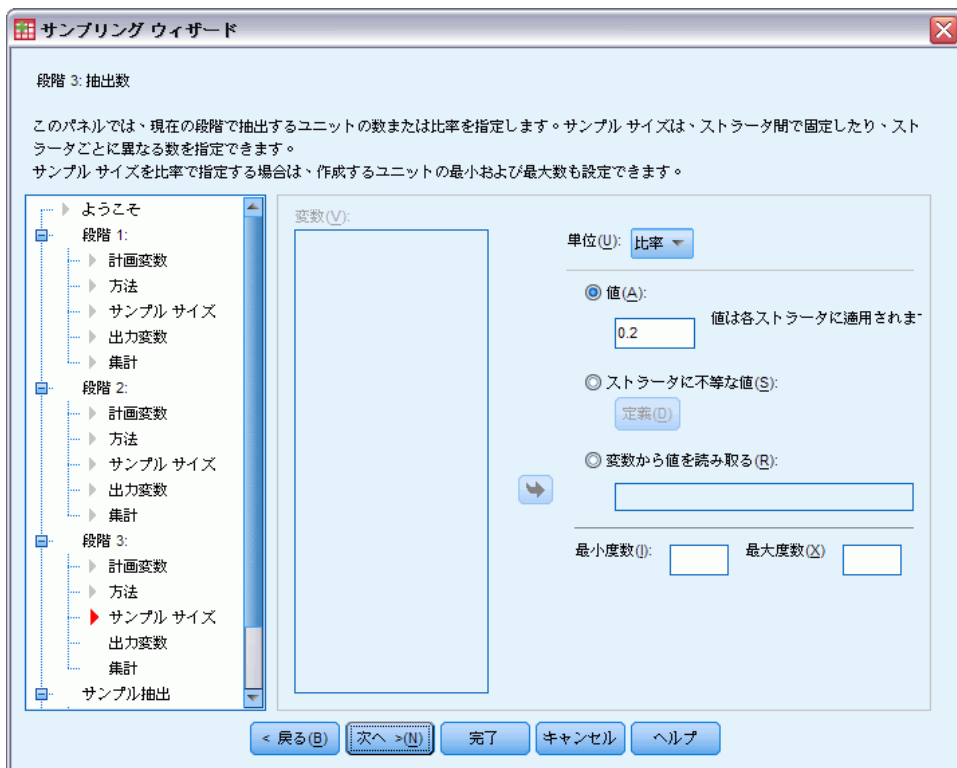
図 13-21  
サンプリング ウィザードの [計画変数] ステップ (第 3 段階)



- ▶ 「再分割」を層化（ストラータ）変数として選択します。
- ▶ [次へ] をクリックし、[抽出方法] ステップで [次へ] をクリックします。

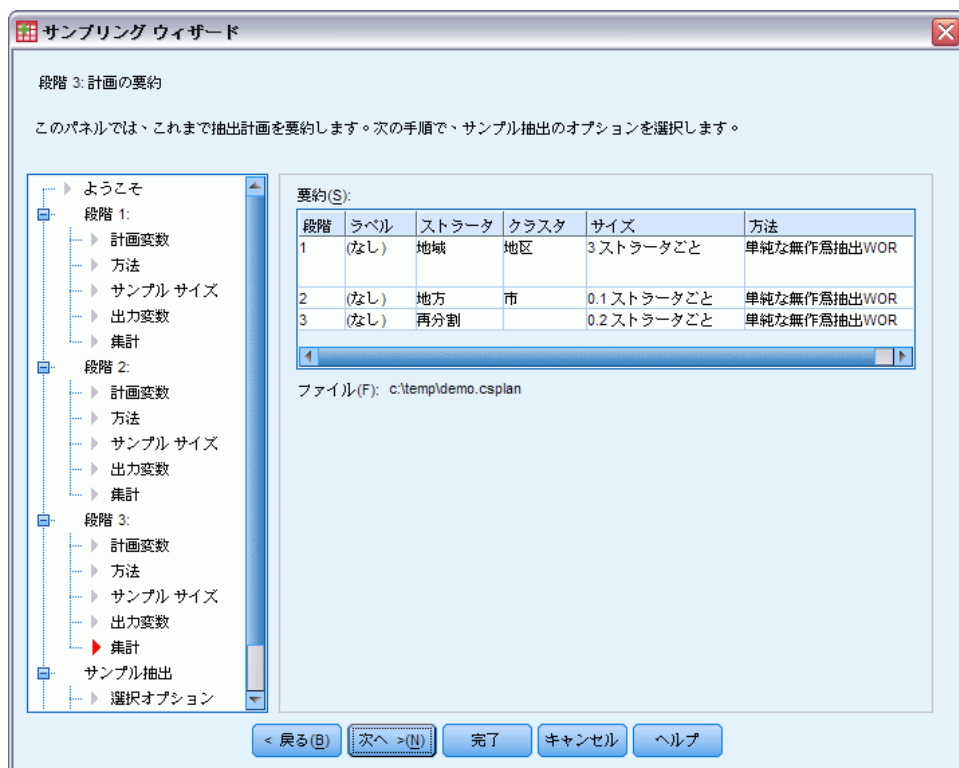
この計画構造は、各区画に対して独立サンプルが抽出されることを意味します。この段階（ステージ）では、世帯単位はデフォルトの [単純な無作為抽出] 方法により主要な抽出単位として抽出されます。

図 13-22  
サンプリング ウィザードの [サンプル サイズ] ステップ (第 3 段階)



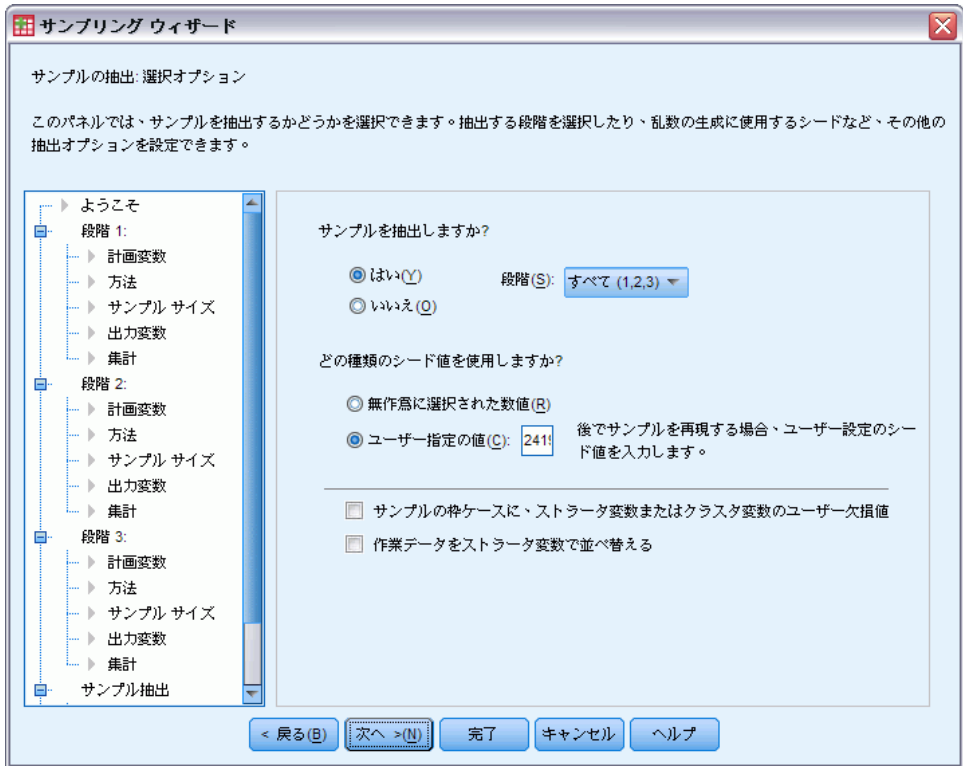
- ▶ [ユニット] ドロップダウン リストから、[比率] を選択します。
- ▶ この段階 (ステージ) で選択するユニットの割合の値として、「0.2」と入力します。
- ▶ [次へ] をクリックし、[出力変数] ステップで [次へ] をクリックします。

図 13-23  
サンプリング ウィザードの [計画の要約] ステップ (第 3 段階)



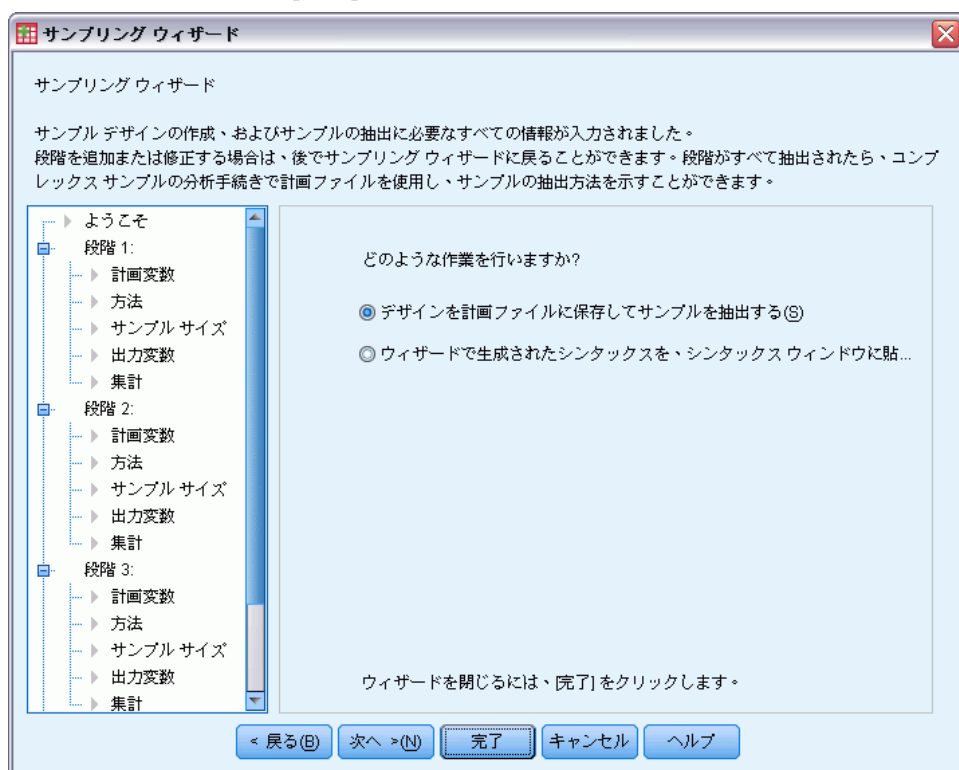
- ▶ 抽出計画を確認し、[次へ] をクリックします。

図 13-24  
サンプリング ウィザードの [サンプルの抽出: 選択オプション] ステップ



- ▶ ここで抽出する段階として [1, 2] を選択します。
- ▶ 使用する乱数シードのタイプについて [ユーザー指定の値] を選択し、「241972」を値として入力します。  
カスタム値を使用すると、この例の結果をそのまま複製できます。
- ▶ [次へ] をクリックし、[サンプルの抽出: 出力ファイル] ステップで [次へ] をクリックします。

図 13-25  
サンプリング ウィザードの [完了] ステップ



- ▶ [完了] をクリックします。

以上の選択により、抽出計画ファイルの `demo.csplan` が作成され、計画の最初の 2 つの段階 (ステージ) に基づいてサンプルが抽出されます。

## 抽出結果

図 13-26  
データ エディタの抽出結果

	地域	地方	地区	市	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	SampleWeight_Final_
295	1	2	10	295	.	.	.	.	.
296	1	2	10	296	.	.	.	.	.
297	1	2	10	297	.	.	.	.	.
298	1	2	10	298	0.20	5.00	0.10	50.00	50.00
299	1	2	10	299	.	.	.	.	.
300	1	2	10	300	0.20	5.00	0.10	50.00	50.00
301	1	2	11	301	.	.	.	.	.
302	1	2	11	302	.	.	.	.	.
303	1	2	11	303	.	.	.	.	.
304	1	2	11	304	.	.	.	.	.
305	1	2	11	305	.	.	.	.	.
306	1	2	11	306	.	.	.	.	.
307	1	2	11	307	0.20	5.00	0.10	50.00	50.00
308	1	2	11	308	.	.	.	.	.

データ ビュー(D) 変数 ビュー(V)

データ エディタで抽出結果を参照できます。新しい5つの変数が作業ファイルに保存され、各段階（ステージ）の包含確率や累積抽出重み付けと、最初の2つの段階の「最終的な」抽出重み付けが示されています。

- これらの変数の値を持つ町がサンプルに選択されました。
- 変数に対してシステム欠損値を持つ町は選択されませんでした。

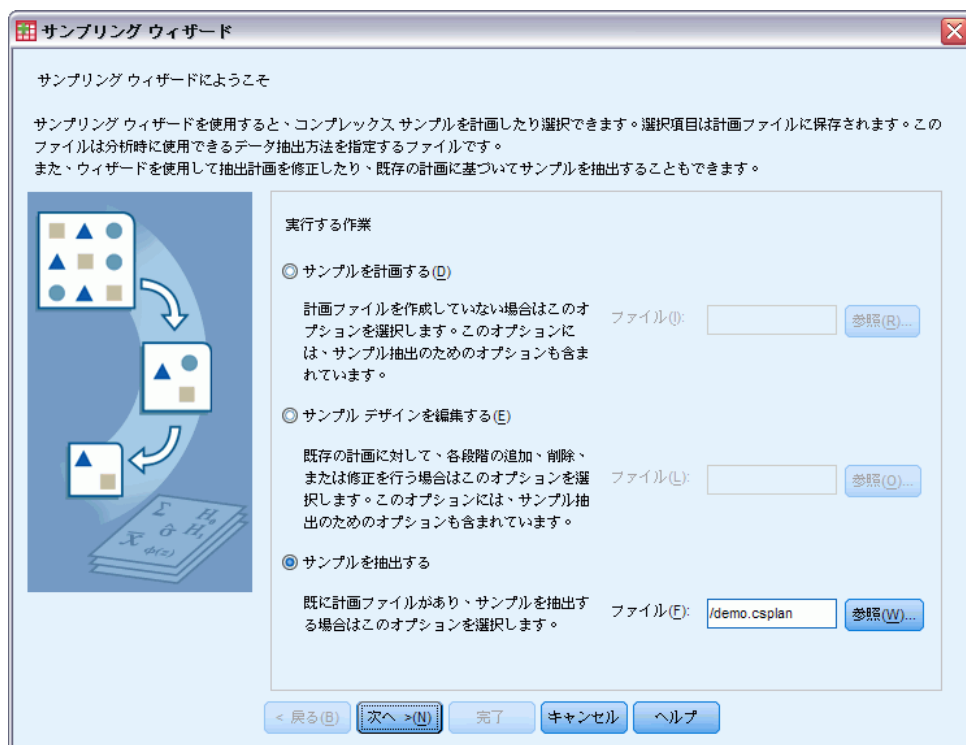
選択された各市に対して、企業は区画単位および世帯単位の情報を取得し、demo\_cs\_2.sav に保存しました。このファイルと、この計画の第3段階（ステージ）を抽出するサンプリング ウィザードを使用します。

## ウィザードによる2番目の部分枠からの抽出

- ▶ コンプレックス サンプルのサンプリング ウィザードを実行するには、メニューから次の項目を選択します。  
分析(A) > コンプレックス サンプル > サンプルの選択...

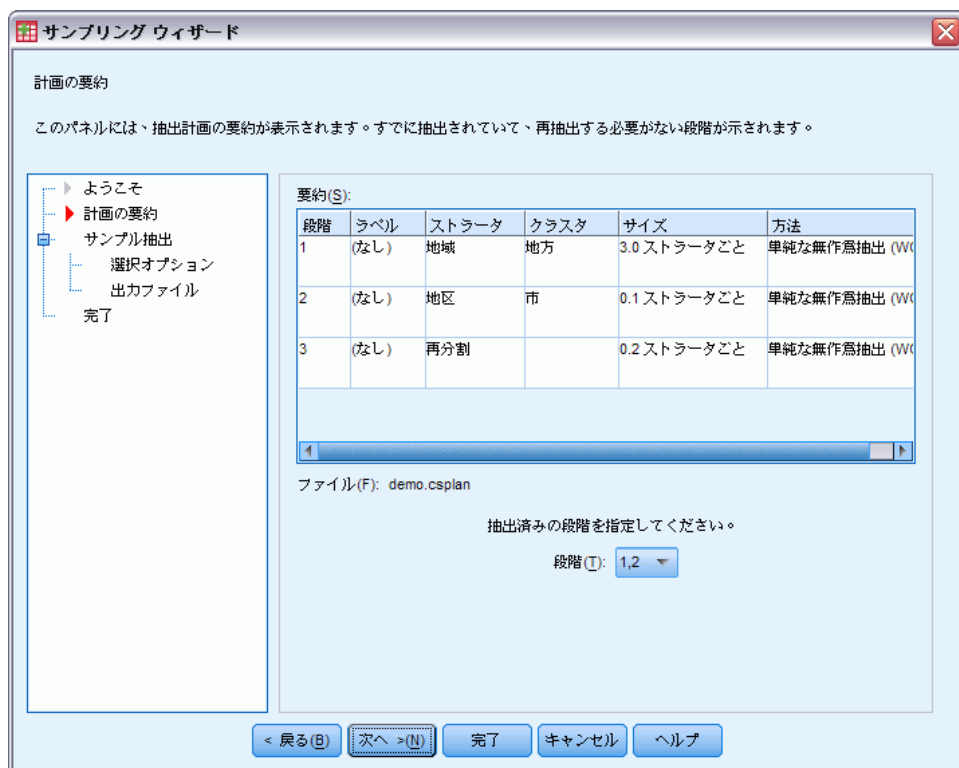


図 13-27  
サンプリング ウィザードの [ようこそ] ステップ



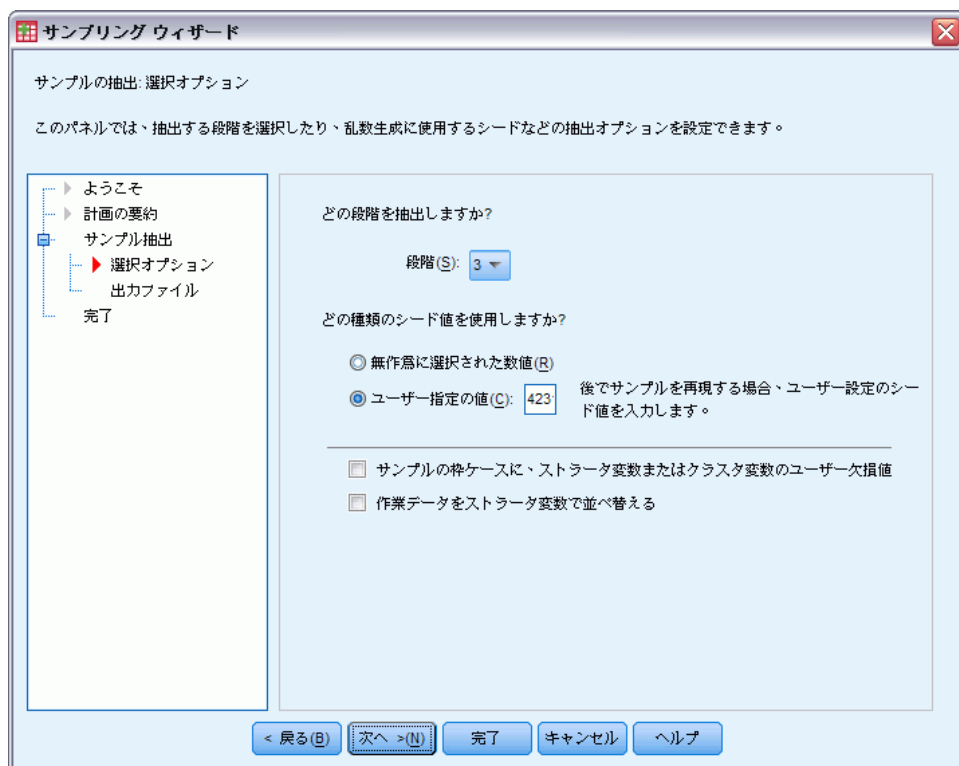
- ▶ [サンプルを抽出する] を選択し、ファイルの保存場所を参照して、作成した demo.csplan 計画ファイルを選択します。
- ▶ [次へ] をクリックします。

図 13-28  
サンプリング ウィザードの [計画の要約] ステップ (第 3 段階)



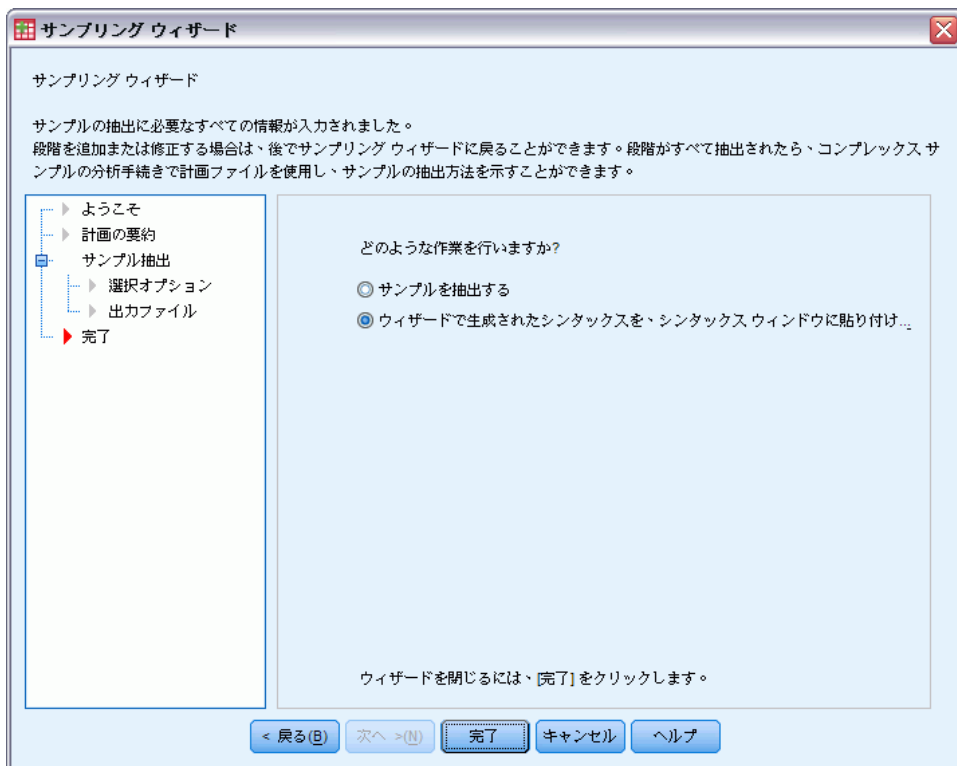
- ▶ 抽出済みの段階として [1, 2] を選択します。
- ▶ [次へ] をクリックします。

図 13-29  
サンプリング ウィザードの [サンプルの抽出: 選択オプション] ステップ



- ▶ 使用する乱数シードのタイプに対する[ユーザー指定の値]を選択し、「4231946」を値として入力します。
- ▶ [次へ]をクリックし、[サンプルの抽出: 出力ファイル]ステップで[次へ]をクリックします。

図 13-30  
サンプリング ウィザードの [完了] ステップ



- ▶ [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] を選択します。
- ▶ [完了] をクリックします。

次のシンタックスが生成されます。

\* サンプリング ウィザード .

```
CSSELECT
/PLAN FILE='demo.csplan'
/CRITERIA STAGES = 3 SEED = 4231946
/CLASSMISSING EXCLUDE
/DATA RENAMEVARS
/PRINT SELECTION.
```

この場合、抽出の要約を出力すると、有用性の低い表が作成され出力ビューアで問題が生じます。抽出の要約の表示を無効にするには、PRINT サブコマンドの SELECTION を CPS に置換します。次に、シンタックス ウィンドウ内でシンタックスを実行します。

以上の選択により、demo.csplan 抽出計画の第 3 段階（ステージ）に基づいてサンプルが抽出されます。

## 抽出結果

図 13-31  
データ エディタの抽出結果

	市	再分割	ユニット	InclusionProbability_2_	SampleWeightCumulative_2_	InclusionProbability_3_1_	SampleWeightCumulative_3_1_	SampleWeight_Final_1_	Inclab
14	190	946	94514	0.10	50.00	.	.	.	
15	190	946	94515	0.10	50.00	.	.	.	
16	190	946	94516	0.10	50.00	0.20	244.44	244.44	
17	190	946	94517	0.10	50.00	.	.	.	
18	190	946	94518	0.10	50.00	.	.	.	
19	190	946	94519	0.10	50.00	.	.	.	
20	190	946	94520	0.10	50.00	.	.	.	
21	190	946	94521	0.10	50.00	.	.	.	
22	190	946	94522	0.10	50.00	.	.	.	
23	190	946	94523	0.10	50.00	.	.	.	
24	190	946	94524	0.10	50.00	0.20	244.44	244.44	
25	190	946	94525	0.10	50.00	.	.	.	
26	190	946	94526	0.10	50.00	.	.	.	
27	190	946	94527	0.10	50.00	.	.	.	
28	190	946	94528	0.10	50.00	.	.	.	
29	190	946	94529	0.10	50.00	0.20	244.44	244.44	
30	190	946	94530	0.10	50.00	.	.	.	

データ エディタで抽出結果を参照できます。新しい 3 つの変数が作業ファイルに保存され、第 3 段階 (ステージ) の包含確率や累積抽出重み付けが示されています。これらの新しい重み付けは、最初の 2 つの段階 (ステージ) の抽出中に計算された重み付けを考慮しています。

- これらの変数の値を持つ単位がサンプルに選択されました。
- これらの変数に対してシステム欠損値を持つ単位は選択されませんでした。

企業は、企業のリソースを使用して、サンプルで選択された世帯単位の調査情報を収集します。これらの評価が収集されたら、サンプルを [コンプレックス サンプル] 分析手続きで処理し、抽出計画の demo.csplan を使用して抽出仕様を指定できます。

## 確率比例 (PPS) 法を使用した抽出

議会開会前に法案の提出を検討している議員は、市民が法案を支持しているかどうか、法案への支持がどの程度有権者の人口統計に関連しているかに関心を持っています。そこで、世論調査会社が複合抽出計画に基づいて対面式の調査を実施することになりました。

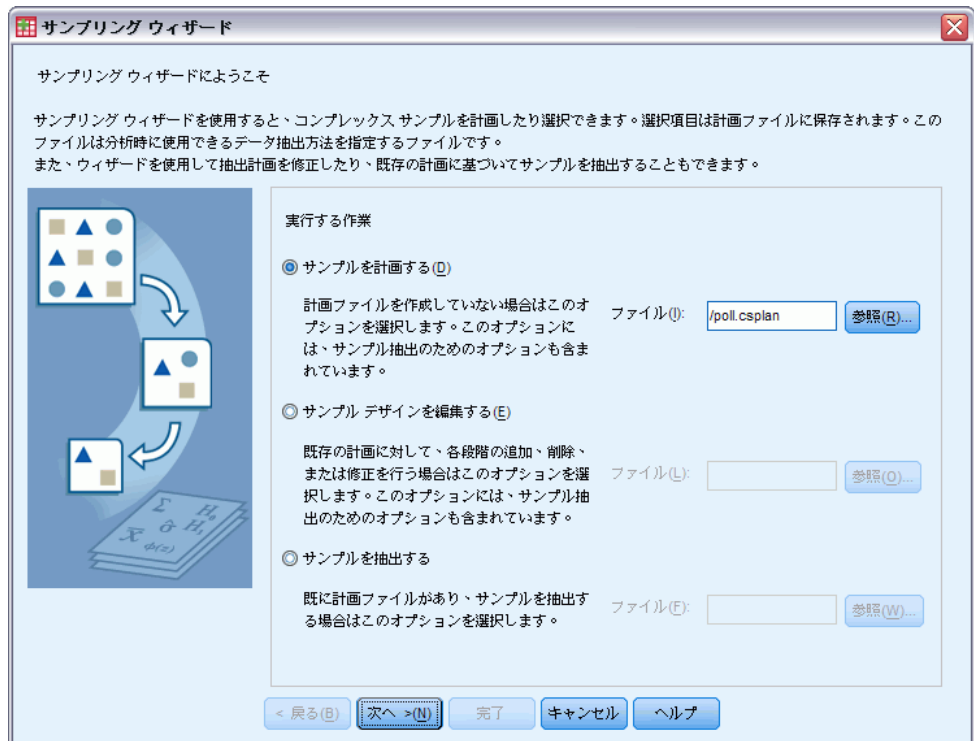
登録有権者のリストは、poll\_cs.sav に収集されています。詳細は、A 付録 p.281 サンプル ファイル を参照してください。コンプレックス サンプルのサンプリング ウィザードを使用して、分析を進めるためのサンプルを選択します。

## ウィザードの使用方法

- ▶ コンプレックス サンプルのサンプリング ウィザードを実行するには、メニューから次の項目を選択します。

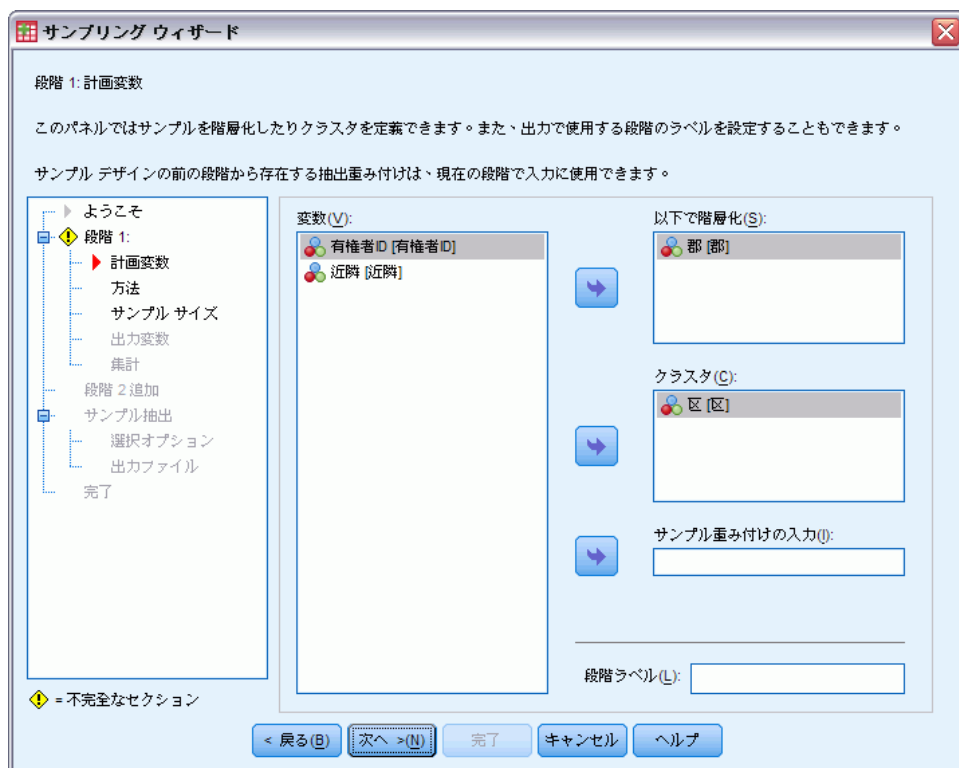
分析(A) > コンプレックス サンプル > サンプルの選択...

図 13-32  
サンプリング ウィザードの [ようこそ] ステップ



- ▶ [サンプルを計画する] を選択し、ファイルの保存場所を参照して、計画ファイル名として「poll.csplan」と入力します。
- ▶ [次へ] をクリックします。

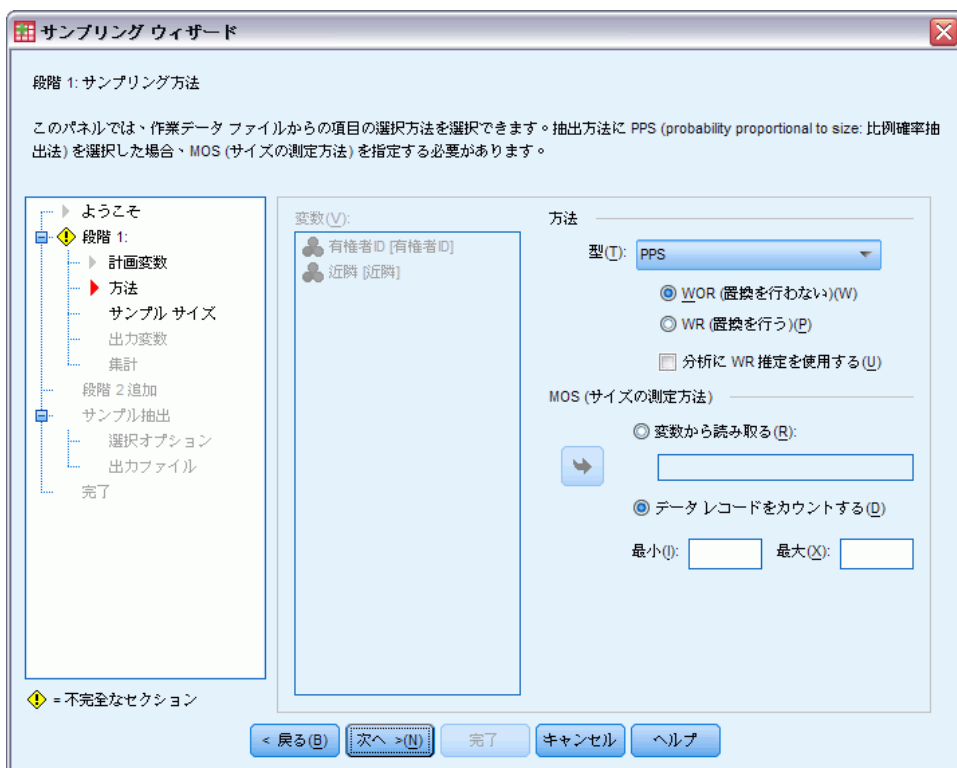
図 13-33  
サンプリング ウィザードの [計画変数] ステップ (第 1 段階)



- ▶ 「郡」を層化（ストラータ）変数として選択します。
- ▶ 「町」をクラスタ変数として選択します。
- ▶ [次へ] をクリックします。

この計画構造は、各郡に対して独立サンプルが抽出されることを意味します。この段階（ステージ）では、町は主要な抽出単位として抽出されます。

図 13-34  
サンプリング ウィザードの [抽出方法] ステップ (第 1 段階)

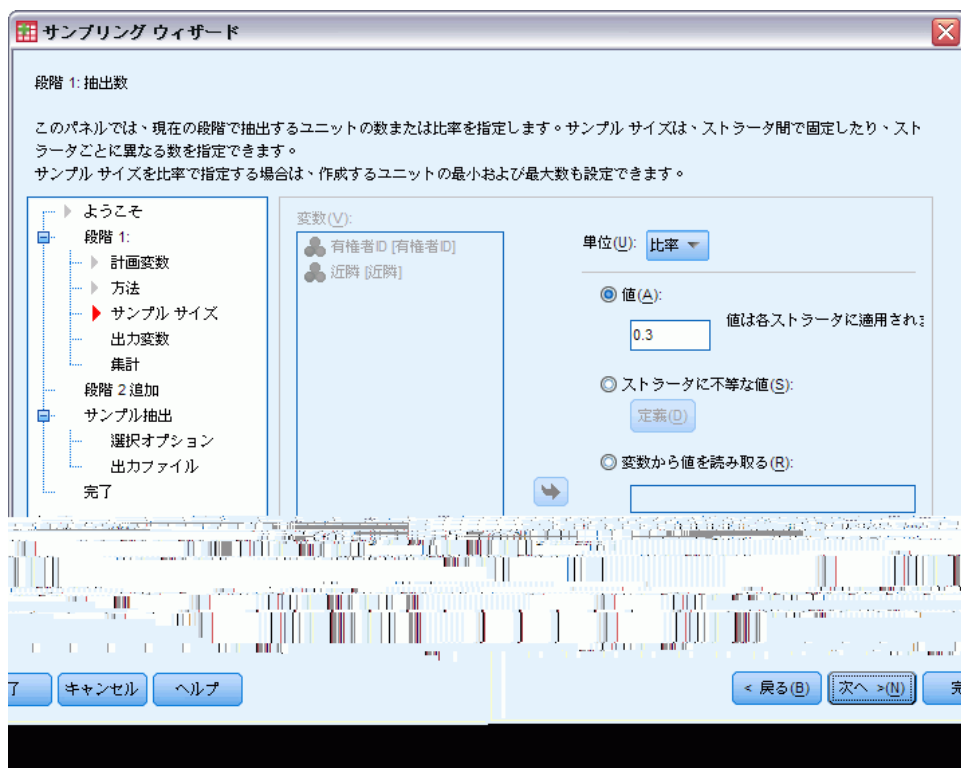


- ▶ 抽出方法として、[PPS] を選択します。
- ▶ サイズの測定方法として、[データレコードをカウントする] を選択します。
- ▶ [次へ] をクリックします。

各町のレコード数に比例した確率で置換を行わずに、郡ごとに町が抽出されます。PPS 法を使用すると、町についての結合抽出確率が生成されます。この値の保存先は、[出力ファイル] ステップで指定します。



図 13-35  
サンプリング ウィザードの [サンプル サイズ] ステップ (第 1 段階)

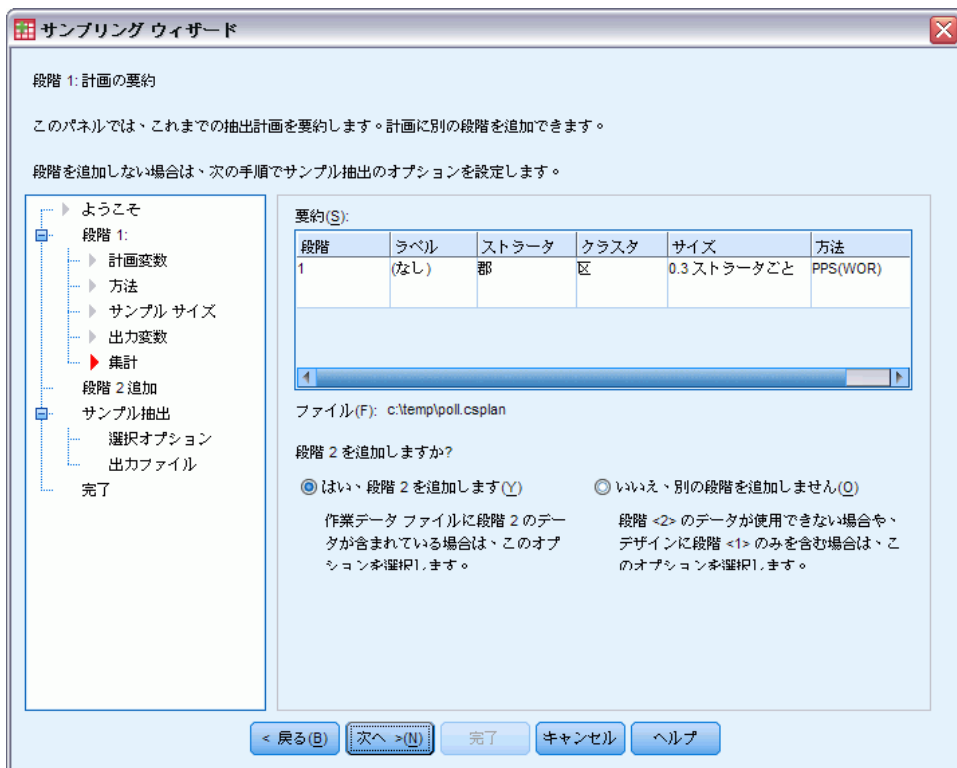


- ▶ [ユニット] ドロップダウン リストから、[比率] を選択します。
- ▶ この段階 (ステージ) で郡ごとに選択する町の割合の値として、「0.3」と入力します。

西部郡選出の議員は、西部郡の町の数比其他の郡よりも少ないことに注目しました。そこで、実態を適切に表現できるように、各郡から抽出する町の最小数を 3 に定義することにします。

- ▶ 選択する町の最小数として「3」、最大数として「5」を入力します。
- ▶ [次へ] をクリックし、[出力変数] ステップで [次へ] をクリックします。

図 13-36  
サンプリング ウィザードの [計画の要約] ステップ (第 1 段階)



- ▶ [はい。段階 2 を追加します] を選択します。
- ▶ [次へ] をクリックします。

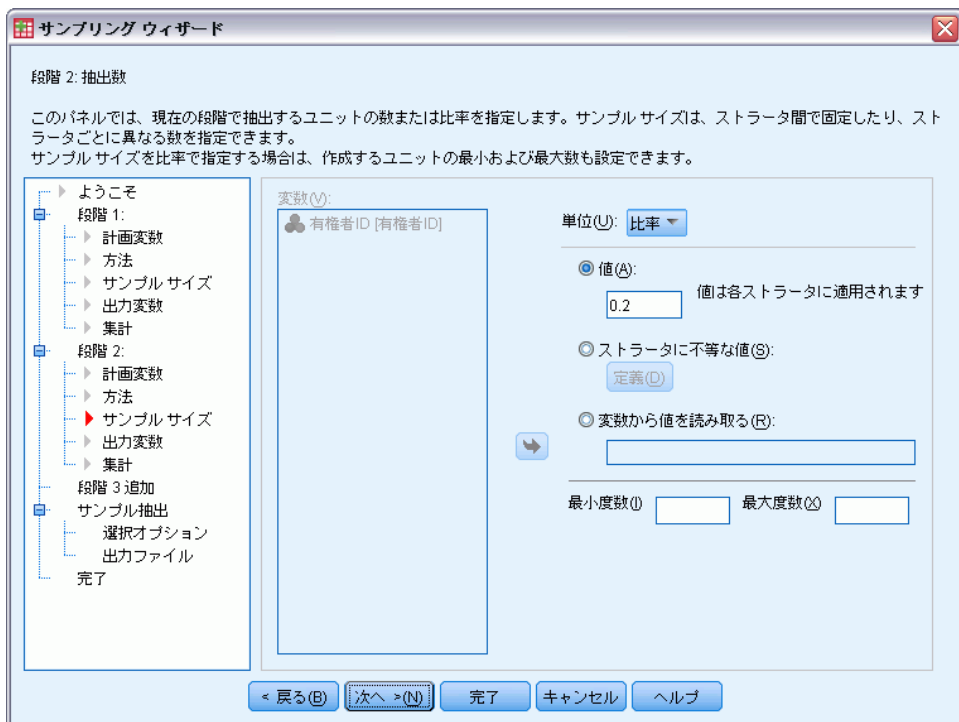
図 13-37  
サンプリング ウィザードの [計画変数] ステップ (第 2 段階)



- ▶ 「区域」を層化（ストラータ）変数として選択します。
- ▶ [次へ] をクリックし、[抽出方法] ステップで [次へ] をクリックします。

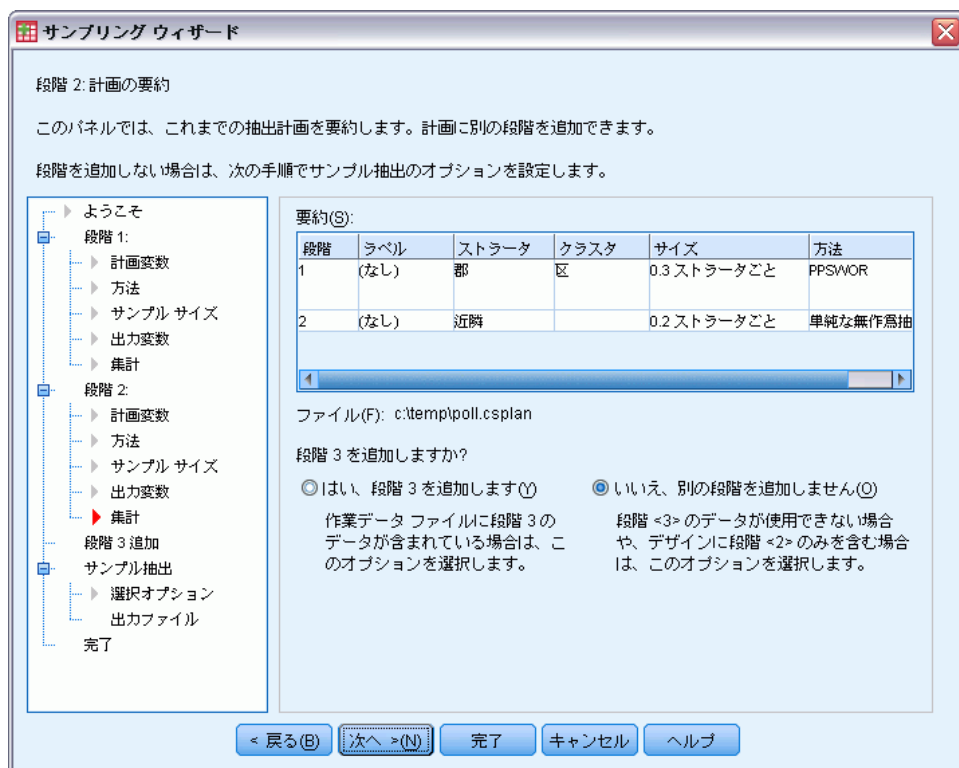
この計画構造は、第 1 段階（ステージ）で抽出された町の各区域に対して独立サンプルが抽出されることを意味します。この段階（ステージ）では、有権者は [置換を行わない単純な無作為抽出] により主要な抽出ユニットとして抽出されます。

図 13-38  
サンプリング ウィザードの [サンプル サイズ] ステップ (第 2 段階)



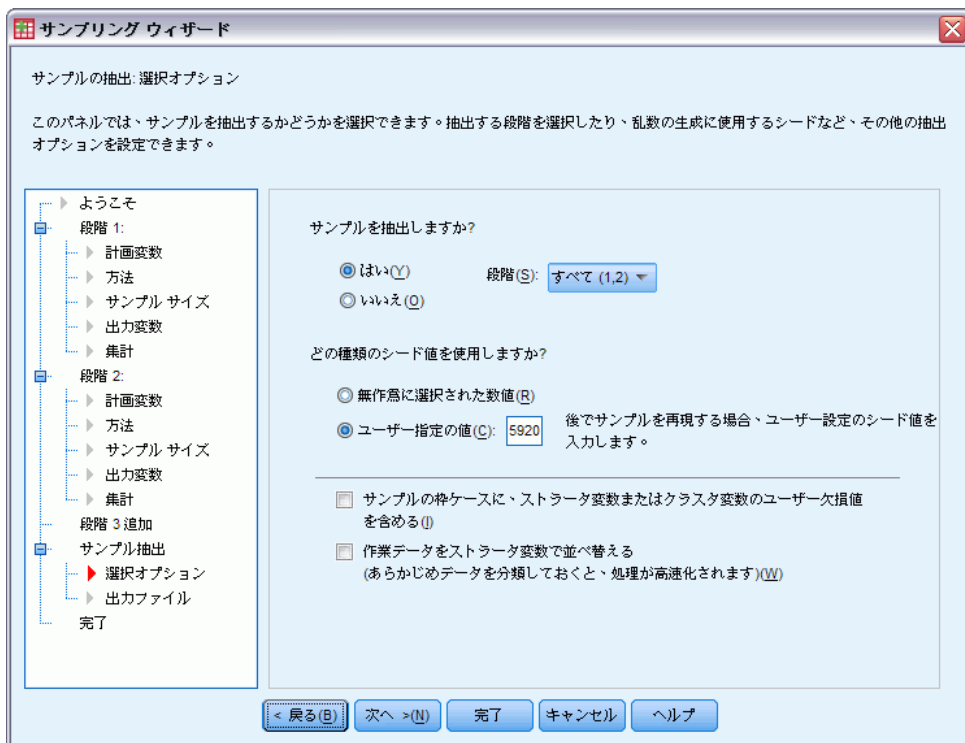
- ▶ [ユニット] ドロップダウン リストから、[比率] を選択します。
- ▶ 各ストラータからのサンプルに対するユニットの割合の値として、「0.2」と入力します。
- ▶ [次へ] をクリックし、[出力変数] ステップで [次へ] をクリックします。

図 13-39  
サンプリング ウィザードの [計画の要約] ステップ (第 2 段階)



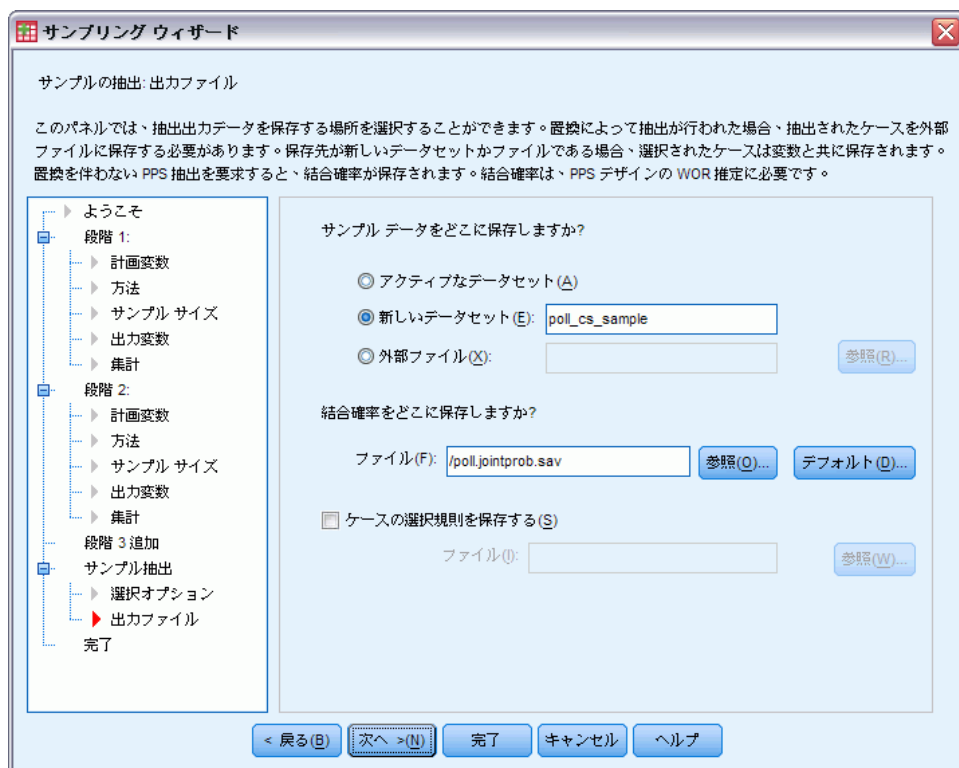
- ▶ 抽出計画を確認し、[次へ] をクリックします。

図 13-40  
サンプリング ウィザードの [サンプルの抽出: 選択オプション] ステップ



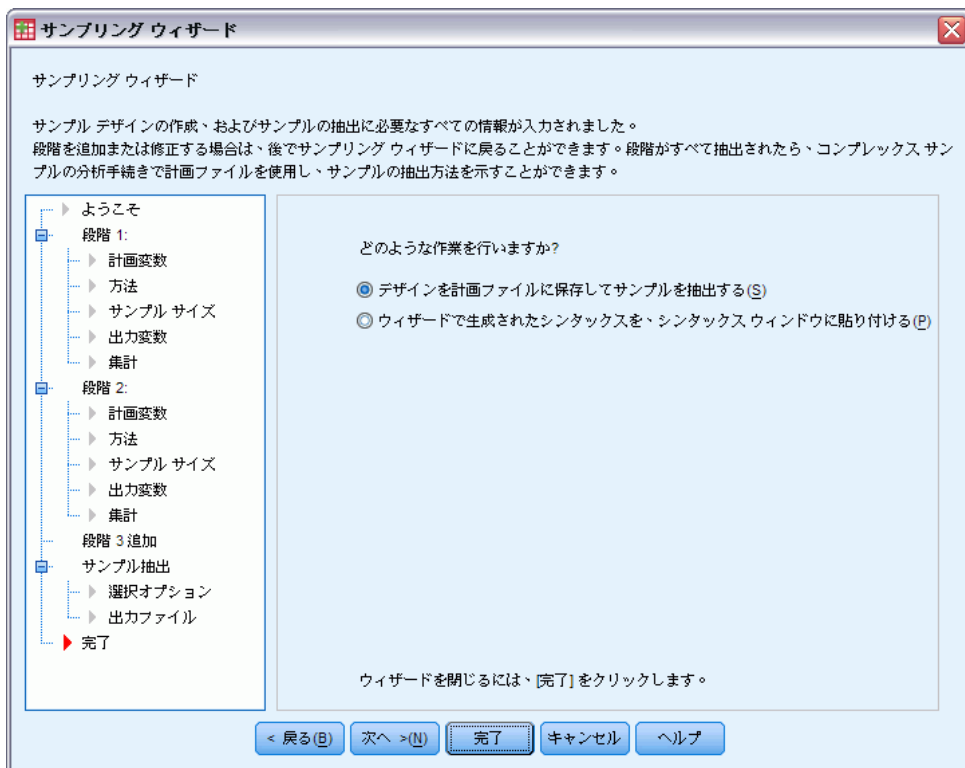
- ▶ 使用する乱数シードのタイプについて [ユーザー指定の値] を選択し、「592004」を値として入力します。  
カスタム値を使用すると、この例の結果をそのまま複製できます。
- ▶ [次へ] をクリックします。

図 13-41  
サンプリング ウィザードの [サンプルの抽出: 選択オプション] ステップ



- ▶ 新しいデータセットにサンプルを保存するように選択し、データセットの名前として「/poll\_cs\_sample」と入力します。
- ▶ 結合確率の保存場所を参照し、結合確率ファイルの名前として「poll\_jointprob.sav」と入力します。
- ▶ [次へ] をクリックします。

図 13-42  
サンプリング ウィザードの [完了] ステップ



- ▶ [完了] をクリックします。

以上の選択により、抽出計画ファイル `poll.esplan` が作成され、計画に基づいてサンプルが抽出されます。また、抽出結果が新しいデータセットの `poll_cs_sample` に保存され、結合確率ファイルが外部データ ファイルの `poll_jointprob.sav` に保存されます。



## 計画の要約

図 13-43  
計画の要約

			第 1 段階	第 2 段階
計画変数	ストラータ	1	County	Neighborhood
	クラスタ	1	Township	
サンプル 情報	選択方法		置換を行 わない、 PPS 抽出	置換を行 わない、単 純な無作 為抽出
	サイズの測定		データか ら取得	
	抽出されたユニットの比率		.3	.2
	抽出されたユニットの最小数		3	
	抽出されたユニットの最大数		5	
	作成または変更された変 数	段階に関連する包含 (選 択) 確率	Inclusion Probability _1_	Inclusion Probability _2_
		段階に関連する累積抽出 重み付け	Sample Weight Cumulativ e_1_	Sample Weight Cumulativ e_2_
分析情報	推定量の仮定		置換を行 わない、不 等確率抽 出 (結合 包含確率 を使用)	置換を行 わない、等 確率抽出
	包含確率		変数 Inclusion Probability _1_ から取 得	変数 Inclusion Probability _2_ から取 得

計画ファイル: C:\Program Files\SPSS\Tutorial\sample\_files\poll.csplan  
重み付け変数: SampleWeight\_Final\_

要約表は抽出計画を確認し、計画が固有の目的を表しているかを調べるために便利です。

## 抽出の要約

図 13-44  
段階の集計

County	抽出されたユニットの数		抽出されたユニットの比率	
	要求数	実際数	要求数	実際数
Eastern	4	4	30.0%	30.8%
Central	4	4	30.0%	30.8%
Western	3	3	30.0%	50.0%
Northern	5	5	30.0%	33.3%
Southern	3	3	30.0%	50.0%

計画ファイル: C:\Program Files\SPSS\Tutorial\sample\_files\poll.csplan

この要約表は、抽出の最初の段階（ステージ）を確認します。抽出が計画どおりに行われたかを調べるために便利です。郡ごとに 30% の町のサンプルを要求したことを思い出してください。西部郡と南部郡を除き、抽出された実際の割合は 30% に近くなります。これは、西部郡と南部郡に町がそれぞれ 6 つしかなく、郡ごとに最低 3 つの町を選択するよう指定したためです。

図 13-45  
段階の集計

County	Township	Neighborhood	抽出されたユニットの数		抽出されたユニットの比率		
			要求数	実際数	要求数	実際数	
Eastern	9	1	49	49	20.0%	19.9%	
		2	143	143	20.0%	20.0%	
		3	113	113	20.0%	20.0%	
		4	77	77	20.0%	20.0%	
		5	139	139	20.0%	20.0%	
		6	120	120	20.0%	20.0%	
	10	1	149	149	20.0%	20.1%	
		2	117	117	20.0%	20.0%	
		3	116	116	20.0%	20.0%	
		4	69	69	20.0%	19.9%	
	11	1	65	65	20.0%	19.9%	
		2	72	72	20.0%	19.9%	
		3	109	109	20.0%	20.0%	
		4	140	140	20.0%	20.0%	
		5	42	42	20.0%	19.8%	
		6	142	142	20.0%	20.0%	
	12	1	145	145	20.0%	20.1%	
		2	69	69	20.0%	20.1%	
		3	98	98	20.0%	20.1%	
		4	134	134	20.0%	20.0%	
		5	114	114	20.0%	20.0%	
		6	137	137	20.0%	19.9%	
	Central	2	1	119	119	20.0%	20.1%
			2	153	153	20.0%	19.9%
3			101	101	20.0%	20.0%	
4			52	52	20.0%	19.8%	
5			144	144	20.0%	20.0%	

この要約表（ここで示される最上位の部分）は、抽出の第 2 段階（ステージ）を確認します。これは、抽出が計画どおりに実行されたかどうかを調べる場合にも便利です。要求に応じて、第 1 段階（ステージ）で抽出された各町の各区域から有権者の約 20% が抽出されました。

## 抽出結果

図 13-46  
データ エディタの抽出結果

	有権者ID	近隣	区	郡	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	SampleWeight_Final_
376	368	4	9	1	0.44	2.26	0.20	11.28	11.26
377	369	4	9	1	0.44	2.26	0.20	11.28	11.26
378	374	4	9	1	0.44	2.26	0.20	11.28	11.26
379	376	4	9	1	0.44	2.26	0.20	11.28	11.26
380	379	4	9	1	0.44	2.26	0.20	11.28	11.26
381	380	4	9	1	0.44	2.26	0.20	11.28	11.26
382	382	4	9	1	0.44	2.26	0.20	11.28	11.26
383	13	5	9	1	0.44	2.26	0.20	11.26	11.26
384	18	5	9	1	0.44	2.26	0.20	11.26	11.26
385	23	5	9	1	0.44	2.26	0.20	11.26	11.26
386	38	5	9	1	0.44	2.26	0.20	11.26	11.26
387	39	5	9	1	0.44	2.26	0.20	11.26	11.26
388	40	5	9	1	0.44	2.26	0.20	11.26	11.26
389	41	5	9	1	0.44	2.26	0.20	11.26	11.26
390	43	5	9	1	0.44	2.26	0.20	11.26	11.26

新たに作成したデータセットで抽出結果を参照できます。新しい5つの変数が作業ファイルに保存され、各段階（ステージ）の包含確率や累積抽出重み付け、および最終的な抽出重み付けを示しています。サンプルに選択されなかった有権者は、このデータセットから除外されます。

最終的な抽出重み付けは、単純な無作為抽出方法に基づいて有権者が区域内で選択されるので、同一の区域内の有権者に対して同じになります。ただし、同一の町内の区域間では異なります。これは、抽出した割合がすべての区域で正確に 20% になっていないためです。

図 13-47  
データ エディタの抽出結果

	有権者ID	近隣	区	郡	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	SampleWeight_Final_
635	577	6	9	1	0.44	2.26	0.20	11.30	11.30
636	578	6	9	1	0.44	2.26	0.20	11.30	11.30
637	582	6	9	1	0.44	2.26	0.20	11.30	11.30
638	590	6	9	1	0.44	2.26	0.20	11.30	11.30
639	594	6	9	1	0.44	2.26	0.20	11.30	11.30
640	597	6	9	1	0.44	2.26	0.20	11.30	11.30
641	600	6	9	1	0.44	2.26	0.20	11.30	11.30
642	4	1	10	1	0.31	3.21	0.20	16.00	16.00
643	5	1	10	1	0.31	3.21	0.20	16.00	16.00
644	9	1	10	1	0.31	3.21	0.20	16.00	16.00
645	10	1	10	1	0.31	3.21	0.20	16.00	16.00
646	12	1	10	1	0.31	3.21	0.20	16.00	16.00
647	16	1	10	1	0.31	3.21	0.20	16.00	16.00
648	17	1	10	1	0.31	3.21	0.20	16.00	16.00
649	19	1	10	1	0.31	3.21	0.20	16.00	16.00

第 2 段階での有権者の場合と異なり、第 1 段階の抽出重み付けは、確率比例法に基づいて有権者が選択されるため、同一の郡内の町に対して同じではありません。

図 13-48  
[結合確率] ファイル

	郡	区	Unit_No_	Joint_Prob_1	Joint_Prob_2	Joint_Prob_3	Joint_Prob_4	Joint_Prob_5	var
1	1	10	1	0.31	0.10	0.11	0.12	.	.
2	1	11	2	0.10	0.39	0.15	0.16	.	.
3	1	9	3	0.11	0.15	0.44	0.21	.	.
4	1	12	4	0.12	0.16	0.21	0.48	.	.
5	2	12	1	0.22	0.04	0.07	0.08	.	.
6	2	6	2	0.04	0.23	0.07	0.08	.	.
7	2	7	3	0.07	0.07	0.41	0.19	.	.
8	2	2	4	0.08	0.08	0.19	0.45	.	.
9	3	5	1	0.58	0.31	0.32	.	.	.
10	3	3	2	0.31	0.61	0.36	.	.	.
11	3	4	3	0.32	0.36	0.63	.	.	.
12	4	14	1	0.26	0.06	0.06	0.07	0.09	.
13	4	8	2	0.06	0.29	0.07	0.08	0.10	.
14	4	4	3	0.06	0.07	0.29	0.08	0.10	.
15	4	2	4	0.07	0.08	0.08	0.33	0.12	.
16	4	13	5	0.09	0.10	0.10	0.12	0.43	.
17	5	3	1	0.74	0.25	0.27	.	.	.
18	5	6	2	0.25	0.41	0.13	.	.	.
19	5	4	3	0.27	0.13	0.43	.	.	.

poll\_jointprob.sav ファイルには、郡内の選択された町に対する第 1 段階の結合確率が含まれています。この表では、「郡」が第 1 段階のストラタ変数を、「町」がクラスタ変数を表しています。これらの変数を組み合わせると、すべての第 1 段階の PSU を一意に識別できます。PSU

は Unit\_No\_ によって各ストラータ内にラベル付けされ、Joint\_Prob\_1\_、Joint\_Prob\_2\_、Joint\_Prob\_3\_、Joint\_Prob\_4\_、および Joint\_Prob\_5\_ で使用されます。最初の 2 つのストラータにはそれぞれ 4 つの PSU があります。このため、結合包含確率行列は  $4 \times 4$  になり、Joint\_Prob\_5\_ 列が空白の列となっています。同様に、ストラータ 3 と 5 は  $3 \times 3$  結合包含確率行列に、ストラータ 4 は  $5 \times 5$  結合包含確率行列になっています。

結合確率ファイルの必要性は、結合包含確率行列の値を詳細に調べるとわかります。抽出方法が PPS WOR 推定 以外のとき、ある PSU の選択は他の PSU の選択から独立し、これらの PSU の結合包含確率は単純に包含確率の積となります。対照的に、[郡 1] の [町 9] と [町 10] の結合含有確率は約 0.11 (Joint\_Prob\_3\_ の上から 1 番目のケース、または Joint\_Prob\_1\_ の上から 3 番目のケースを参照)、または [町 9] と [町 10] の個別の包含確率の積 (Joint\_Prob\_1\_ の上から 1 番目のケースと Joint\_Prob\_3\_ の上から 3 番目のケースの積 ( $0.31 \times 0.44=0.1364$ )) 未満となります。

世論調査会社は、選択したサンプルを対象に対面式の調査を実施することになりました。調査結果が判明したら、サンプルを [コンプレックス サンプル] 分析手続きで処理し、抽出計画の poll.csplan を使用して抽出仕様を指定できます。また、poll\_jointprob.sav を使用して必要な結合包含確率を指定できます。

## 関連手続き

コンプレックス サンプルのサンプリング ウィザード手続きは、抽出計画ファイルの作成とサンプル抽出のための便利なツールです。

- 抽出計画ファイルにアクセスできない場合に分析用サンプルを準備するには、[分析準備ウィザード](#)を使用します。

# コンプレックス サンプルの分析 準備ウィザード

分析準備ウィザードでは、さまざまなコンプレックス サンプル分析手続きで使用される分析計画の作成と修正のステップを、順を追って行います。サンプル抽出に使用する抽出計画ファイルにアクセスできない場合には特に便利です。

## コンプレックス サンプルの分析準備ウィザードによる NHIS 一般使用データの準備

National Health Interview Survey (NHIS) は、米国国民を対象とした人口ベースの大規模な調査です。全国の代表的な世帯サンプルについて対面式で調査が行われます。各世帯のメンバーに関して、人口統計情報、健康に関する行動および状態の観測値が得られます。

2000 年の調査のサブセットは、nhis2000\_subset.sav に収集されています。詳細は、A 付録 p. 281 [サンプル ファイル](#) を参照してください。コンプレックス サンプルの分析準備ウィザードを使用し、このデータ ファイルの分析計画を作成して、データ ファイルを [コンプレックス サンプル] 分析手続きで処理できるようにします。

### ウィザードの使用方法

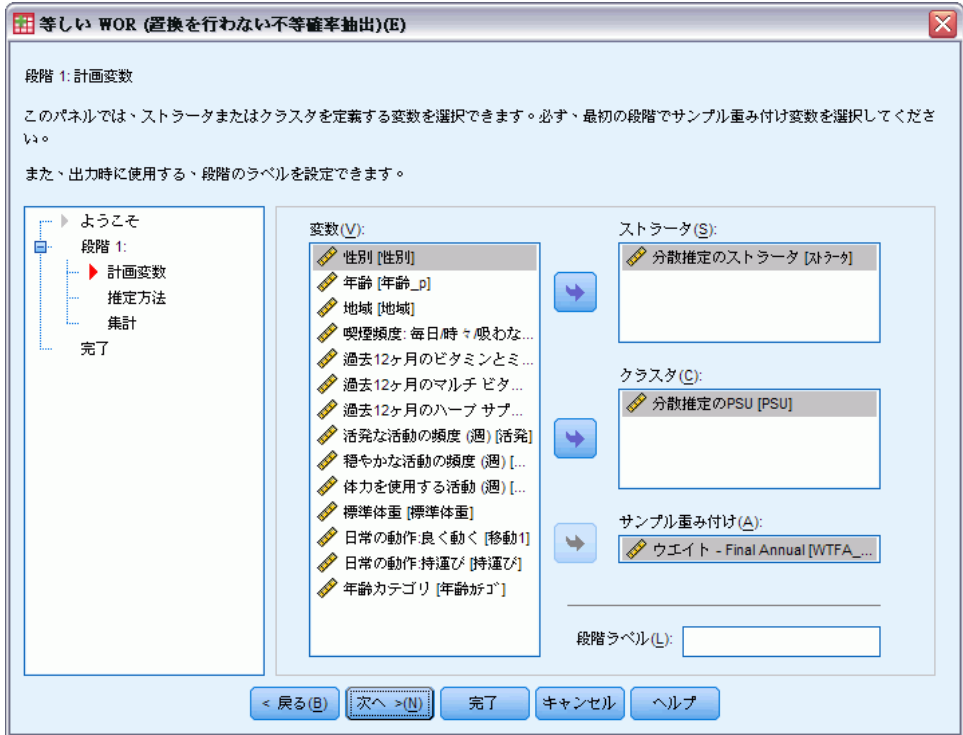
- ▶ コンプレックス サンプルの分析準備ウィザードを使用してサンプルを準備するには、メニューから次の項目を選択します。  
分析(A) > コンプレックス サンプル > 分析の準備...

図 14-1  
分析準備ウィザードの [ようこそ] ステップ



- ▶ 計画ファイルの保存場所を参照し、分析計画ファイルの名前として「nhis2000\_subset.csaplan」と入力します。
- ▶ [次へ] をクリックします。

図 14-2  
分析準備ウィザードの [計画変数] ステップ (第 1 段階)



データは複雑な複数段階サンプルで取得されますが、。エンドユーザーのために、元の NHIS 計画変数が変換され、計画変数および重み付け変数の単純なセットになっています。得られる結果は元の計画構造の場合とほぼ同じです。

- ▶ 「分散推定のストラータ」を層化（ストラータ）変数として選択します。
- ▶ 「分散推定の PSU」をクラスタ変数として選択します。
- ▶ 「ウエイト - Final Annual」をサンプル重み付け変数として選択します。
- ▶ [完了] をクリックします。



## 要約表

図 14-3  
要約表

			第1段階
計画変数	ストラータ	1	分散推定のストラータ
	クラスタ	1	分散推定のPSU
分析情報	推定量の仮定		置換を行う抽出

計画ファイル: C:\nhis2000\_subset.csaplan

重み付け変数: ウェイト - Final Annual

- a. 要求された推定量によって、第1段階では置換を行う抽出が仮定されます。その後の段階は、推定に使用されません。

要約表で、分析計画を確認できます。計画は、1つの層化（ストラータ）変数および1つのクラスタ変数を含んだ1つの段階（ステージ）で構成されています。置換を行う（WR）推定が使用され、計画は c:\nhis2000\_subset.csaplan に保存されます。これで、この計画ファイルを使用して [コンプレックス サンプル] 分析手続きで nhis2000\_subset.sav を処理できるようになりました。

## 抽出重み付けがデータ ファイル内に存在しない場合の分析準備

ある融資担当者はコンプレックス デザインに従って取得した顧客記録のコレクションを持っていますが、そのファイルには抽出重み付けが含まれていません。この情報は bankloan\_cs\_nowights.sav に含まれています。詳細は、A 付録 p.281 [サンプル ファイル](#) を参照してください。その融資担当者は抽出計画について知ることから始めて、コンプレックス サンプルの分析準備ウィザードを使用し、このデータ ファイルの分析計画を作成して、データ ファイルをコンプレックス サンプル分析手続きで処理できるようにすることを望んでいます。

その融資担当者は、顧客記録が2段階にわたって選択されたことを把握しています。まず、第1段階では、銀行支店100店舗の中から15店舗が等しい確率で選択され、置換は行われませんでした。第2段階では、それらの店舗から100人の顧客が等しい確率で選択され、置換は行われませんでした。店舗ごとの顧客数に関する情報は、データ ファイルに含まれています。分析計画を作成するための第1段階は、段階ごとの包含確率と最終的な抽出重み付けの計算です。

## 包含確率と抽出重み付けの計算

- ▶ 第1段階の包含確率を計算するには、メニューから次の項目を選択します。  
変換(T) > 変数の計算(C)...

図 14-4  
[変数の計算] ダイアログ ボックス



第 1 段階では銀行支店 100 店舗の中から 15 店舗が置換なしで選択されたため、特定の店舗が選択される確率は  $15 / 100 = 0.15$  です。

- ▶ 目標変数として「inclprob\_s1」と入力します。
- ▶ 数式として「0.15」と入力します。
- ▶ [OK] をクリックします。

図 14-5  
[変数の計算] ダイアログ ボックス



第 2 段階では各店舗から 100 人の顧客が選択されたため、特定の銀行の特定の顧客に関する包含確率は、「100 / 当該銀行の顧客数」となります。

- ▶ もう一度、[変数の計算] ダイアログ ボックスを表示します。
- ▶ 目標変数として「inclprob\_s2」と入力します。
- ▶ 数式として「100 / 顧客数」と入力します。
- ▶ [OK] をクリックします。

図 14-6  
[変数の計算] ダイアログ ボックス



これで段階ごとの包含確率がわかったので、最終的な抽出重み付けは簡単に計算できます。

- ▶ もう一度、[変数の計算] ダイアログ ボックスを表示します。
- ▶ 目標変数として「finalweight」と入力します。
- ▶ 数式として「1 / (inclprob\_s1 \* inclprob\_s2)」と入力します。
- ▶ [OK] をクリックします。

これで、分析計画を作成する準備が整いました。

## ウィザードの使用方法

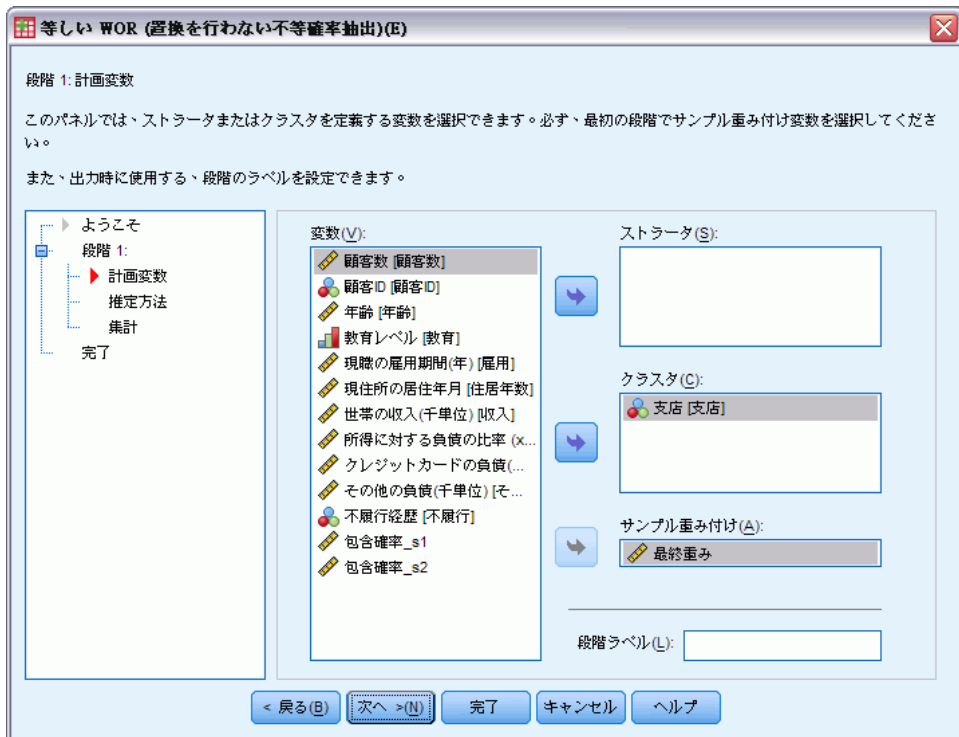
- ▶ コンプレックス サンプルの分析準備ウィザードを使用してサンプルを準備するには、メニューから次の項目を選択します。  
分析(A) > コンプレックス サンプル > 分析の準備...

図 14-7  
分析準備ウィザードの [ようこそ] ステップ



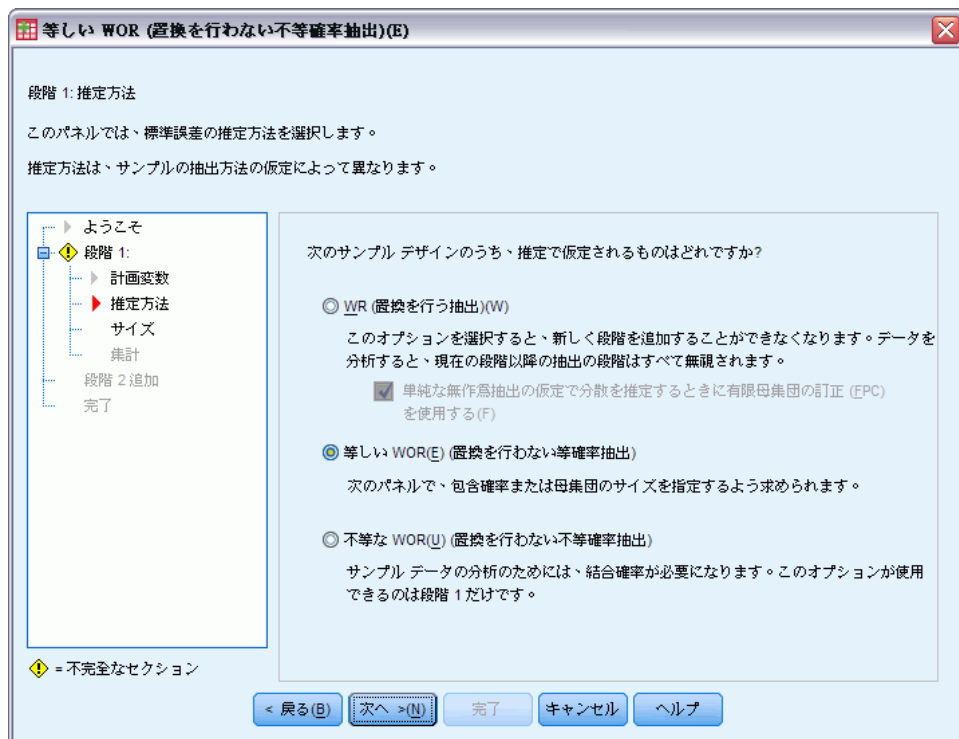
- ▶ 計画ファイルの保存場所を参照し、分析計画ファイルの名前として「bankloan.csplan」と入力します。
- ▶ [次へ] をクリックします。

図 14-8  
分析準備ウィザードの [計画変数] ステップ (第 1 段階)



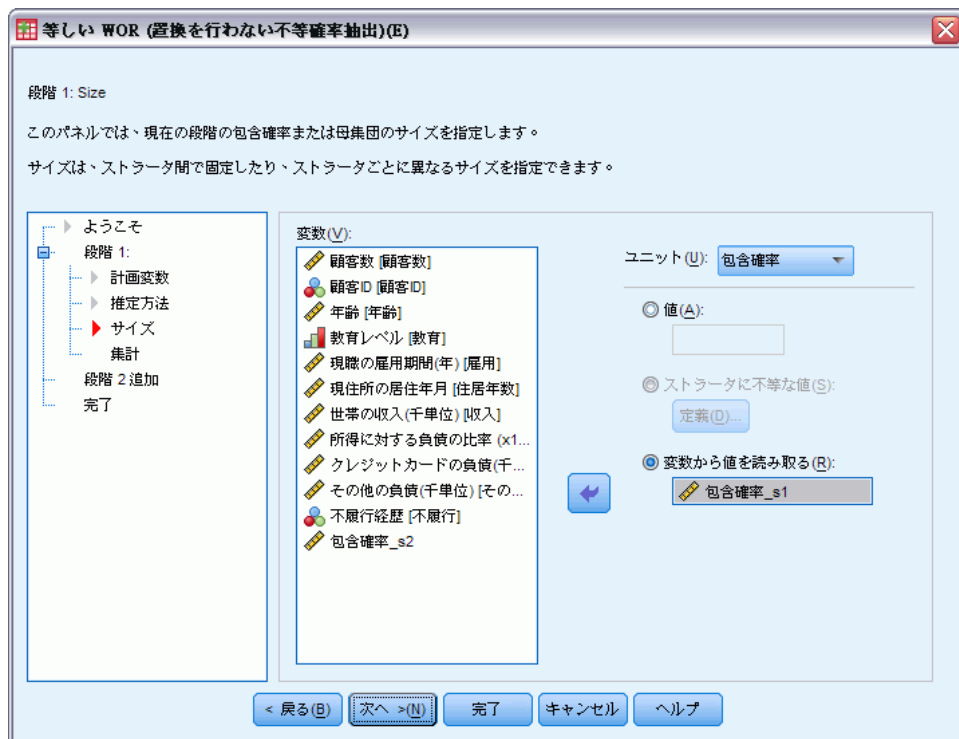
- ▶ 「支店」をクラスタ変数として選択します。
- ▶ 「finalweight」をサンプル重み付け変数として選択します。
- ▶ [次へ] をクリックします。

図 14-9  
分析準備ウィザードの [推定方法] ステップ (第 1 段階)



- ▶ 第 1 段階の推定方法として [等しい WOR] を選択します。
- ▶ [次へ] をクリックします。

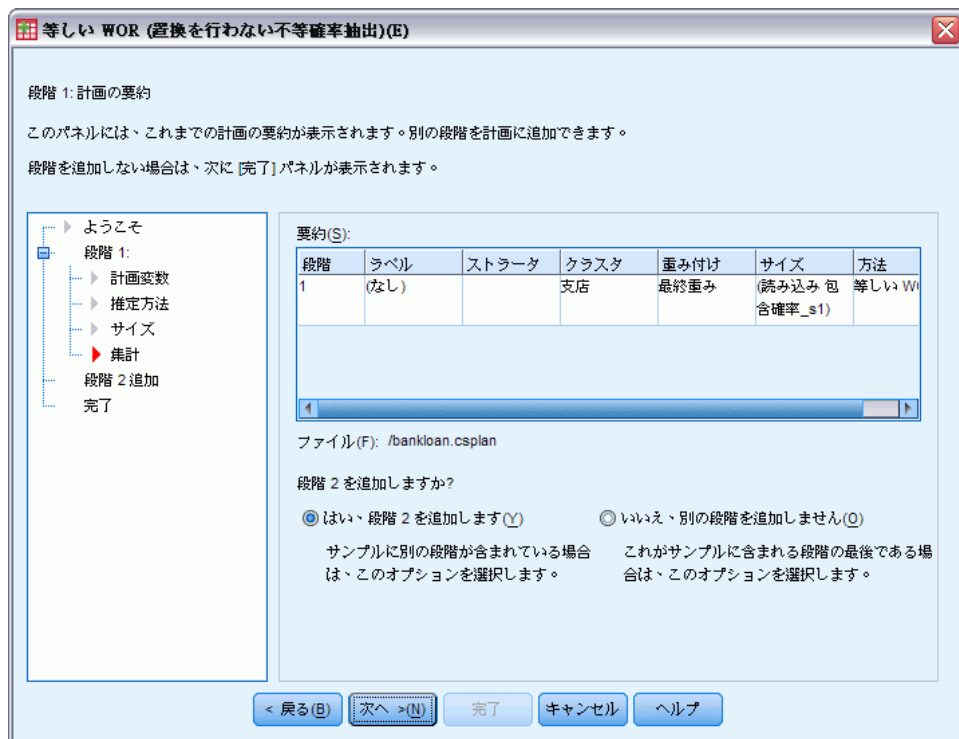
図 14-10  
分析準備ウィザードの [サイズ] ステップ (第 1 段階)



- ▶ [変数から値を読み取る] を選択し、第 1 段階の包含確率を含める変数として「inclprob\_s1」を選択します。
- ▶ [次へ] をクリックします。

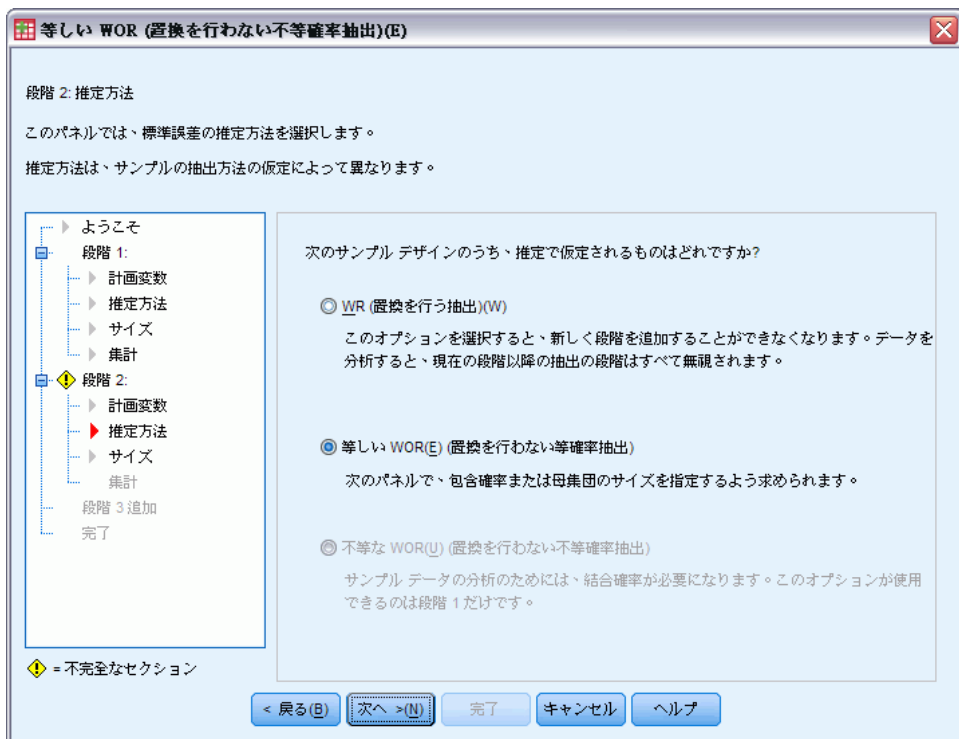


図 14-11  
分析準備ウィザードの [計画の要約] ステップ (第 1 段階)



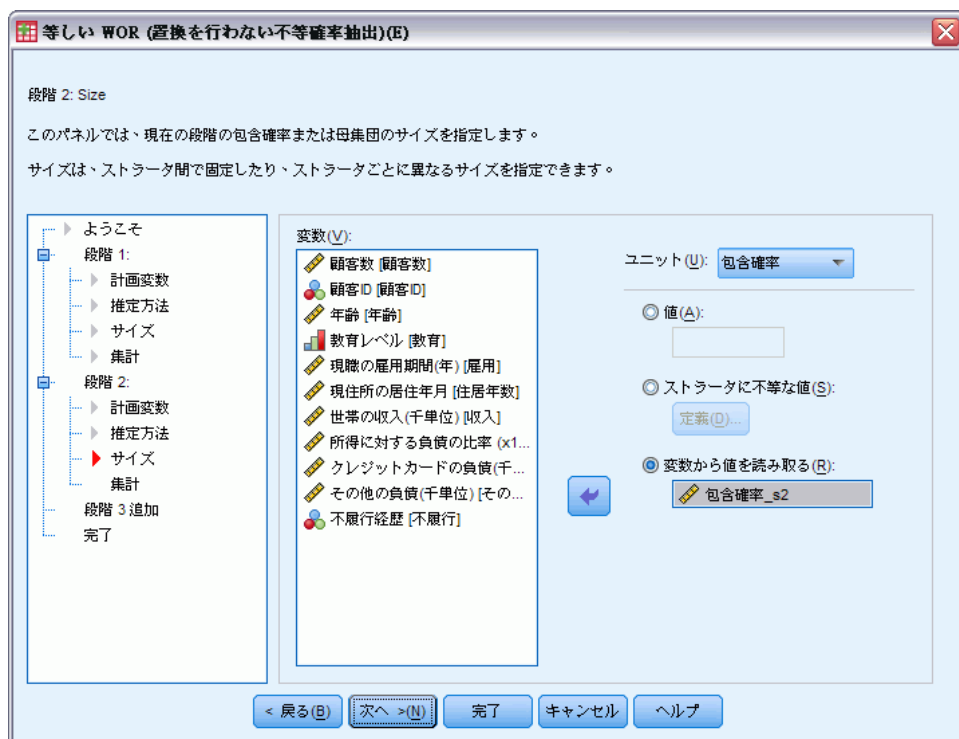
- ▶ [はい。段階 2 を追加します] を選択します。
- ▶ [次へ] をクリックし、[計画変数] ステップで [次へ] をクリックします。

図 14-12  
分析準備ウィザードの [推定方法] ステップ (第 2 段階)



- ▶ 第 2 段階の推定方法として [等しい WOR] を選択します。
- ▶ [次へ] をクリックします。

図 14-13  
分析準備ウィザードの [サイズ] ステップ (第 2 段階)



- ▶ [変数から値を読み取る] を選択し、第 2 段階の包含確率を含める変数として「inclprob\_s2」を選択します。
- ▶ [完了] をクリックします。

## 要約表

図 14-14  
集計表(U)

		第 1 段階	第 2 段階
計画変数	クラス	1	
分析情報	推定量の仮定	支店 置換を行 わない等 確率抽出 変数 inclprob s1 から取 得	置換を行 わない等 確率抽出 変数 inclprob s2 から取 得
	包含確率		

計画ファイル: C:\%bankloan.csaplans  
重み付け変数: finalweight

要約表で、分析計画を確認できます。計画は、1つのクラスター変数の計画を含む2つの段階から構成されます。置換を行わない(WOR)等確率推定が使用され、計画は `c:\¥bankloan.csaplan` に保存されます。これで、この計画ファイルを使用してコンプレックス サンプル分析手続きで(計算した包含確率と抽出重み付けを使用して) `bankloan_noweights.sav` を処理できるようになりました。

## 関連手続き

コンプレックス サンプルの分析準備ウィザード手続きは、抽出計画ファイルにアクセスできない場合に分析用サンプルを準備するための便利なツールです。

- 出計画ファイルを作成してサンプルを抽出するには、[サンプリングウィザード](#)を使用します。

# コンプレックス サンプル度数分布表

[コンプレックス サンプルの度数分布表] 手続きでは、選択した変数の度数分布表が作成され、1 変量の統計量が表示されます。1 つ以上のカテゴリ変数で定義したサブグループによって、統計を要求することもできます。

## コンプレックス サンプルの度数分布表による栄養補給剤使用の分析

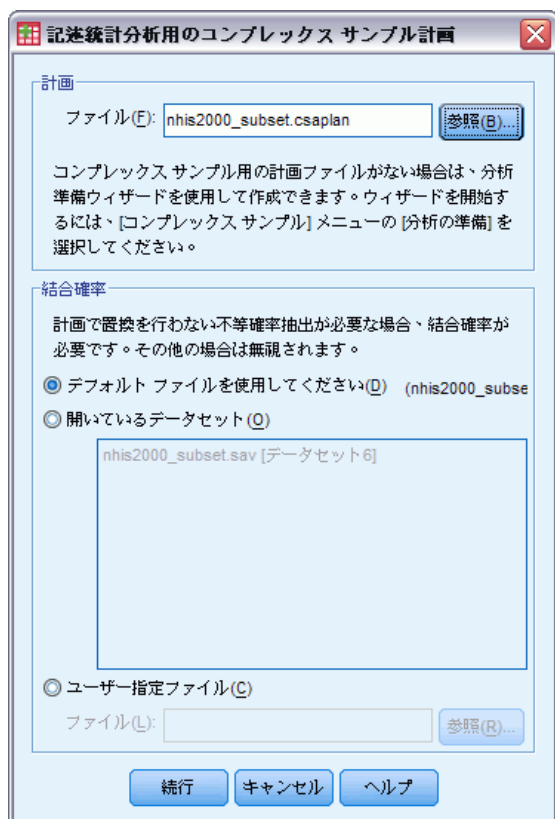
研究者は、National Health Interview Survey (NHIS) の結果、および事前に作成した分析計画を用いて、米国国民の栄養補給剤使用について調べたいと考えています。詳細は、14 章 p.152 [コンプレックス サンプルの分析準備ウィザードによる NHIS 一般使用データの準備](#) を参照してください。

2000 年の調査のサブセットは、nhis2000\_subset.sav に収集されています。分析計画は、nhis2000\_subset.csaplan に格納されています。詳細は、A 付録 p.281 [サンプル ファイル](#) を参照してください。コンプレックス サンプルの度数分布表を使用し、栄養補給剤使用の統計量を作成します。

### 分析の実行

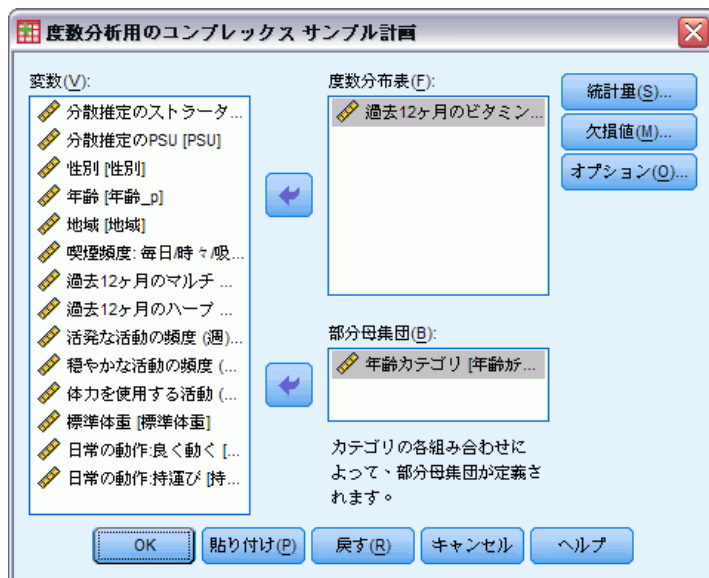
- ▶ コンプレックス サンプルの度数分布表の分析を実行するには、メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > 度数分布表...

図 15-1  
[コンプレックス サンプル計画] ダイアログ ボックス



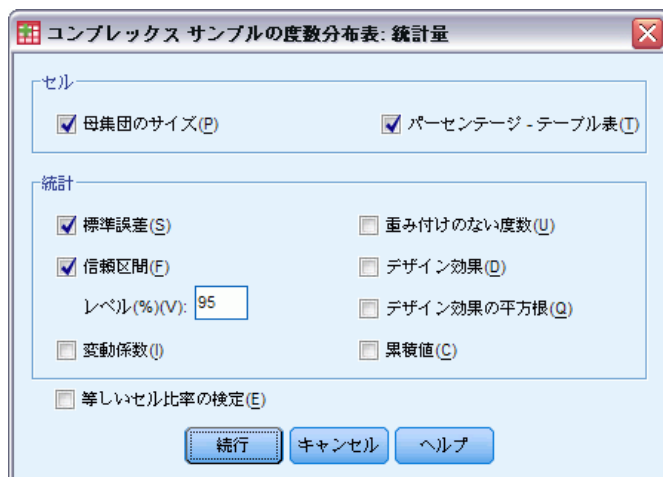
- ▶ nhis2000\_subset.csaplan を参照して選択します。詳細は、A 付録 p. 281 [サンプル ファイル](#) を参照してください。
- ▶ [続行] をクリックします。

図 15-2  
[度数分布表] ダイアログ ボックス



- ▶ 「過去 12 ヶ月のビタミンとミネラル サプリメントの摂取」を度数変数として選択します。
- ▶ 「年齢カテゴリ」を部分母集団変数として選択します。
- ▶ [統計量] をクリックします。

図 15-3  
[度数分布表: 統計] ダイアログ ボックス



- ▶ [セル] グループの [パーセンテージ - テーブル表] を選択します。
- ▶ [統計量] グループの [信頼区間] を選択します。

- ▶ [続行] をクリックします。
- ▶ [度数分布表] ダイアログ ボックスで [OK] をクリックします。

## 度数分布表

図 15-4  
変数/状況の度数分布表

		推定値	標準誤差	95% 信頼区間	
				下段	上段
母集団のサイズ	はい	1.03E+08	1185126.7	1.00E+08	1.05E+08
	いいえ	9.08E+07	1094401.9	8.86E+07	9.29E+07
	合計 (ピボットテーブル)	1.94E+08	1789098.7	1.90E+08	1.97E+08
合計数に対する割合 (%)	はい	53.1%	.4%	52.4%	53.8%
	いいえ	46.9%	.4%	46.2%	47.6%
	合計 (ピボットテーブル)	100.0%	.0%	100.0%	100.0%

選択された各統計量は、それぞれに指定されたセルの測定方法で計算されます。最初の列には、ビタミン/ミネラル栄養補給剤を使用する/使用しない母集団の数とパーセンテージの推定値が含まれています。信頼区間は重複していません。つまり、米国国民はビタミン/ミネラル栄養補給剤を使用している人々のほうが多いと結論付けることができます。

## 部分母集団による度数分布表

図 15-5  
部分母集団による度数分布表

年齢カテゴリー		推定値	標準誤差	95% 信頼区間		
				下段	上段	
18-24	母集団のサイズ	はい	1.00E+07	350602.352	9.33E+06	1.07E+07
		いいえ	1.55E+07	499182.391	1.45E+07	1.65E+07
		合計 (ピボットテーブル)	2.55E+07	680732.812	2.42E+07	2.68E+07
	合計数に対する割合 (%)	はい	39.3%	1.0%	37.4%	41.2%
		いいえ	60.7%	1.0%	58.8%	62.6%
		合計 (ピボットテーブル)	100.0%	.0%	100.0%	100.0%
25-44	母集団のサイズ	はい	3.92E+07	660855.719	3.79E+07	4.05E+07
		いいえ	3.95E+07	645934.187	3.82E+07	4.08E+07
		合計 (ピボットテーブル)	7.87E+07	961114.325	7.68E+07	8.06E+07
	合計数に対する割合 (%)	はい	49.8%	.6%	48.7%	50.9%
		いいえ	50.2%	.6%	49.1%	51.3%
		合計 (ピボットテーブル)	100.0%	.0%	100.0%	100.0%
45-64	母集団のサイズ	はい	3.42E+07	598603.728	3.30E+07	3.53E+07
		いいえ	2.40E+07	497723.833	2.30E+07	2.50E+07
		合計 (ピボットテーブル)	5.82E+07	814680.415	5.66E+07	5.98E+07
	合計数に対する割合 (%)	はい	58.7%	.6%	57.5%	60.0%
		いいえ	41.3%	.6%	40.0%	42.5%
		合計 (ピボットテーブル)	100.0%	.0%	100.0%	100.0%
65+	母集団のサイズ	はい	1.94E+07	439459.793	1.86E+07	2.03E+07
		いいえ	1.18E+07	314238.078	1.12E+07	1.24E+07
		合計 (ピボットテーブル)	3.12E+07	587623.439	3.01E+07	3.24E+07
	合計数に対する割合 (%)	はい	62.2%	.7%	60.7%	63.6%
		いいえ	37.8%	.7%	36.4%	39.3%
		合計 (ピボットテーブル)	100.0%	.0%	100.0%	100.0%



部分母集団による統計量を計算する場合は、「年齢カテゴリ」の値に基づき、指定されたセルの測定方法で計算されます。最初の列には、ビタミン/ミネラル栄養補給剤を使用する/使用しない各カテゴリの母集団の数とパーセンテージの推定値が含まれています。表パーセントの信頼区間はすべて重複していません。つまり、ビタミン/ミネラル栄養補給剤の使用は年齢とともに増加すると結論付けることができます。

## 集計 (報告書 データ列)

[コンプレックス サンプルの度数分布表] 手続きを使用して、米国国民の栄養補給剤使用の統計量を作成しました。

- 全体的に、ビタミン/ミネラル栄養補給剤を使用する割合が高い。
- 年齢カテゴリ別に見た場合、年齢が高くなるにつれてビタミン/ミネラル栄養補給剤を使用する割合が増大する。

## 関連手続き

[コンプレックス サンプルの度数分布表] 手続きは、コンプレックス サンプル計画で得られた観測値に関するカテゴリ変数の 1 変量の記述統計量を取得するための便利なツールです。

- [コンプレックス サンプルのサンプリング ウィザード](#)は、コンプレックス サンプル計画仕様の指定およびサンプル抽出に使用されます。サンプリング ウィザードにより作成される抽出計画ファイルは、デフォルト分析計画を含んでおり、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- [コンプレックス サンプルの分析準備ウィザード](#)は、既存のコンプレックス サンプル用の分析仕様の設定に使用されます。サンプリング ウィザードにより作成される分析計画ファイルは、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- [\[コンプレックス サンプルのクロス集計\]](#) 手続きでは、カテゴリ変数のクロス集計に関する記述統計量が提供されます。
- [\[コンプレックス サンプルの記述統計\]](#) 手続きでは、スケール変数に関する 1 変量の記述統計量が提供されます。

# コンプレックス サンプル記述統計

[コンプレックス サンプルの記述統計] 手続きでは、複数の変数に関する 1 変量の要約統計量が表示されます。1 つ以上のカテゴリ変数で定義したサブグループによって、統計を要求することもできます。

## コンプレックス サンプルの記述統計による運動水準の分析

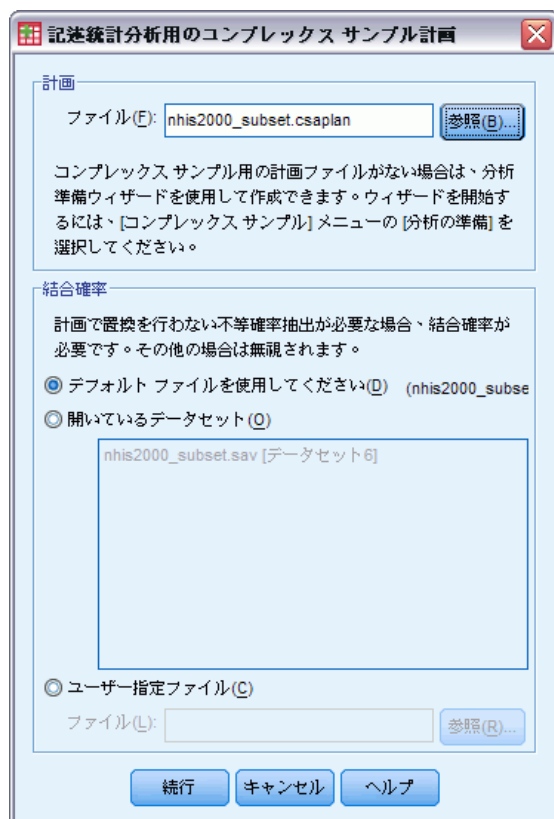
研究者は、National Health Interview Survey (NHIS) の結果、および事前に作成した分析計画を用いて、米国国民の運動水準を調べたいと考えています。詳細は、14 章 p.152 [コンプレックス サンプルの分析準備ウィザードによる NHIS 一般使用データの準備](#) を参照してください。

2000 年の調査のサブセットは、nhis2000\_subset.sav に収集されています。分析計画は、nhis2000\_subset.csaplan に格納されています。詳細は、A 付録 p.281 [サンプル ファイル](#) を参照してください。[コンプレックス サンプルの記述統計] 手続きでは、運動水準に関する 1 変量の記述統計量が提供されます。

### 分析の実行

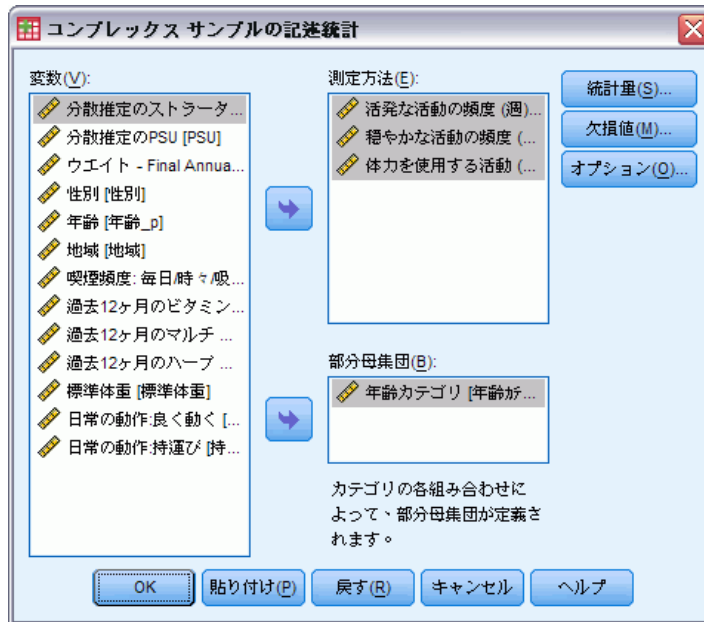
- ▶ コンプレックス サンプルの記述統計の分析を実行するには、メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > 記述統計...

図 16-1  
[コンプレックス サンプル計画] ダイアログ ボックス



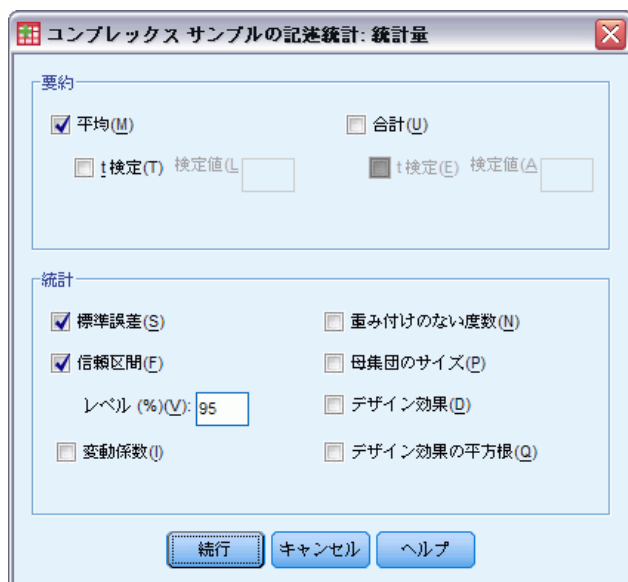
- ▶ nhis2000\_subset.csaplan を参照して選択します。詳細は、A 付録 p. 281 [サンプル ファイル](#) を参照してください。
- ▶ [続行] をクリックします。

図 16-2  
[記述統計] ダイアログ ボックス



- ▶ 「活発な活動の頻度 (週)」から「体力を使用する活動 (週)」を測定変数として選択します。
- ▶ 「年齢カテゴリ」を部分母集団変数として選択します。
- ▶ [統計量] をクリックします。

図 16-3  
[記述統計] ダイアログ ボックス



- ▶ [統計量] グループの [信頼区間] を選択します。
- ▶ [続行] をクリックします。
- ▶ [コンプレックス サンプルの記述統計] ダイアログ ボックスで、[OK] をクリックします。

## 1 変量統計

図 16-4  
1 変量統計

	推定値	標準誤差	95% 信頼区間	
			下段	上段
平均値 (ラン検定)				
活発な活動の頻度 (週)	3.73	.033	3.66	3.79
穏やかな活動の頻度 (週)	4.90	.041	4.82	4.98
体力を使用する活動 (週)	3.52	.042	3.43	3.60

選択された各統計量は、それぞれの測定変数で計算されます。最初の列には、各人が特定のタイプの運動を行う週ごとの平均回数の推定値が含まれています。平均値の信頼区間は重複していません。つまり、全体的に米国民は、体力増強運動よりも激しい運動を行う頻度が高く、激しい運動よりも適度な運動を行う頻度が高いと結論付けることができます。

## 部分母集団による 1 変量統計

図 16-5  
部分母集団による 1 変量統計

年齢カテゴリ			推定値	標準誤差	95% 信頼区間	
					下段	上段
18-24	平均値 (ラン検定)	活発な活動の頻度 (週)	3.92	.087	3.75	4.09
		穏やかな活動の頻度 (週)	5.18	.137	4.91	5.45
		体力を使用する活動 (週)	3.45	.085	3.28	3.62
25-44	平均値 (ラン検定)	活発な活動の頻度 (週)	3.55	.048	3.46	3.65
		穏やかな活動の頻度 (週)	4.73	.056	4.62	4.84
		体力を使用する活動 (週)	3.28	.052	3.18	3.38
45-64	平均値 (ラン検定)	活発な活動の頻度 (週)	3.79	.063	3.66	3.91
		穏やかな活動の頻度 (週)	4.88	.070	4.74	5.02
		体力を使用する活動 (週)	3.65	.092	3.47	3.84
65+	平均値 (ラン検定)	活発な活動の頻度 (週)	4.18	.111	3.96	4.39
		穏やかな活動の頻度 (週)	5.22	.084	5.06	5.39
		体力を使用する活動 (週)	4.66	.155	4.36	4.97

選択された各統計量は、「年齢カテゴリ」の値に基づき、それぞれの測定変数で計算されます。最初の列には、各カテゴリの人々が特定のタイプの運動を行う週ごとの平均回数の推定値が含まれています。平均値の信頼区間により、興味深い結論を導き出すことができます。

- 激しい運動と適度な運動に関しては、25～44才の人々よりも18～24才および45～64才の人々の方が積極的であり、45～64才の人々よりも65才以上の人々の方が積極的です。
- 体力増強運動に関しては、25～44才の人々よりも45～64才の人々の方が積極的であり、18～24才および45～64才の人々よりも65才以上の人々の方が積極的です。

## 集計 (報告書 データ列)

[コンプレックス サンプルの記述統計] 手続きを使用して、米国国民の運動水準の統計量が得られました。

- 全体的に、運動のタイプによって参加時間数が異なります。
- 年齢ごとに見た場合、社会人になってから最初のうちは学生時代ほど積極的ではありませんが、年齢を重ねるに従い、運動に真剣に取り組む傾向があります。

## 関連手続き

[コンプレックス サンプルの記述統計] 手続きは、複合抽出計画で得られた観測値に関する尺度変数の 1 変量の記述統計量を取得するための便利なツールです。

- **コンプレックス サンプルのサンプリング ウィザード**は、コンプレックス サンプル計画仕様の指定およびサンプル抽出に使用されます。サンプリング ウィザードにより作成される抽出計画ファイルは、デフォルト分析計画を含んでおり、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- **コンプレックス サンプルの分析準備ウィザード**は、既存のコンプレックス サンプル用の分析仕様の設定に使用されます。サンプリング ウィザードにより作成される分析計画ファイルは、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- **[コンプレックス サンプルの比率分析]** 手続きでは、尺度変数の比率分析に関する記述統計量が提供されます。
- **[コンプレックス サンプルの度数分布表]** 手続きでは、カテゴリ変数に関する 1 変量の記述統計量が提供されます。

# コンプレックス サンプルのクロス集計表

[コンプレックス サンプルのクロス集計] 手続きでは、選択した変数組のクロス集計表が生成され、2次元統計量が表示されます。1つ以上のカテゴリ変数で定義したサブグループによって、統計を要求することもできます。

## コンプレックス サンプルのクロス集計によるイベントの相対リスクの測定

雑誌を定期購読で販売する企業は、従来より、購入した顧客名データベースに基づいて毎月ダイレクトメールを送っています。一般に回答率が低いため、購入の可能性が高い顧客を対象を絞り込む方法を見つける必要があります。そこで、新聞を読む人々の方が雑誌を購読する可能性が高いと考え、ダイレクトメールを新聞購読者に発送する方法が提案されました。

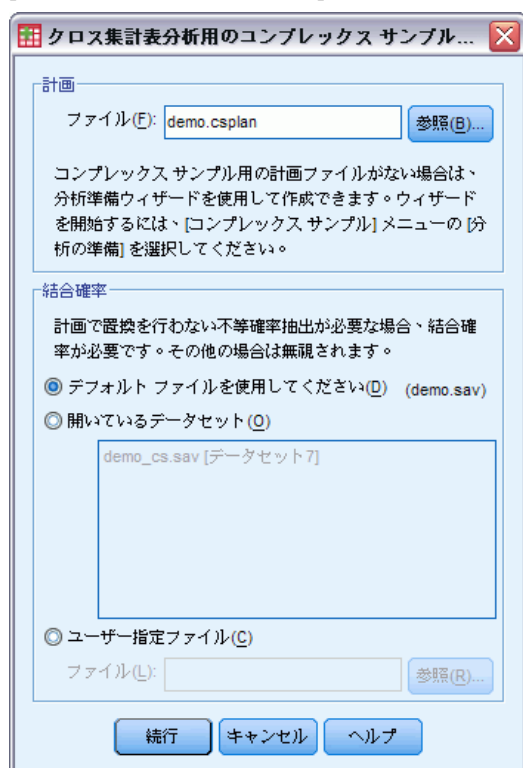
[コンプレックス サンプルのクロス集計] 手続きを使用してこの方法をテストするには、[新聞の購読] と [回答] による  $2 \times 2$  の表を作成し、各新聞購読者がダイレクトメールに回答する相対リスクを計算します。この情報は demo\_cs.sav に収集され、抽出計画ファイル demo.csplan を使用して分析する必要があります。詳細は、A 付録 p.281 [サンプル ファイル](#) を参照してください。

## 分析の実行

- ▶ コンプレックス サンプルのクロス集計の分析を実行するには、メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > クロス集計表...

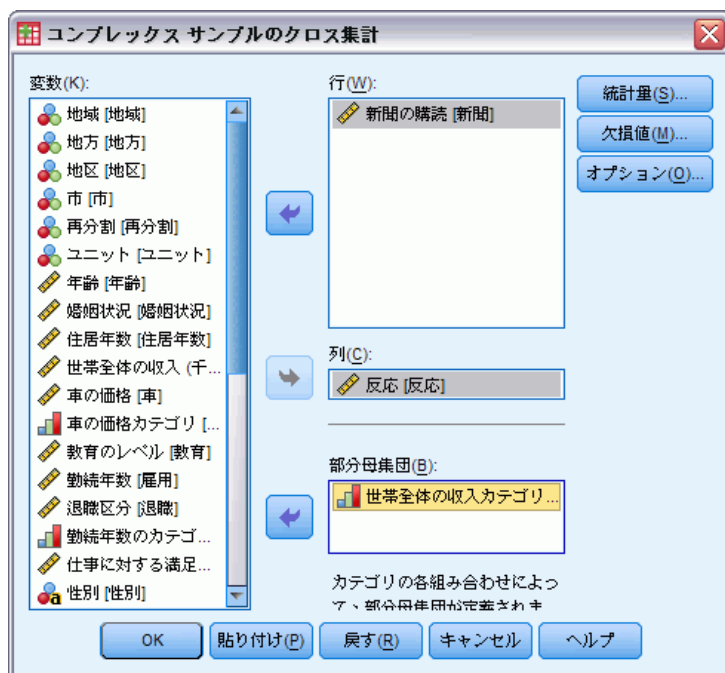


図 17-1  
[コンプレックス サンプル計画] ダイアログ ボックス



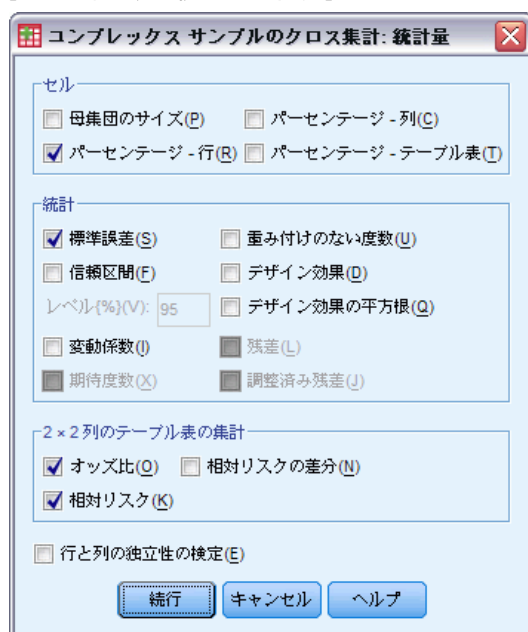
- ▶ demo.csplan を参照して選択します。詳細は、A 付録 p. 281 サンプル ファイル を参照してください。
- ▶ [続行] をクリックします。

図 17-2  
[クロス集計表] ダイアログ ボックス



- ▶ 「新聞の購読」を行変数として選択します。
- ▶ 「回答」を列変数として選択します。
- ▶ 収入カテゴリごとの結果も確認する必要があるので、「世帯全体の収入カテゴリ (千ドル)」を部分母集団変数として選択します。
- ▶ [統計量] をクリックします。

図 17-3  
[クロス集計表: 統計量の指定] ダイアログ ボックス



- ▶ [セル] グループの [母集団のサイズ] の選択を解除し、[パーセンテージ - 行] を選択します。
- ▶ [2 × 2 列のテーブル表の集計] グループの [オッズ比] および [相対リスク] を選択します。
- ▶ [続行] をクリックします。
- ▶ [コンプレックス サンプルのクロス集計] ダイアログ ボックスで、[OK] をクリックします。

以上の選択により、「新聞の購読」と「回答」のクロス集計表およびリスク推定値が作成されます。「世帯全体の収入カテゴリ (千ドル)」ごとの結果を示す別の表も作成されます。

## クロス表

図 17-4  
新聞購読者と回答のクロス集計

新聞の購読			回答		
			あり	なし	合計 (ピボットテーブル)
購読している	新聞の購読内での割合 (%)	推定値	18.6%	81.4%	100.0%
		標準誤差	1.2%	1.2%	.0%
購読していない	新聞の購読内での割合 (%)	推定値	10.6%	89.4%	100.0%
		標準誤差	.7%	.7%	.0%
合計 (ピボットテーブル)	新聞の購読内での割合 (%)	推定値	13.5%	86.5%	100.0%
		標準誤差	.7%	.7%	.0%

このクロス集計表は、全体的にダイレクトメールに回答している人々が少ないことを示しています。ただし、新聞購読者による回答の割合は高くなっています。

## リスク推定値

図 17-5  
新聞購読者と回答のリスク推定値

新聞の購読 * 回答			推定値
オッズ比			1.912
相対リスク	コーホートに対して回答 = あり		1.743
ク	コーホートに対して回答 = なし		.911

統計量が計算されるのは、2 × 2 列のテーブルですべてのセルが観測されている場合だけです。

相対リスクは、イベントの確率の比率です。ダイレクトメールへの回答の相対リスクは、非新聞購読者が回答する確率に対する、新聞購読者が回答する確率の比率です。つまり、相対リスクの推定値は単純に  $17.2\%/10.3\% = 1.673$  となります。同様に、無回答の相対リスクは、非購読者が回答しない確率に対する、購読者が回答しない確率の比率です。この相対リスクの推定値は、0.923 です。この結果から、ダイレクトメールに回答する新聞購読者は非新聞購読者の 1.673 倍で、ダイレクトメールに回答しない新聞購読者は非新聞購読者の 0.923 倍であることが推定できます。

オッズ比は、イベントオッズの比率です。イベントのオッズは、イベントが発生しない確率に対する、イベントが発生する確率の比率です。つまり、新聞購読者がダイレクトメールに回答するオッズの推定値は、 $17.2\%/82.8\% = 0.208$  となります。同様に、非購読者が回答するオッズの推定値は、 $10.3\%/89.7\% = 0.115$  となります。したがって、オッズ比の推定値は、 $0.208/0.115 = 1.812$  となります（この手順を進める中で若干の丸め誤差が生じることに注意してください）。このオッズ比は、非回答の相対リスクに対する、回答の相対リスクの比率 ( $1.673/0.923 = 1.812$ ) でもあります。

## オッズ比と相対リスク

オッズ比は比率の比率であり、理解するのが非常に困難です。相対リスクの方が理解しやすく、オッズ比単独ではあまり有用とはいえません。ただし、特定の状況において相対リスクの推定値があまり適切でないことがあり、その場合はオッズ比を使用して対象イベントの相対リスクを概算できます。次の条件が2つとも適合する場合に、対象イベントの相対リスクの近似値としてオッズ比を使用します。

- 該当イベントの確率が小さい ( $< 0.1$ )。この条件では、オッズ比が相対リスクの適切な近似値になることが確実です。この例の場合、対象イベントはダイレクトメールへの回答です。
- 研究計画がケースコントロールである。この条件は、相対リスクの通常の推定値が適切でないことを示しています。ケースコントロール研究は回顧的であり、たいていは、対象イベントが発生しそうな場合や、見込み実験の計画が非現実的または非倫理的である場合に使用されます。

この例では回答者全体の割合が 12.8% であり、研究計画がケースコントロールではないため、いずれの条件も適合しないので、オッズ比の値ではなく相対リスクの 1.673 を報告する方が安全です。

## 部分母集団によるリスク推定値

図 17-6  
新聞購読者と回答のリスク推定値、収入カテゴリのための制御

家族全体の収入			推定値
25未満	新聞の購読 * 回答	オッズ比	2.608
		相対リスク	2.149
		コーホートに対して回答 = あり コーホートに対して回答 = なし	.824
25 - 49	新聞の購読 * 回答	オッズ比	1.923
		相対リスク	1.737
		コーホートに対して回答 = あり コーホートに対して回答 = なし	.903
50 - 74	新聞の購読 * 回答	オッズ比	1.538
		相対リスク	1.493
		コーホートに対して回答 = あり コーホートに対して回答 = なし	.971
75以上	新聞の購読 * 回答	オッズ比	1.200
		相対リスク	1.182
		コーホートに対して回答 = あり コーホートに対して回答 = なし	.985

統計量が計算されるのは、 $2 \times 2$  列のテーブルですべてのセルが観測されている場合だけです。

相対リスクの推定値は、各収入カテゴリごとに個別に計算されます。新聞購読者による肯定的な回答の相対リスクが、収入が増えるにつれて徐々に下がっていくことに注意してください。この結果に基づき、ダイレクトメールの対象をさらに絞り込むことができます。

## 集計 (報告書 データ列)

コンプレックス サンプルのクロス集計のリスク推定値を使用した結果、ダイレクトメールの対象を新聞購読者に絞ることにより回答率を上げられることがわかりました。また、リスク推定値が収入カテゴリ全体で一定ではないという証拠もいくつか得られました。収入が低い新聞購読者を対象とすることにより、さらに回答率を向上させる可能性があります。

## 関連手続き

[コンプレックス サンプルのクロス集計] 手続きは、コンプレックス サンプル計画で得られた観測値に関するカテゴリ変数のクロス集計の記述統計量を取得するための便利なツールです。

- **コンプレックス サンプルのサンプリング ウィザード**は、コンプレックス サンプル計画仕様の指定およびサンプル抽出に使用されます。サンプリング ウィザードにより作成される抽出計画ファイルは、デフォルト分析計画を含んでおり、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- **コンプレックス サンプルの分析準備ウィザード**は、既存のコンプレックス サンプル用の分析仕様の設定に使用されます。サンプリング ウィザードにより作成される分析計画ファイルは、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- **[コンプレックス サンプルの度数分布表]** 手続きでは、カテゴリ変数に関する 1 変量の記述統計量が提供されます。

# コンプレックス サンプル比率

[コンプレックス サンプルの比率分析] 手続きでは、変数の比率に関する 1 変量の要約統計量が表示されます。1 つ以上のカテゴリ変数で定義したサブグループによって、統計を要求することもできます。

## コンプレックス サンプルの比率分析による財産価値評価の支援

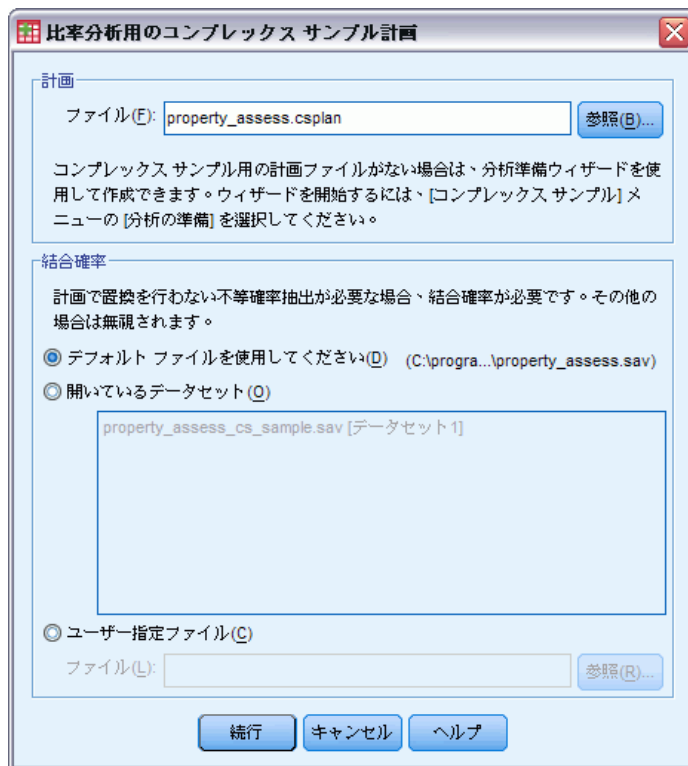
州当局は、財産税の評価が各郡の間で公正に行われていることを確認する責任があります。税は財産の評価額に基づいているため、各郡の財産価値を追跡し、各郡のレコードが同等の最新状態であることを確認する必要があります。現在の評価額を抽出するリソースは限られているため、州当局は財産の選択に複合抽出方式を用いることを決定しました。

選択した財産のサンプルと現在の評価額の情報は、`property_assess_cs_sample.sav` に収集されています。詳細は、A 付録 p. 281 [サンプル ファイル](#) を参照してください。コンプレックス サンプルの比率分析を使用し、5 つの郡について、最終評価額以降の財産価値の変化を評価します。

### 分析の実行

- ▶ コンプレックス サンプルの比率分析を実行するには、メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > 比率...

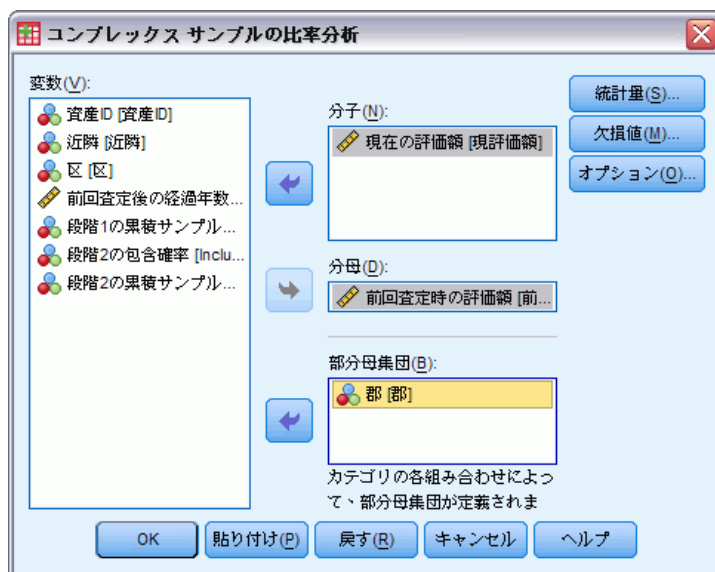
図 18-1  
[コンプレックス サンプル計画] ダイアログ ボックス



- ▶ property\_assess.csplan を参照して選択します。詳細は、A 付録 p.281 サンプル ファイル を参照してください。
- ▶ [続行] をクリックします。

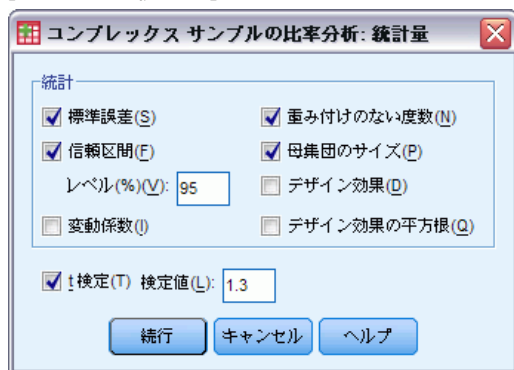


図 18-2  
[比率分析] ダイアログ ボックス



- ▶ 「現在の価値」を分子変数として選択します。
- ▶ 「最後の査定の値」を分母変数として選択します。
- ▶ 「郡」を部分母集団変数として選択します。
- ▶ [統計量] をクリックします。

図 18-3  
[比率分析: 統計量] ダイアログ ボックス



- ▶ [統計量] グループの [信頼区間]、[重み付けのない度数]、および [母集団のサイズ] を選択します。
- ▶ [t 検定] を選択し、検定値として「1.3」と入力します。
- ▶ [続行] をクリックします。

- ▶ [コンプレックス サンプルの比率分析] ダイアログ ボックスで、[OK] をクリックします。

## 比率

図 18-4  
比率分析表

郡	分子	分母	比率の推定	標準誤差	95% 信頼区間	
					下段	上段
東部	現在の価値	最後の査定の値	1.381	.068	1.236	1.525
中部	現在の価値	最後の査定の値	1.364	.064	1.227	1.502
西部	現在の価値	最後の査定の値	1.524	.053	1.410	1.638
北部	現在の価値	最後の査定の値	1.277	.032	1.208	1.346
南部	現在の価値	最後の査定の値	1.195	.029	1.134	1.256

表のデフォルト表示は非常に幅が広いので、ピボットして見やすくする必要があります。

### 比率分析表のピボット

- ▶ テーブルをダブルクリックしてアクティブにします。
- ▶ ビューアのメニューから、次の項目を選択します。  
ピボット > ピボットトレイ
- ▶ 行から層に [分子] をドラッグしてから、[分母] をドラッグします。
- ▶ 行から列に [郡] をドラッグします。
- ▶ 行から列に [統計量] をドラッグします。
- ▶ [ピボット トレイ] ウィンドウを閉じます。

## ピボットされた比率分析表

図 18-5  
ピボットされた比率分析表

分子: 現在の価値  
分母: 最後の査定値

	郡					
	東部	中部	西部	北部	南部	
比率の推定	1.381	1.364	1.524	1.277	1.195	
標準誤差	.068	.064	.053	.032	.029	
95% 信頼区間	下段	1.236	1.227	1.410	1.208	1.134
	上段	1.525	1.502	1.638	1.346	1.256
仮説の検定	検定値 (t サンプルの t 検定)	1.3	1.3	1.3	1.3	1.3
	t	1.191	.997	4.201	-.702	-3.646
	自由度	15	15	15	15	15
	有意確率	.252	.334	.001	.493	.002
母集団のサイズ	1983.250	1557.500	4044.000	2306.250	2204.000	
重み付けのない度数	168	179	202	205	220	

比率分析表がピボットされ、各郡の統計量を比較しやすくなりました。

- 比率の推定値は、最小が Southern 郡の 1.195 で、最大が Western 郡の 1.524 です。
- 標準誤差にも多少の幅があります。最小が Southern 郡の 0.029 で、最大が Eastern 郡の 0.068 です。
- 信頼区間の一部は重複していません。つまり、Western 郡の比率が Northern 郡および Southern 郡の比率よりも高いと結論付けることができます。
- 最後に、より客観的な測定値として、Western 郡と Southern 郡の t 検定の有意確率が 0.05 未満であることに注目してください。これにより、Western 郡の比率は 1.3 より大きく、Southern 郡の比率は 1.3 未満であると結論付けることができます。

## 集計 (報告書 データ列)

[コンプレックス サンプルの比率分析] 手続きを使用して、「現在の値」の「最後の査定値」に対する比率に関してさまざまな統計量を作成しました。その結果は、各郡の財産税の評価の間に一定程度の不公正があることを示唆しています。

- Western 郡の比率が高く、財産価値の評価については、この郡のレコードの更新状態が他の郡より古いことを示しています。この郡の財産税が低すぎる可能性があります。

- Southern 郡の比率が低く、財産価値の評価については、この郡のレコードの更新状態が他の郡よりも新しいことを示しています。この郡の財産税が高すぎる可能性があります。
- Southern 郡の比率は Western 郡の比率よりも低くなっていますが、客観値 (1.3) 以内です。

Southern 郡の財産価値の追跡に使われたリソースが Western 郡に再割り当てされると、これらの郡の比率が他の郡の比率および客観値 (1.3) と合致します。

## 関連手続き

[コンプレックス サンプルの比率分析] 手続きは、複合抽出計画で得られた観測値に関する尺度変数の比率の 1 変量の記述統計量を取得するための便利なツールです。

- **コンプレックス サンプルのサンプリング ウィザード**は、コンプレックス サンプル計画仕様の指定およびサンプル抽出に使用されます。サンプリング ウィザードにより作成される抽出計画ファイルは、デフォルト分析計画を含んでおり、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- **コンプレックス サンプルの分析準備ウィザード**は、既存のコンプレックス サンプル用の分析仕様の設定に使用されます。サンプリング ウィザードにより作成される分析計画ファイルは、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- **[コンプレックス サンプルの記述統計]** 手続きでは、尺度変数に関する記述統計量が提供されます。

# コンプレックス サンプルの一般線型モデル

コンプレックス サンプルの一般線型モデル (CSGLM) 手続きでは、コンプレックス サンプル法によって抽出したサンプルに対して、分散分析や共分散分析だけではなく、線型回帰分析を実行します。オプションとして、部分母集団の分析も実行できます。

## コンプレックス サンプルの一般線型モデルを使用した二元配置分散分析の適合

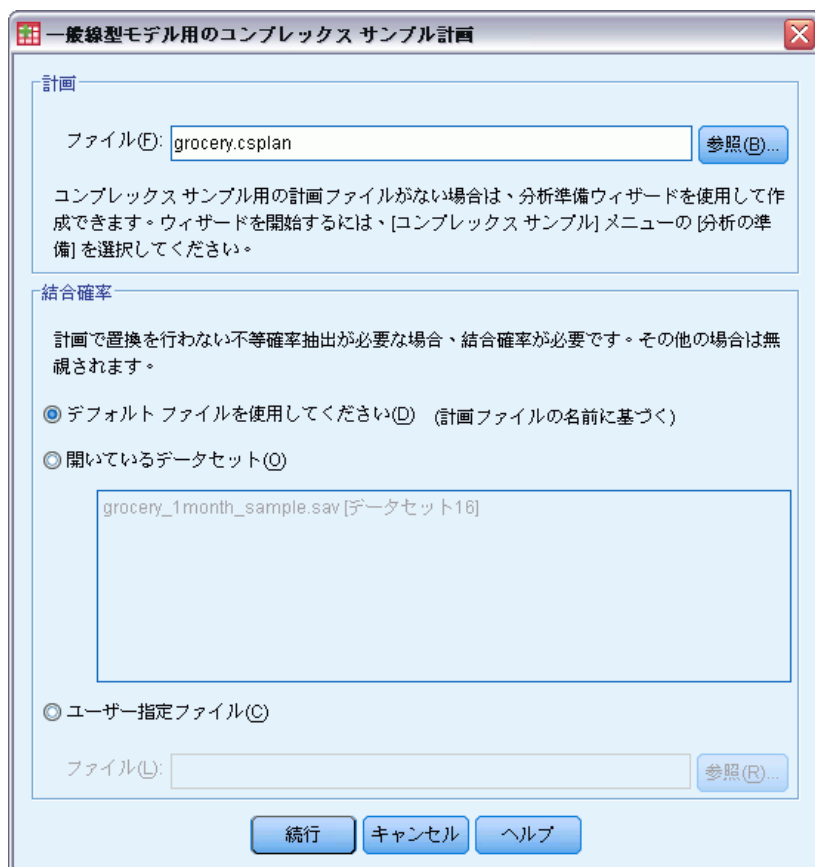
ある食料雑貨店チェーンが、特定の顧客グループを対象に、購買習慣についてコンプレックス デザインに従って調査しました。その食料雑貨店は、調査結果と各顧客の前月の購買額を前提に、顧客の性別による分類や抽出計画を取り入れながら、専門店が顧客の 1 か月間の購買額にどの程度関わっているかを調査します。

この情報は、grocery\_1month\_sample.sav に収集されています。詳細は、A 付録 p.281 [サンプル ファイル](#) を参照してください。購買額について二元配置分散分析を実行するには、コンプレックス サンプルの一般線型モデル手続きを使用します。

### 分析の実行

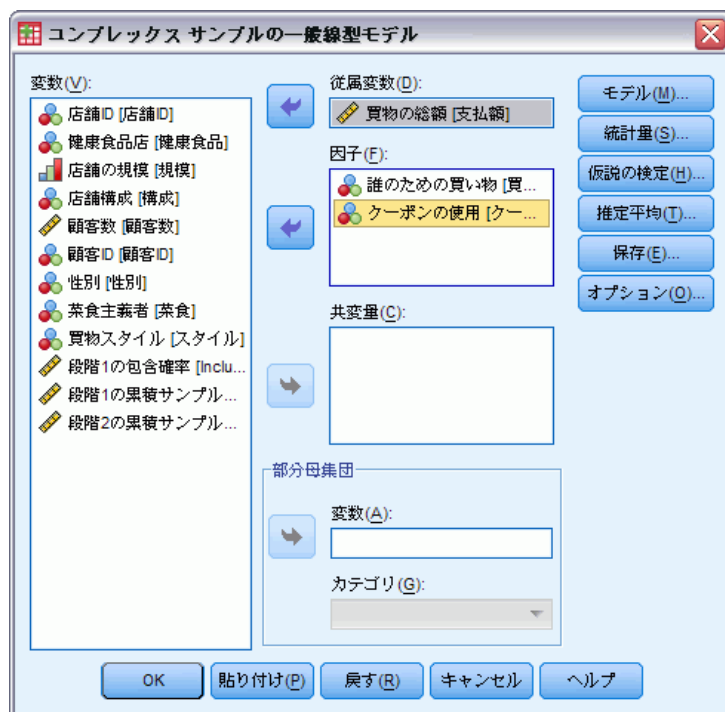
- ▶ コンプレックス サンプルの一般線型モデルの分析を実行するには、メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > 一般線型モデル

図 19-1  
[コンプレックス サンプル計画] ダイアログ ボックス



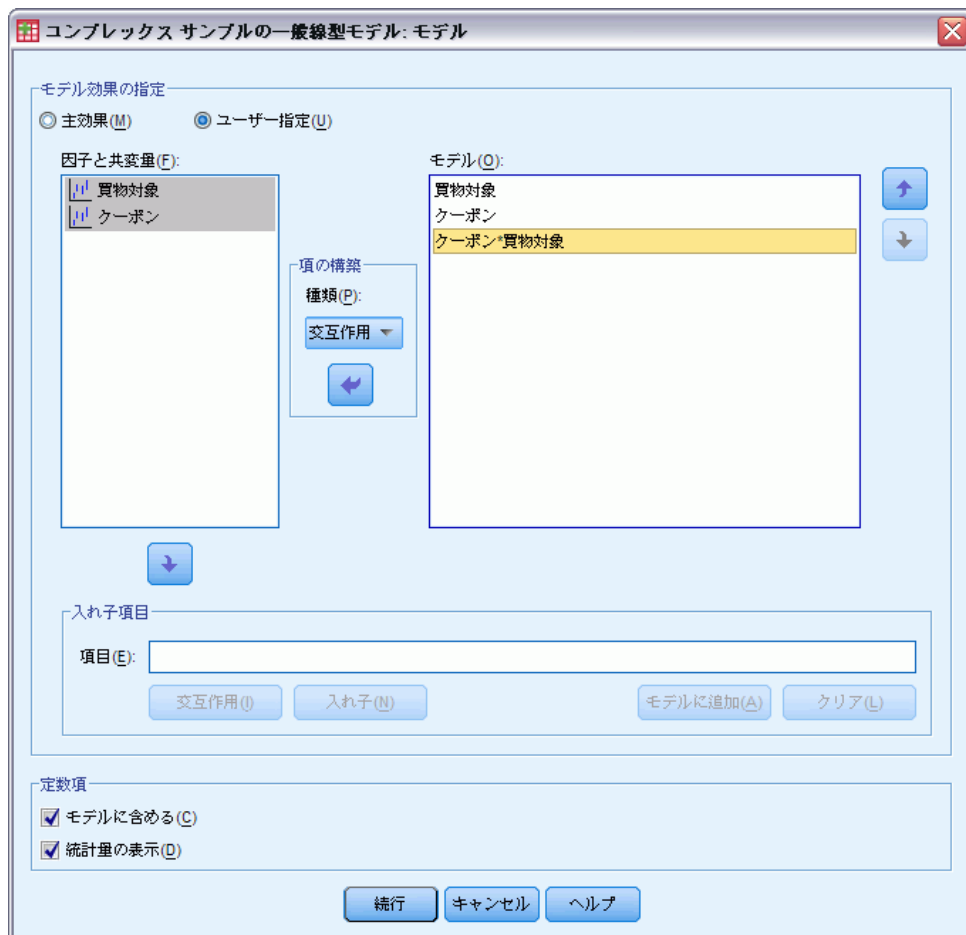
- ▶ grocery.csplan を参照して選択します。詳細は、A 付録 p.281 [サンプルファイル](#) を参照してください。
- ▶ [続行] をクリックします。

図 19-2  
[一般線型モデル] ダイアログ ボックス



- ▶ 従属変数として「買物の総額」を選択します。
- ▶ 因子として「誰のための買い物」と「クーポンの使用」を選択します。
- ▶ [モデル] をクリックします。

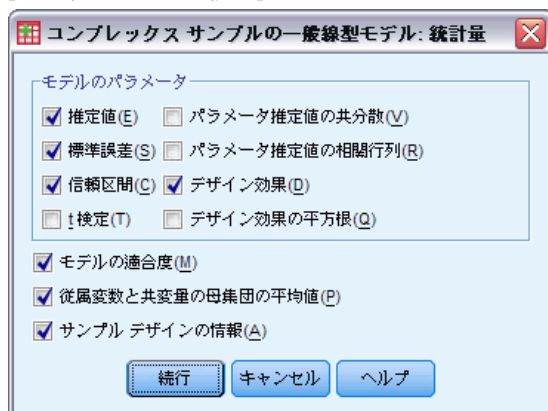
図 19-3  
[モデル] ダイアログ ボックス



- ▶ [ユーザー指定] モデルの構築を選択します。
- ▶ 構築する項の種類として[主効果]を選択し、モデル項として「買物対象」と「クーポン」を選択します。
- ▶ 構築する項の種類として[交互作用]を選択し、モデル項として「買物対象\*クーポン」交互作用を追加します。
- ▶ [続行] をクリックします。
- ▶ [一般線型モデル] ダイアログ ボックスで[統計] をクリックします。

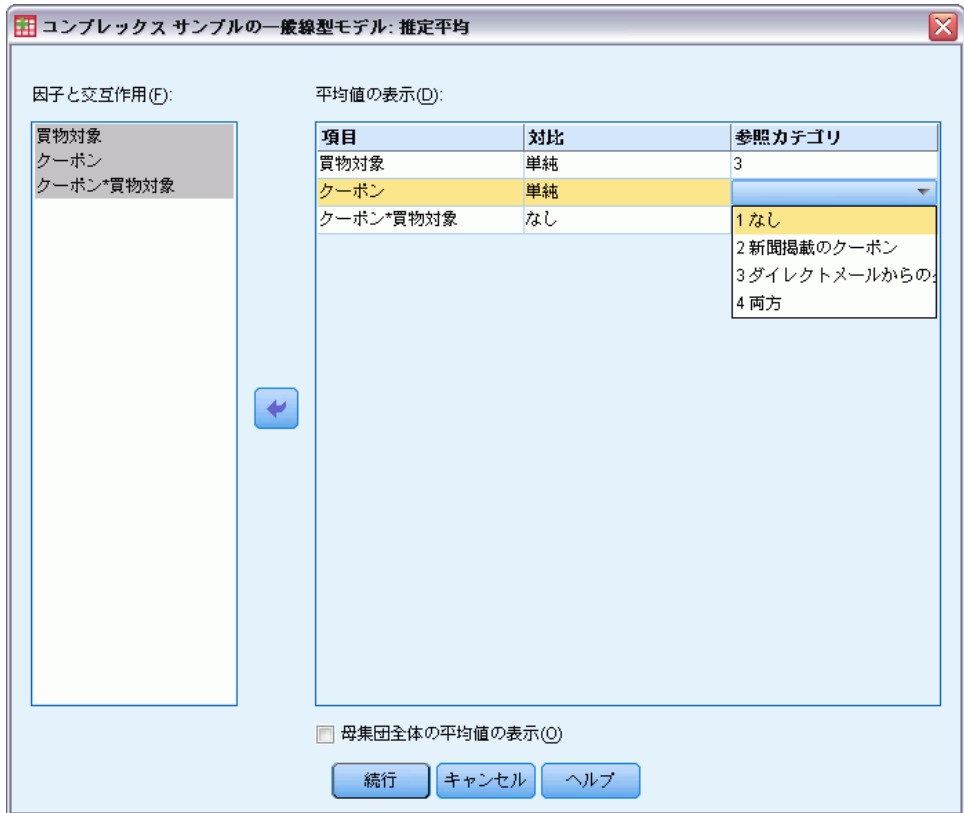


図 19-4  
[一般線型モデル: 統計] ダイアログ ボックス



- ▶ [モデルのパラメータ] グループで、[推定値]、[標準誤差]、[信頼区間]、[デザイン効果] を選択します。
- ▶ [続行] をクリックします。
- ▶ [一般線型モデル] ダイアログ ボックスで [推定平均] をクリックします。

図 19-5  
[一般線型モデル: 推定平均] ダイアログ ボックス



- ▶ 「買物対象」、「クーポン」、および「買物対象\*クーポン」交互作用の平均値の表示を選択します。
- ▶ 「買物対象」では、対比として[単純]を選択し、参照カテゴリとして[3 自分と家族]を選択します。カテゴリを選択すると、ダイアログ ボックスには「3」が表示されることに注意してください。
- ▶ 「クーポン」では、対比として[単純]を選択し、参照カテゴリとして[1 なし]を選択します。
- ▶ [続行] をクリックします。
- ▶ [一般線型モデル] ダイアログ ボックスで [OK] をクリックします。

## モデルの要約 (ピボットテーブル 回帰)

図 19-6  
R2 乗統計量

R 2 乗 (決定係数)	.601
--------------	------

a. モデル: 買物の総額 = (Intercept) + 買物対象 + クーボン + 買物対象 \* クーボン

[R2 乗] は、決定係数として、モデル適合度の強さの測定値になります。この値は、「買物の総額」の変動の約 60% がモデルで説明されることを示し、適切な説明として使用できます。モデルに他の予測変数を追加して、適合度をさらに高めることもできます。

## モデル効果の検定

図 19-7  
被験者間効果の検定

ソース	df1	df2	Wald F	有意確率
ソース (補正後のモデル)	11.000	3.000	127.231	.001
(切片)	1.000	13.000	6321.597	.000
買物対象	2.000	12.000	643.593	.000
クーポン	3.000	11.000	87.453	.000
買物対象 * クーボン	6.000	8.000	10.688	.002

a. モデル: 買物の総額 = (Intercept) + 買物対象 + クーボン + 買物対象 \* クーボン

モデル内の各項およびモデル全体に対して、その効果の値が 0 であるかどうかを検定します。有意確率が 0.05 未満の項には、明確な効果があります。したがって、すべてのモデル項がモデルに影響します。

## パラメータ推定値

図 19-8  
パラメータ推定値

パラメータ	推定値	Std Error	95% 信頼区間		デザイン効果
			下段	上段	
(Intercept)	518.249	11.731	492.905	543.592	1.387
[買物対象=1]	-174.757	10.762	-198.006	-151.508	.950
[買物対象=2]	-129.443	11.455	-154.191	-104.696	.925
[買物対象=3]	.000 <sup>a</sup>	.	.	.	.
[クーポン=1]	-140.838	10.180	-162.830	-118.846	.649
[クーポン=2]	-63.026	13.195	-91.531	-34.520	.940
[クーポン=3]	-31.375	9.726	-52.387	-10.363	.564
[クーポン=4]	.000 <sup>a</sup>	.	.	.	.
[買物対象=1] * [クーポン=1]	41.693	11.170	17.562	65.824	.606
[買物対象=1] * [クーポン=2]	44.505	18.068	5.471	83.539	1.413
[買物対象=1] * [クーポン=3]	9.204	11.057	-14.684	33.092	.594
[買物対象=1] * [クーポン=4]	.000 <sup>a</sup>	.	.	.	.
[買物対象=2] * [クーポン=1]	89.211	10.967	65.518	112.903	.533
[買物対象=2] * [クーポン=2]	54.267	14.949	21.972	86.562	.836
[買物対象=2] * [クーポン=3]	17.884	13.753	-11.828	47.595	.797
[買物対象=2] * [クーポン=4]	.000 <sup>a</sup>	.	.	.	.
[買物対象=3] * [クーポン=1]	.000 <sup>a</sup>	.	.	.	.
[買物対象=3] * [クーポン=2]	.000 <sup>a</sup>	.	.	.	.
[買物対象=3] * [クーポン=3]	.000 <sup>a</sup>	.	.	.	.
[買物対象=3] * [クーポン=4]	.000 <sup>a</sup>	.	.	.	.

a. このパラメータは余分であるため、0に設定されます。

b. モデル: 買物の総額 = (Intercept) + 買物対象 + クーポン + 買物対象 \* クーポン

パラメータ推定値は、「買物の総額」に対する各予測変数の効果を示します。定数項の値が 518.249 の場合は、扶養家族のいる顧客が新聞やダイレクトメールに付いている優待券を使用して、平均で 518.25 ドル購買することを食料雑貨店チェーンが期待できることを示しています。因子レベルのパラメータが冗長であるため、定数項はこれらの因子レベルと関連付けられていると言えます。

- 「買物対象」係数は、送付されたクーポンと新聞から切り抜いたクーポンを両方使用する顧客の中で、単身者は配偶者のいる人より購買額が少ない傾向があり、同様に、配偶者のいる人は扶養家族のいる人よりも購買額が少ない傾向にあることを示しています。モデル効果の検定では、この項がモデルに影響することが示されたため、これらの差は偶然の結果ではありません。
- 「クーポン」係数は、扶養家族のいる顧客の間では、クーポンの使用が減少すると、購買額も減少することを示しています。推定値にはある程度不確定な部分がありますが、信頼区間には 0 は含まれません。

- 交互作用係数は、クーポンを使用しなかったり、新聞から切り抜くだけの顧客で、扶養家族のいない顧客の購買額が予想以上に大きい傾向にあることを示します。交互作用パラメータのどこか一部分でも冗長である場合、その交互作用パラメータは冗長です。
- デザイン効果の値の 1 からの偏差は、これらのパラメータ推定値について計算された標準誤差には、測定値が単純な無作為抽出によるものであると想定した場合に得られる標準誤差よりも大きくなるものもあれば、小さくなるものもあることを示します。抽出計画情報を分析に取り入れることは特に重要です。たとえば、「クーポン=3」係数が 0 になる可能性があるからです。

パラメータ推定値は、各モデル項の効果を数量化する場合に役立ちますが、推定周辺平均テーブルを使用すると、モデルの結果をより簡単に解釈できるようになります。

## 推定周辺平均 (GLM)

図 19-9  
[誰のための買物] のレベルによる推定周辺平均

誰のための買物	平均値 (ラウンド検定)	Std Error	95%信頼区間	
			下段	上段
自分	308.5326	3.94286	300.0145	317.0506
自分と配偶者	370.3361	4.87908	359.7955	380.8767
自分と家族	459.4392	7.19769	443.8895	474.9888

このテーブルでは、「誰のための買物」の因子レベルで、「買物の総額」のモデル推定の周辺平均と標準誤差が表示されます。このテーブルは、この因子のレベル間の差を探索する場合に役立ちます。この例では、単身者の購買額は約 308.53 ドルと予想されます。さらに、配偶者のいる顧客の購買額は 370.34 ドル、扶養家族のいる顧客の購買額は 459.44 ドルと予想されます。この差が実際の差を表しているのか、または偶発的な変動で起きたものかどうかを確認するには、検定結果を確認します。

図 19-10  
性別の推定周辺平均に関する個別の検定結果

個々の検定結果								
誰のための買物 単純対比 <sup>a</sup>	対比推定値	仮説値	差分 (推定値 - 仮説値)	標準誤差	df1	df2	Wald F	有意確率
レベル 自分とレベル 自分と家族	-150.907	.000	-150.907	4.903	1.000	13.000	947.409	.000
レベル 自分と配偶者 とレベル 自分と家族	-89.103	.000	-89.103	5.903	1.000	13.000	227.842	.000

a. 参照カテゴリ = 自分と家族

個別の検定テーブルには、購買額に関する 2 つの単純対比が表示されます。

- 対比推定値は、「誰のための買物」に表示されているレベルの購買額の差です。

- 仮説値が 0.00 の場合は、購買額に差がないという確信を表しています。
- Wald F 統計量は、表示されている自由度を持っており、対比推定値と仮説値の差が偶発的な変動によるものかどうかを検定する場合に使用されます。
- 有意確率が 0.05 未満なので、購買額には差があると結論付けることができます。

対比推定値は、パラメータ推定値と異なります。これは、「誰のための買物」効果を含む交互作用項があるためです。したがって、「買物対象=1」のパラメータ推定値は、変数「クーポンの使用」の[両方]レベルでの[自分]レベルと[自分と家族]レベルの単純対比になります。このテーブルの対比推定値の平均は、「クーポンの使用」のレベルより上になります。

図 19-11  
性別の推定周辺平均に関する全体的な検定結果

df1	df2	Wald F	有意確率
2.000	12.000	643.593	.000

全体的な検定テーブルでは、個別の検定テーブルのすべての対比の検定結果が表示されます。その有意確率が 0.05 未満の場合は、「誰のための買物」のレベル間で購買額に差があることを立証しています。

図 19-12  
購買スタイルのレベルによる推定周辺平均

クーポンの使用	平均値 (円検定)	Std Error	95% 信頼区間	
			下段	上段
なし	319.6455	6.51429	305.5722	333.7188
新聞掲載のクーポン	386.7469	4.32295	377.4077	396.0861
ダイレクトメールからのクーポン	394.5028	5.54218	382.5297	406.4760
両方	416.8486	6.51260	402.7790	430.9182

このテーブルでは、「クーポンの使用」の因子レベルで、「買物の総額」のモデル推定の周辺平均と標準誤差が表示されます。このテーブルは、この因子のレベル間の差を探索する場合に役立ちます。この例では、クーポンを使用しない顧客の購買額は約 319.65 ドル、さらに、クーポンを使用する顧客の購買額はクーポンを使用しない顧客の購買額よりも大幅に大きいことが予想されます。

図 19-13  
 購買スタイルの推定周辺平均に関する個別の検定結果

**個々の検定結果**

クーポンの使用 単純対比 <sup>a</sup> レベル 新聞掲載のクーポン とレベル なし	対比推定値	仮説値	差分 (推定値 - 仮説値)	標準誤差	df1	df2	Wald F	有意確率
レベル ダイレクトメールか らのクーポンとレベル なし	67.101	.000	67.101	6.537	1.000	13.000	105.352	.000
レベル 両方とレベル なし	74.857	.000	74.857	5.875	1.000	13.000	162.328	.000
	97.203	.000	97.203	5.603	1.000	13.000	300.921	.000

a. 参照カテゴリ = なし

個別の検定テーブルには、クーポンを使用しない顧客と使用する顧客の購買額を比較して、3つの単純対比が表示されます。

検定の有意確率が 0.05 未満なので、クーポンを使用する顧客はクーポンを使用しない顧客よりも購買額が大きい傾向にあると結論付けることができます。

図 19-14  
 購買スタイルの推定周辺平均に関する全体的な検定結果

**全体的な検定結果**

df1	df2	Wald F	有意確率
3.000	11.000	87.453	.000

全体的な検定テーブルでは、個別の検定テーブルのすべての対比の検定結果が表示されます。その有意確率が 0.05 未満の場合は、「クーポンの使用」のレベル間で購買額に差があることを立証しています。「クーポンの使用」と「誰のための買物」の全体的な検定は、仮説対比値が 0 に等しくなるため、モデル効果の検定に相当することに注意してください。

図 19-15  
 購買スタイル別の性別のレベルによる推定周辺平均

誰のための買い物	クーポンの使用	平均値 (ラン検定)	Std Error	95%信頼区間	
				下段	上段
自分	なし	244.3471	6.00949	231.3644	257.3298
	新聞掲載のクーポン	324.9708	5.94134	312.1353	337.8063
	ダイレクトメールからのクーポン	321.3207	4.11028	312.4410	330.2005
	両方	343.4916	6.57845	329.2797	357.7034
自分と配偶者	なし	337.1783	7.12181	321.7925	352.5640
	新聞掲載のクーポン	380.0468	7.91038	362.9574	397.1361
	ダイレクトメールからのクーポン	375.3141	6.22468	361.8665	388.7617
	両方	388.8054	7.12101	373.4214	404.1894
自分と家族	なし	377.4111	11.58215	352.3894	402.4328
	新聞掲載のクーポン	455.2232	6.14420	441.9494	468.4969
	ダイレクトメールからのクーポン	486.8736	10.76529	463.6166	510.1306
	両方	518.2488	11.73120	492.9050	543.5925

このテーブルでは、「誰のための買物」と「クーポンの使用」という因子の組み合わせで、「買物の総額」のモデル推定の周辺平均、標準誤差、および信頼区間が表示されます。このテーブルは、モデル効果の検定で検出された、これら2つの因子間の交互作用効果を探索する場合に役立ちます。

## 集計 (報告書 データ列)

この例では、推定周辺平均により、「誰のための買物」と「クーポンの使用」のさまざまなレベルで、顧客間の購買額の差を明らかにしました。モデル効果の検定により、このことだけではなく、「誰のための買物\*クーポンの使用」の交互作用効果があると見られることも確認されました。モデルの要約表では、現在のモデルでデータの変動の半数以上が説明され、予測変数をさらに追加することでモデルが改善される可能性があることが示されました。

## 関連手続き

コンプレックス サンプルの一般線型モデル手続きは、コンプレックス サンプルリング法に従ってケースが抽出された場合にスケール変数をモデリングするための便利なツールです。

- **コンプレックス サンプルのサンプリング ウィザード**は、コンプレックス サンプル計画仕様の指定およびサンプル抽出に使用されます。サンプリング ウィザードにより作成される抽出計画ファイルは、デフォルト分析計画を含んでおり、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。



- **コンプレックス サンプルの分析準備ウィザード**は、既存のコンプレックス サンプル用の分析仕様の指定に使用されます。サンプリング ウィザードにより作成される分析計画ファイルは、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- **コンプレックス サンプルのロジスティック回帰**手続きでは、カテゴリ応答をモデル化できます。
- **コンプレックス サンプルの順序回帰**手続きでは、順序応答をモデル化できます。

# コンプレックス サンプルのロジスティック回帰

[コンプレックス サンプルのロジスティック回帰] 手続きでは、コンプレックス サンプル法によって抽出されたサンプルに対して、2 値または多項の従属変数のロジスティック回帰分析を実行します。オプションとして、部分母集団の分析も実行できます。

## コンプレックス サンプルのロジスティック回帰を使用した信用リスクへのアクセス

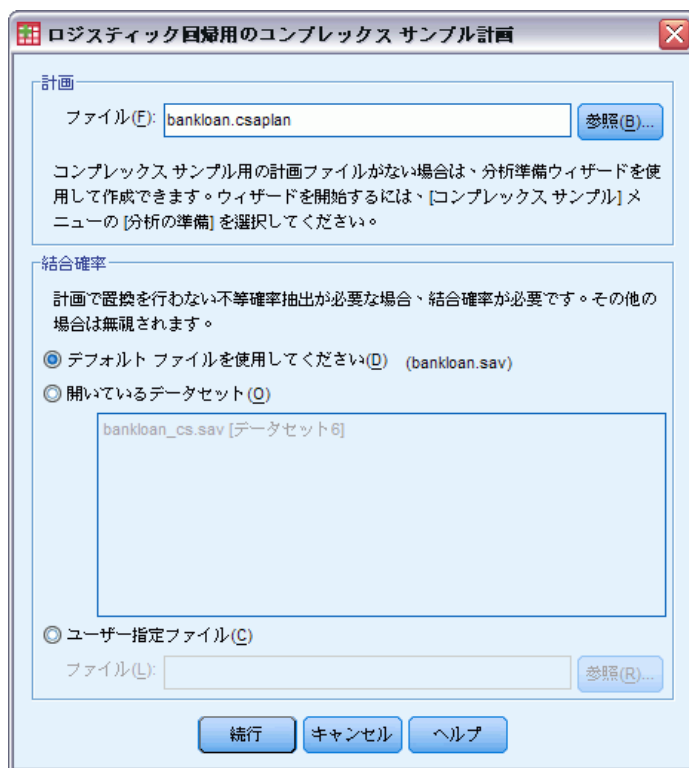
銀行の融資担当者の場合、債務不履行になる可能性がある人物を示す特徴を特定し、その特徴を使用して信用リスクの良し悪しを識別できます。

ある融資担当者が、コンプレックス デザインに従って、複数の支店で融資を受けた顧客の過去の記録を収集したとします。この情報は bankloan\_cs.sav に含まれています。詳細は、A 付録 p. 281 [サンプル ファイル](#) を参照してください。担当者は、顧客の債務不履行が年齢、職歴、およびクレジットカードの負債額に関連している可能性があるかどうかを調べ、抽出計画を取り入れようとしています。

### 分析の実行

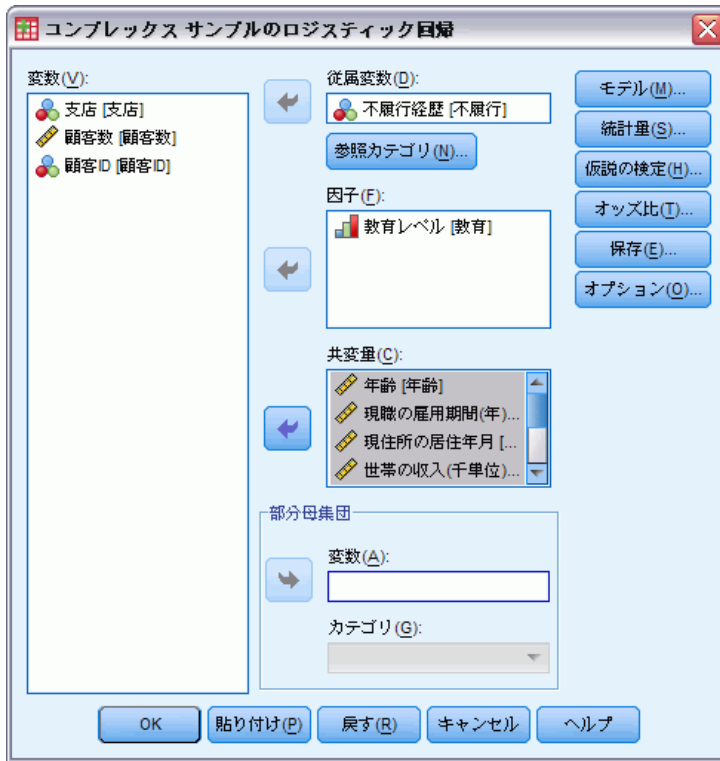
- ▶ ロジスティック回帰モデルを作成するには、メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > ロジスティック回帰...

図 20-1  
[コンプレックス サンプル計画] ダイアログ ボックス



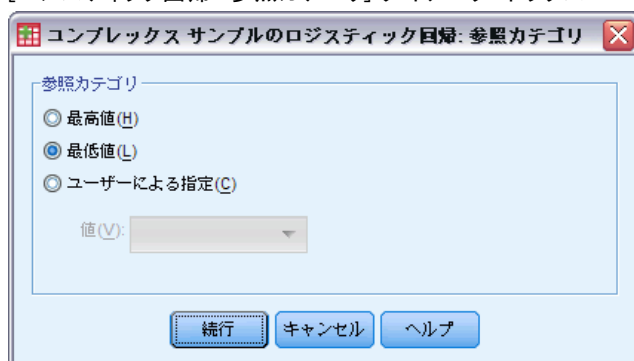
- ▶ bankloan.csaplan を参照して選択します。詳細は、A 付録 p. 281 [サンプル ファイル](#) を参照してください。
- ▶ [続行] をクリックします。

図 20-2  
[ロジスティック回帰] ダイアログ ボックス



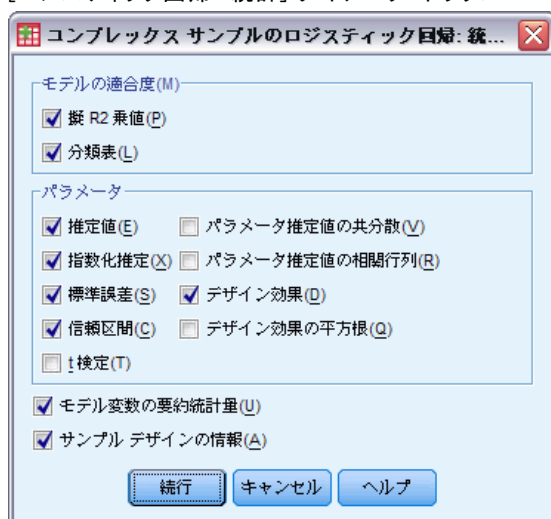
- ▶ [従属変数] として「不履行履歴」を選択します。
- ▶ 因子として「教育レベル」を選択します。
- ▶ [共変量] として「年齢」から「その他の負債 (千単位)」までを選択します。
- ▶ 「不履行経歴」を選択して、[参照カテゴリ] をクリックします。

図 20-3  
[ロジスティック回帰: 参照カテゴリ] ダイアログ ボックス



- ▶ 参照カテゴリとして [最低値] を選択します。  
これにより、「不履行なし」カテゴリが参照カテゴリとして設定されます。そのため、出力で表示されたオッズ比には、債務不履行の可能性が高まるにつれて、オッズ比が大きくなるという性質があります。
- ▶ [続行] をクリックします。
- ▶ [ロジスティック回帰] ダイアログ ボックスで、[統計] をクリックします。

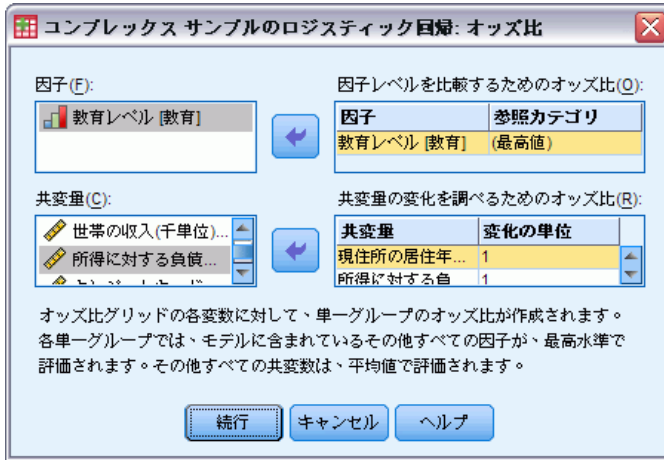
図 20-4  
[ロジスティック回帰: 統計] ダイアログ ボックス



- ▶ [モデルの適合度] グループで、[分類テーブル] を選択します。
- ▶ [パラメータ] グループで、[推定値]、[指数化推定]、[標準誤差]、[信頼区間]、[デザイン効果] を選択します。
- ▶ [続行] をクリックします。

- ▶ [ロジスティック回帰] ダイアログ ボックスで、[オッズ比] をクリックします。

図 20-5  
[ロジスティック回帰: オッズ比] ダイアログ ボックス



- ▶ 因子「教育」と、共変量「雇用」および「負債比」のオッズ比の作成を選択します。
- ▶ [続行] をクリックします。
- ▶ [ロジスティック回帰] ダイアログ ボックスで、[OK] をクリックします。

## 擬 R<sup>2</sup> 乗

図 20-6  
擬 R<sup>2</sup> 乗統計量

Cox と Snell	.330
Nagelkerke	.451
McFadden	.304

従属変数: 不品行経歴 (参照カテゴリ = なし)  
モデル: (Intercept), 教育, 年齢, 雇用, 居住年  
数, 収入, 負債比, クレジット負債, その他負債

線型回帰モデルでの決定係数  $R^2$  乗は、予測 (独立) 変数に対する従属変数の分散の比率を要約します。 $R^2$  乗の値が大きくなるほど、変動全体のうち、モデルによって説明される部分が多くなります (最大で 1)。カテゴリ従属変数がある回帰モデルの場合、線型回帰モデルにおける  $R^2$  乗の特性をすべて備えた単一の  $R^2$  乗統計量は計算できません。このため、近似値が計算されます。決定係数の推定には次の方法が使用されます。

- Cox と Snell の  $R^2$  乗 (Cox および Snell, 1989)。ベースライン モデルの対数尤度と比較されるモデルの対数尤度に基づいています。ただし、カテゴリ型の結果では、モデルが「完璧」な場合でも理論上の最大値は 1 未満です。
- Nagelkerke の  $R^2$  乗 (Nagelkerke, 1991)。Cox と Snell の  $R^2$  乗の修正版で、0 から 1 までの範囲に完全に対応するよう統計量の尺度を調整します。
- McFadden の  $R^2$  乗 (McFadden, 1974)。別の方法による修正版で、定数項のみのモデルおよび完全推定モデルの対数尤度カーネルに基づきます。

$R^2$  乗値が「適切」かどうかの基準は、適用する領域によって異なります。この統計量は、それ自体で示唆に富みますが、同じデータの競合するモデルを比較する場合に最も有効です。こうした観点に基づいた場合、 $R^2$  乗統計量が最も大きいモデルが「最適」であると言えます。

## 分類

図 20-7  
分類テーブル

観測	予測値		
	なし	あり	正確な パーセント
なし	188289.667	31871.267	85.5%
あり	49970.600	77675.133	60.9%
全体のパーセント	68.5%	31.5%	76.5%

従属変数: 不履行経歴 (参照カテゴリ = なし)  
モデル: (Intercept), 教育, 年齢, 雇用, 居住年数, 収入, 負債  
比: 外債/負債, その他負債

分類テーブルでは、実際にロジスティック回帰モデルを使用した結果が表示されます。各ケースでは、そのケースのモデル予測ロジットが 0 より大きい場合、予測応答は [はい] になります。ケースは「finalweight」によって重み付けされているため、分類テーブルでは、母集団で予測されるモデルのパフォーマンスが示されます。

- 対角上のセルが正しい予測値です。
- 対角線から外れたセルは不正な予測値です。

モデルの作成に使用されるケースを基に、このモデルを使用して、債務不履行者ではない人の 85.5% が母集団内で正しく分類されることを期待できます。同様に、債務不履行者の 60.9% も正しく分類されることを期待できます。全体的には、ケースの 76.5% が正しく分類されることを期待できます。ただし、このテーブルは、モデルの作成に使用したケースで作成されたため、ケースの推定値は過度に楽観的である可能性があります。

## モデル効果の検定

図 20-8  
被験者間効果の検定

ソース	df1	df2	Wald F	有意確率
(補正後のモデル) (切片)	11.000	4.000	14.669	.010
教育	1.000	14.000	5.777	.031
年齢	4.000	11.000	1.683	.224
雇用	1.000	14.000	5.352	.036
居住年数	1.000	14.000	88.244	.000
収入	1.000	14.000	1.123	.307
負債比	1.000	14.000	.007	.932
クレジット負債	1.000	14.000	27.632	.000
クレジット負債	1.000	14.000	33.402	.000
その他負債	1.000	14.000	.709	.414

従属変数: 不履行経歴 (参照カテゴリ = なし)

モデル: (Intercept), 教育, 年齢, 雇用, 居住年数, 収入, 負債比, クレジット負債, その他負債

モデル内の各項およびモデル全体に対して、その効果が 0 であるかどうかを検定します。有意確率が 0.05 未満の項には、明確な効果があります。したがって、「年齢」、「雇用」、「負債比」、「クレジット負債」はモデルに影響しますが、他の主効果は影響しません。さらにデータを分析すると、モデルの考慮事項から「教育」、「居住年数」、「収入」、「その他負債」が削除される可能性があります。

## パラメータ推定値

図 20-9  
パラメータ推定値

不履行経歴	パラメータ	回帰係数	Std Error	95% 信頼区間		デザイン効果	Exp (B)	Exp (B) の 95% 信頼区間	
				下段	上段			下段	上段
あり	(Intercept)	-1.140	.399	-1.995	-.284	.665	.320	.136	.753
	[教育=1]	.720	.340	-.010	1.449	.862	2.054	.990	4.259
	[教育=2]	.684	.371	-.112	1.481	1.247	1.983	.894	4.397
	[教育=3]	.518	.307	-.140	1.177	.813	1.679	.869	3.244
	[教育=4]	.789	.302	.142	1.437	.817	2.202	1.152	4.208
	[教育=5]	.000 <sup>a</sup>	.	.	.	.	1.000	.	.
	年齢	-.023	.010	-.043	-.002	.418	.978	.958	.998
	雇用	-.225	.024	-.277	-.174	1.200	.798	.758	.840
	居住年数	-.028	.026	-.085	.029	.651	.972	.919	1.029
	収入	.000	.003	-.007	.006	1.410	1.000	.993	1.006
	負債比	.095	.018	.056	.134	1.222	1.100	1.058	1.143
	クレジット負債	.493	.085	.310	.676	1.373	1.637	1.363	1.966
	その他負債	.026	.031	-.041	.094	1.219	1.027	.960	1.098

従属変数: 不履行経歴 (参照カテゴリ = なし)

モデル: (Intercept), 教育, 年齢, 雇用, 居住年数, 収入, 負債比, クレジット負債, その他負債

a. このパラメータは余分であるため、0 に設定されます。

パラメータ推定値テーブルは、各予測変数の効果をまとめたものです。パラメータ値は、「不履行なし」カテゴリに対する「不履行あり」カテゴリの尤度に影響することに注意してください。したがって、パラメータが正の係数を持つ場合は不履行の尤度が大きくなり、パラメータが負の係数を持つ場合は不履行の尤度が小さくなります。



ロジスティック回帰係数は、線型回帰係数ほど単純ではありません。[B]はモデルの効果を検定する場合に便利ですが、[Exp(B)]の方が容易に解釈できます。[Exp(B)]は、交互作用項に含まれない予測変数に対する予測変数の単位増加量が1であるために生じる、注目のイベントのオッズ比の変化を示しています。たとえば、「雇用」の[Exp(B)]が0.798だとします。これは、雇用期間が2年の場合の債務不履行のオッズが、雇用期間が1年の場合の債務不履行のオッズの0.798倍であることを意味します。この場合、雇用期間以外の条件はすべて同じです。

デザイン効果は、これらのパラメータ推定値について計算された標準誤差には、測定値が単純な無作為抽出によるものであると想定した場合に得られる標準誤差よりも大きくなるものもあれば、小さくなるものもあることを示します。抽出計画情報を分析に取り入れることは特に重要です。たとえば、年齢係数が0になる可能性があるからです。

## オッズ比

図 20-10  
教育レベル [教育] のオッズ比

	不履行経歴	オッズ比	95% 信頼区間	
			下段	上段
教育 レベル	中学 vs. 大学院 あり	2.054	.990	4.259
	高校 vs. 大学院 あり	1.983	.894	4.397
	短大 vs. 大学院 あり	1.679	.869	3.244
	大学 vs. 大学院 あり	2.202	1.152	4.208

従属変数: 不履行経歴 (参照カテゴリ = なし)

モデル: (Intercept), 教育, 年齢, 雇用, 居住年数, 収入, 負債比, クレジット負債, その他負債

- a. 計算に使用される因子と共変量は値 教育レベル=大学院; 年齢=34.19; 現職の雇用期間(年)=6.99; 現住所の居住年月=6.32; 世帯の収入(千単位)=60.1581; 所得に対する負債の比率(×100)=9.9341; クレジットカードの負債(千単位)=1.9764; その他の負債(千単位)=3.9164に固定されます。

このテーブルでは、「教育のレベル [教育]」の因子レベルで「不履行経歴」のオッズ比が表示されます。表示された値は、「大学院」の債務不履行のオッズと比較した、「中学」から「短大」までの債務不履行のオッズの比率を示します。したがって、テーブルの最初の行に表示されるオッズ比 2.054 は、高卒未満者の債務不履行のオッズが、大卒者の債務不履行のオッズの 2.054 倍であることを意味します。

図 20-11  
現職の雇用期間 (年) のオッズ比

変化の単位	不履行経歴	オッズ比	95% 信頼区間	
			下段	上段
現職の雇用期間(年) 1.000	あり	.798	.758	.840

従属変数: 不履行経歴 (参照カテゴリ = なし)

モデル: (Intercept), 教育, 年齢, 雇用, 居住年数, 収入, 負債比, クレジット負債, その他負債

- a. 計算に使用される因子と共変量は値 教育レベル=大学院; 年齢=34.19; 現職の雇用期間(年)=6.99; 現住所の居住年月=6.32; 世帯の収入(千単位)=60.1581; 所得に対する負債の比率(×100)=9.9341; クレジットカードの負債(千単位)=1.9764; その他の負債(千単位)=3.9164に固定されます。

このテーブルでは、共変量「現職の雇用期間(年)」の単位変更に対する「不履行経歴」のオッズ比が表示されます。表示された値は、雇用期間 6.99 年(平均)の人の債務不履行のオッズと比較した、雇用期間が 7.99 年の人の債務不履行のオッズ比を示しています。

図 20-12  
所得に対する負債の比率のオッズ比

変化の単位	不履行経歴	オッズ比	95%信頼区間	
			下段	上段
所得に対する負債の比率(x100)	1.000 あり	1.100	1.058	1.143

従属変数: 不履行経歴(参照カテゴリ=なし)

モデル: Intercept, 教育, 年齢, 雇用, 居住年数, 収入, 負債比, クレジット負債, その他負債

- a. 計算に使用される因子と共変量は値 教育レベル=大学院; 年齢=34.19; 現職の雇用期間(年)=6.99; 現住所の居住年月=6.32; 世帯の収入(千単位)=60.1581; 所得に対する負債の比率(x100)=9.9341; クレジットカードの負債(千単位)=1.9764; その他の負債(千単位)=3.9164に固定されます。

このテーブルでは、共変量「所得に対する負債の比率(x100)」の単位変更に対する「不履行経歴」のオッズ比が表示されます。表示された値は、負債率または所得率が 9.9341(平均)の人の債務不履行のオッズと比較した、その率が 10.9341の人の債務不履行のオッズ比を示しています。

これらの予測変数が交互作用項に含まれていないため、テーブルに表示されたオッズ比の値は、指数化されたパラメータ推定値と等しくなることに注意してください。予測変数が交互作用項に含まれている場合は、テーブルに表示されるそのオッズ比も、交互作用を構成する他の予測変数の値に左右されます。

## 集計(報告書 データ列)

コンプレックス サンプルのロジスティック回帰手続きを使用して、特定の顧客が債務不履行になる確率を予測するためのモデルを作成しました。

融資担当者にとっての重要な問題は、タイプ I およびタイプ II の誤りにかかるコストです。つまり、債務の不履行者を履行者として分類する際のコスト(タイプ I)はどのようなもので、債務の履行者を不履行者として分類する際のコスト(タイプ II)はどのようなものかが重要になります。不良債権に最も関心がある場合は、タイプ I の誤りを小さくし、**感度**を最大にする必要があります。顧客基盤の拡大を優先する場合は、タイプ II の誤りを小さくし、**特異性**を最大にする必要があります。通常は、両方とも主要な問題なので、感度と特異性が適切に組み合わせられた、顧客を分類するための決定規則を選択する必要があります。

## 関連手続き

コンプレックス サンプルのロジスティック回帰手続きは、コンプレックス サンプル法に従ってケースが抽出された場合にカテゴリ変数をモデリングするための便利なツールです。

- **コンプレックス サンプルのサンプリング ウィザード**は、コンプレックス サンプル計画仕様の指定およびサンプル抽出に使用されます。サンプリング ウィザードにより作成される抽出計画ファイルは、デフォルト分析計画を含んでおり、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- **コンプレックス サンプルの分析準備ウィザード**は、既存のコンプレックス サンプル用の分析仕様の指定に使用されます。サンプリング ウィザードにより作成される分析計画ファイルは、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- **コンプレックス サンプルの一般線型モデル**手続きでは、尺度の応答をモデル化できます。
- **コンプレックス サンプルの順序回帰**手続きでは、順序応答をモデル化できます。

# コンプレックス サンプルの順序 回帰

[コンプレックス サンプルの順序回帰] 手続きでは、コンプレックス サンプルリング法によって抽出されたサンプルに対して、順序従属変数の予測モデルを作成します。オプションとして、部分母集団の分析も実行できます。

## コンプレックス サンプルの順序回帰を使用した調査結果の分析

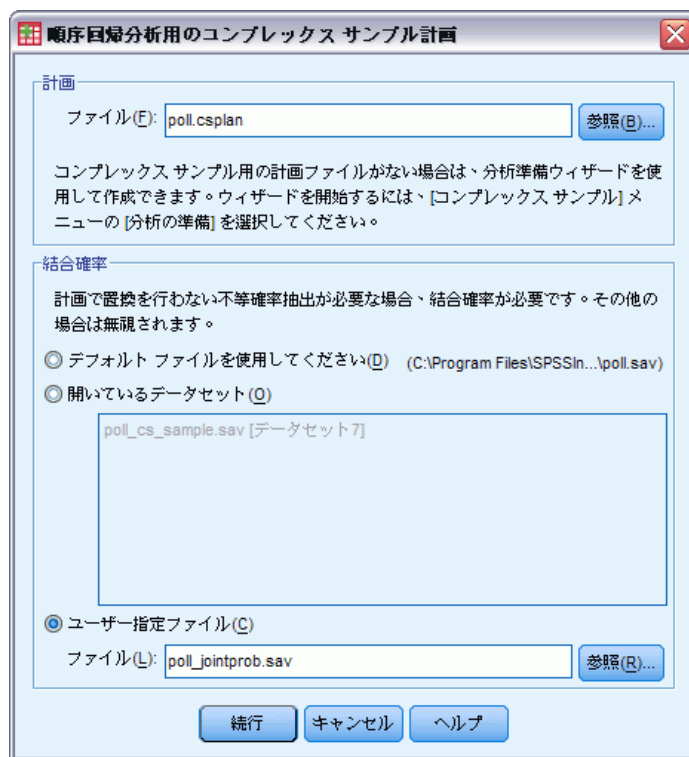
議会開会前に法案の提出を検討している議員は、市民が法案を支持しているかどうか、法案への支持がどの程度有権者の人口統計に関連しているかに関心を持っています。そこで、世論調査会社が複合抽出計画に基づいて対面式の調査を実施することになりました。

調査結果は、poll\_cs\_sample.sav に収集されています。世論調査会社で使用した抽出計画は、poll.csplan に保存されています。調査では確率比例 (PPS) 法を採用したため、結合選択確率を保持したファイル (poll\_jointprob.sav) も用意されています。詳細は、A 付録 p.281 [サンプル ファイル](#) を参照してください。コンプレックス サンプルの順序回帰を使用すると、有権者の人口統計に基づいて法案の支持水準のモデルを当てはめることができます。

## 分析の実行

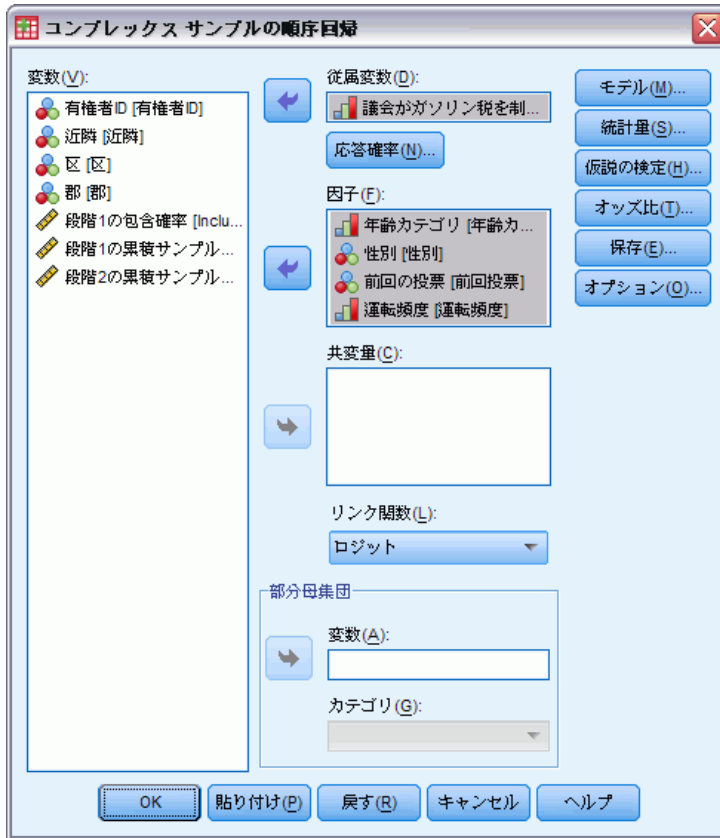
- ▶ コンプレックス サンプルの順序回帰分析を実行するには、メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > 順序回帰...

図 21-1  
[コンプレックス サンプル計画] ダイアログ ボックス



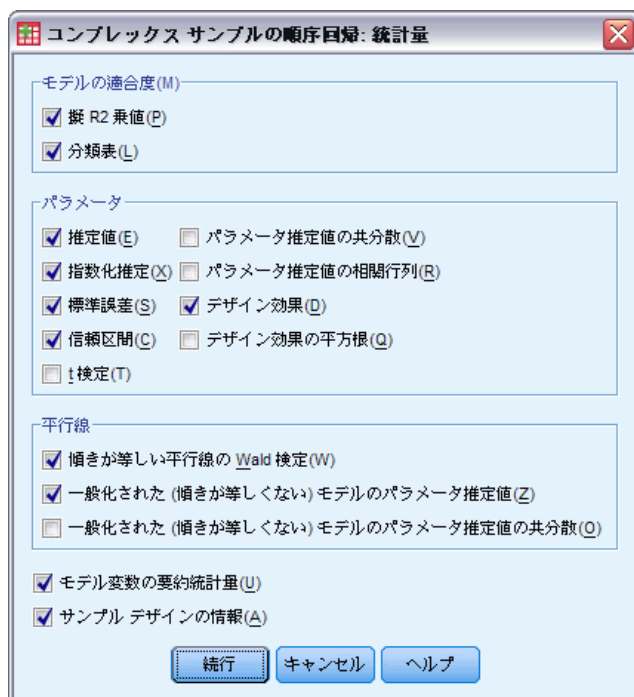
- ▶ 計画ファイルとして poll.csplan を参照して選択します。詳細は、A 付録 p.281 [サンプル ファイル](#) を参照してください。
- ▶ 結合確率ファイルとして poll\_jointprob.sav を選択します。
- ▶ [続行] をクリックします。

図 21-2  
[順序回帰] ダイアログ ボックス



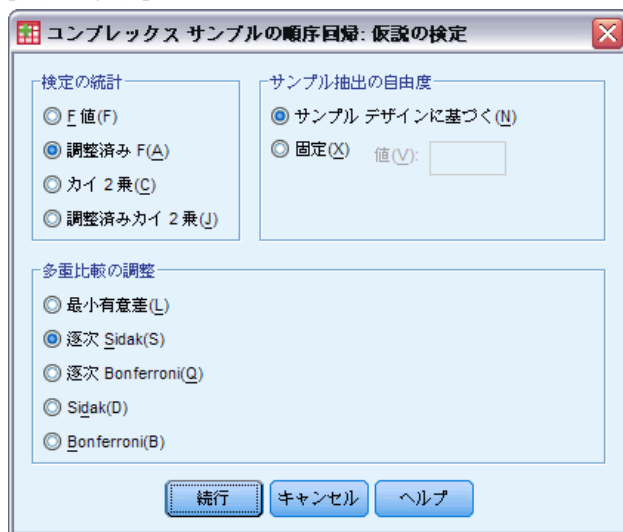
- ▶ 従属変数として、「議会がガソリン税を制定するべきか」を選択します。
- ▶ 因子として、「年齢カテゴリ」から「運転頻度」までを選択します。
- ▶ [統計量] をクリックします。

図 21-3  
[順序回帰: 統計量] ダイアログ ボックス



- ▶ [モデルの適合度] グループで、[分類テーブル] を選択します。
- ▶ [パラメータ] グループで、[推定値]、[指数化推定]、[標準誤差]、[信頼区間]、[デザイン効果] を選択します。
- ▶ [傾きが等しい平行線の Wald 検定] および [一般化された (傾きが等しくない) モデルのパラメータ推定値] を選択します。
- ▶ [続行] をクリックします。
- ▶ [コンプレックス サンプルの順序回帰] ダイアログ ボックスで、[仮説の検定] をクリックします。

図 21-4  
[仮説の検定] ダイアログ ボックス

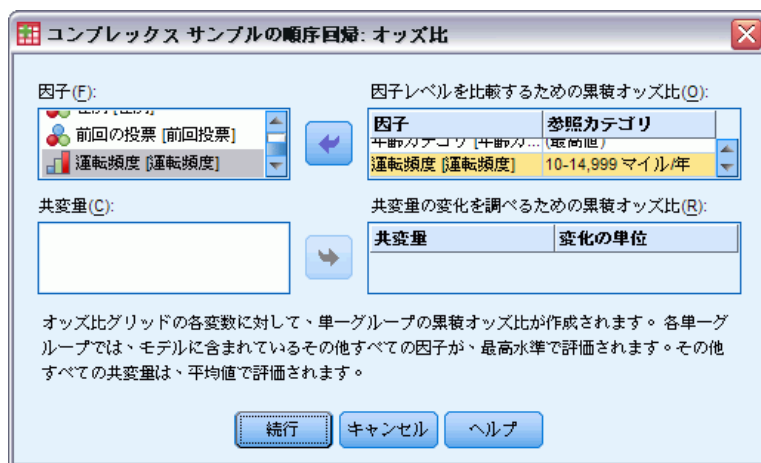


予測変数と応答カテゴリが適度な数であっても、Wald F 検定統計量は、平行線の検定では計測不能となる場合があります。

- ▶ [検定の統計] グループで [調整済み F] を選択します。
- ▶ 多重比較の調整方法として [逐次 Sidak] を選択します。
- ▶ [続行] をクリックします。
- ▶ [コンプレックス サンプルの順序回帰] ダイアログ ボックスで、[オッズ比] をクリックします。



図 21-5  
[順序回帰: オッズ比] ダイアログ ボックス



- ▶ 「年齢カテゴリ」と「運転頻度」の累積オッズ比の作成を選択します。
- ▶ 「運転頻度」の参照カテゴリとして、最大走行距離より「標準的」年間走行距離の「10 ~ 14,999 マイル/年」を選択します。
- ▶ [続行] をクリックします。
- ▶ [コンプレックス サンプルの順序回帰] ダイアログ ボックスで、[OK] をクリックします。

## 擬 R<sup>2</sup> 乗

図 21-6  
擬 R<sup>2</sup> 乗

Cox と Snell	.179
Nagelkerke	.191
McFadden	.071

従属変数: 議会がガソリン税を制定するべきか (昇順)

モデル: (しきい値) 年齢 性別 前回投票 運転頻度  
リンク関数: ロジット

線型回帰モデルでの決定係数  $R^2$  乗は、予測 (独立) 変数に対する従属変数の分散の比率を要約します。 $R^2$  乗の値が大きくなるほど、変動全体のうち、モデルによって説明される部分が多くなります (最大で 1)。カテゴリ従属変数がある回帰モデルの場合、線型回帰モデルにおける  $R^2$  乗の特性をすべて備えた単一の  $R^2$  乗統計量は計算できません。このため、近似値が計算されます。決定係数の推定には次の方法が使用されます。

- Cox と Snell の  $R^2$  乗 (Cox および Snell, 1989)。ベースライン モデルの対数尤度と比較されるモデルの対数尤度に基づいています。ただし、カテゴリ型の結果では、モデルが「完璧」な場合でも理論上の最大値は 1 未満です。
- Nagelkerke の  $R^2$  乗 (Nagelkerke, 1991)。Cox と Snell の  $R^2$  乗の修正版で、0 から 1 までの範囲に完全に対応するよう統計量の尺度を調整します。
- McFadden の  $R^2$  乗 (McFadden, 1974)。別の方法による修正版で、定数項のみのモデルおよび完全推定モデルの対数尤度カーネルに基づきます。

$R^2$  乗値が「適切」かどうかの基準は、適用する領域によって異なります。この統計量は、それ自体で示唆に富みますが、同じデータの競合するモデルを比較する場合に最も有効です。こうした観点に基づいた場合、 $R^2$  乗統計量が最も大きいモデルが「最適」であると言えます。

## モデル効果の検定

図 21-7  
モデル効果の検定

ソース	df1	df2	調整済み Wald F	有意確率	Sidak の逐 次有意確率
年齢	2.283	31.966	6.215	.004	.003
性別	1.000	14.000	.046	.834	.834
前回投票	1.000	14.000	.076	.787	.787
運転頻度	3.785	52.987	228.015	.000	.000

従属変数: 議会がガソリン税を制定するべきか (昇順)  
モデル: しきい値) 年齢 性別 前回投票 運転頻度  
リンク関数: ロジット

モデル内の各項に対して、その効果が 0 であるかどうかを検定します。有意確率が 0.05 未満の項には、明確な効果があります。したがって、「年齢カテゴリ」と「運転頻度」はモデルに影響しますが、他の主効果は影響しません。データの分析を進めていく過程で、「性別」と「前回の投票」は、モデルから削除することになります。

## パラメータ推定値

パラメータ推定値テーブルは、各予測変数の効果をまとめたものです。このモデルの係数の解釈は、リンク関数の性質を考慮すると困難です。ただし、共変量の係数の符号、および因子レベルの係数の相対値によって、モデルの予測変数の効果を理解ことができます。

- 共変量の場合、正 (負) の係数は、予測変数と結果が正 (逆) の関係にあることを示します。正の係数を持つ共変量の値が増加すると、「より大きな」累積結果カテゴリのいずれかに存在する確率が高くなります。

- 因子の場合、因子レベルの係数が大きくなると、「より大きな」累積結果カテゴリのいずれかに存在する確率が高くなります。因子レベルの係数の符号は、参照カテゴリに対する因子レベルの効果によって異なります。

図 21-8  
パラメータ推定値

パラメータ	回帰係数	標準誤差	95%信頼区間		デザイン効果	Exp (B)	Exp (B) の 95%信頼区間		
			下段	上段			下段	上段	
しきい値	[ガソリン税=1]	-3.343	.104	-3.566	-3.120	1.132	.035	.028	.044
	[ガソリン税=2]	-1.910	.098	-2.120	-1.700	1.058	.148	.120	.183
	[ガソリン税=3]	-.674	.090	-.866	-.482	.915	.510	.421	.618
回帰分析	[年齢=1]	-.324	.079	-.494	-.154	1.793	.723	.610	.858
	[年齢=2]	-.138	.054	-.255	-.022	1.158	.871	.775	.978
	[年齢=3]	-.095	.076	-.257	.068	2.206	.909	.773	1.070
	[年齢=4]	.000 <sup>a</sup>	.	.	.	.	1.000	.	.
	[性別=0]	-.008	.035	-.084	.068	.949	.992	.920	1.071
	[性別=1]	.000 <sup>a</sup>	.	.	.	.	1.000	.	.
	[前回投票=0]	-.011	.039	-.095	.073	1.103	.989	.909	1.076
	[前回投票=1]	.000 <sup>a</sup>	.	.	.	.	1.000	.	.
	[運転頻度=1]	-3.751	.153	-4.079	-3.423	1.117	.023	.017	.033
	[運転頻度=2]	-3.003	.116	-3.251	-2.755	1.226	.050	.039	.064
	[運転頻度=3]	-2.295	.114	-2.540	-2.050	1.585	.101	.079	.129
	[運転頻度=4]	-1.570	.092	-1.769	-1.372	1.078	.208	.171	.254
[運転頻度=5]	-.812	.089	-1.003	-.621	.941	.444	.367	.537	
[運転頻度=6]	.000 <sup>a</sup>	.	.	.	.	1.000	.	.	

従属変数: 議会がガソリン税を制定するべきか (昇川)  
モデル: しきい値, 年齢, 性別, 前回投票, 運転頻度  
リンク関数: ロジット

a. このパラメータは余分であるため、0に設定されます。

パラメータ推定値に基づいて次のように解釈できます。

- 年齢カテゴリが最も高い有権者と比較して、有権者の年齢カテゴリが低くなるほど法案の支持率は高くなる。
- 運転の頻度が高い有権者と比較して、有権者の運転の頻度が低くなるほど法案の支持率は高くなる。
- 変数「性別」と「前回の投票」の係数は、統計的に有意ではなく、他の係数と比較して小さいと言える。

デザイン効果は、上記のパラメータ推定値について計算された標準誤差が、単純な無作為抽出を使用した場合に得られる標準誤差よりも大きくなるものもあれば、小さくなるものもあることを示します。抽出計画情報を分析に取り入れることは特に重要です。たとえば、「年齢カテゴリ」の3番目の係数 ([年齢=3]) と 0 との差が有意になる可能性があるためです。

## 分類

図 21-9  
カテゴリ変数情報

		重み付き度数	重み付き パーセント
議会在ガソリン 税を制定するべ きか	大変賛成	25132.955	21.3%
	賛成	32261.425	27.3%
	反対	29477.417	24.9%
	大変反対	31314.203	26.5%
年齢カテゴリ	18-30	20509.504	17.4%
	31-45	35380.506	29.9%
	46-60	34865.792	29.5%
	>60	27430.198	23.2%
	性別	男性	61424.547
	女性	56761.453	48.0%
前回の投票	いいえ	70607.216	59.7%
	はい	47578.784	40.3%
運転頻度	車なし	3437.137	2.9%
	<10,000 マイル/年	10816.349	9.2%
	10-14,999 マイル/年	32539.364	27.5%
	15-19,999 マイル/年	39179.814	33.2%
	20-29,999 マイル/年	25617.804	21.7%
	>=30,000 マイル/年	6595.532	5.6%
母集団のサイズ		118186.000	100.0%

a. 従属変数の値は昇順でソートされます。

「帰無仮説」モデル（つまり予測変数のないモデル）は、与えられた観測データの全有権者を、最頻グループ、つまり、「賛成」に分類します。したがって、「帰無仮説」モデルは 27.3% 正しいということになります。

図 21-10  
分類テーブル

観測	予測値				
	大変賛成	賛成	反対	大変反対	正答率
大変賛成	7067.567	12130.814	3875.825	2058.750	28.1%
賛成	4271.234	14464.286	7320.767	6205.137	44.8%
反対	2024.816	11703.368	7108.487	8640.746	24.1%
大変反対	889.869	8169.109	6946.522	15308.703	48.9%
全体のパーセント	12.1%	39.3%	21.4%	27.3%	37.2%

従属変数: 議会在ガソリン税を制定するべきか (昇順)  
モデル: (しきい値), 年齢, 性別, 前回投票, 運転頻度  
リンク関数: ロジット

分類テーブルでは、実際にモデルを使用した結果が表示されます。ケースごとの予測応答は、モデル予測確率が最も高い応答カテゴリです。ケースは、最終的な抽出の重み付けをされているため、分類テーブルでは、母集団で予測されるモデルのパフォーマンスが示されます。

- 対角上のセルが正しい予測値です。
- 対角線から外れたセルは不正な予測値です。

モデルでは、9.9%以上（つまりケースの37.2%）を正しく分類しています。特に、「賛成」または「大変反対」の分類は非常に優れています。一方、「反対」の分類はやや劣っています。

## オッズ比

累積オッズは、従属変数が特定の応答カテゴリより大きな値を取る確率に対する、応答カテゴリ以下の値を取る確率の比率として定義されます。累積オッズ比は、予測変数値ごとの累積オッズの比率であり、指数化されたパラメータ推定値と密接な関連があります。興味深い点として、累積オッズ比自体は応答カテゴリに依存しません。

図 21-11  
年齢カテゴリの累積オッズ比

	累積オッズ比	95% 信頼区間		デザイン効果	デザイン効果の平方根
		下段	上段		
年齢 18-30 対 >60	1.383	1.166	1.639	1.793	1.339
カテゴリ 31-45 対 >60	1.148	1.022	1.290	1.158	1.076
カテゴリ 46-60 対 >60	1.100	.935	1.294	2.206	1.485

従属変数: 議会がガソリン税を制定するべきか (昇順)  
モデル: (しきい値) 年齢, 性別, 前回投票, 運転頻度  
リンク関数: ロジット

a. 計算に使用される因子と共変量は値 年齢カテゴリ=>60; 性別=女性; 前回の投票=はい; 運転頻度=>30,000 マイル/年 に固定されます。

このテーブルは、「年齢カテゴリ」の因子レベルの累積オッズ比を示しています。表示された値は、[>60] の累積オッズと比較した、[18 ~ 30] から [46 ~ 60] までの累積オッズの比率を示します。したがって、テーブルの最初の行に記載されているオッズ比 1.383 の場合、年齢層が 18 ~ 30 歳の人々の累積オッズは、61 歳以上の人々の累積オッズの 1.383 倍であることを意味します。「年齢カテゴリ」は交互作用項に含まれていないため、これらのオッズ比は指数化されたパラメータ推定値の比率になります。たとえば、[18 ~ 30] と [>60] の累積オッズ比は  $1.00 / 0.723 = 1.383$  となります。

図 21-12  
運転頻度のオッズ比

運転頻度	車なし 対 10-14,999 マイル/年 <sup>a</sup>	累積オッズ比	95% 信頼区間		デザイン効果	デザイン効果の平方根
			下段	上段		
<10,000 マイル/年 対 10-14,999 マイル/年 <sup>a</sup>		4.288	2.878	6.390	2.345	1.531
15-19,999 マイル/年 対 10-14,999 マイル/年 <sup>a</sup>		2.030	1.656	2.488	1.838	1.356
20-29,999 マイル/年 対 10-14,999 マイル/年 <sup>a</sup>		.484	.430	.546	1.450	1.204
>=30,000 マイル/年 対 10-14,999 マイル/年 <sup>a</sup>		.227	.193	.267	2.095	1.448
		.101	.079	.129	1.585	1.259

従属変数: 議会がガソリン税を制定するべきか (昇順)  
モデル: (しきい値) 年齢, 性別, 前回投票, 運転頻度  
リンク関数: ロジット

a. 計算に使用される因子と共変量は値 年齢カテゴリ>60; 性別=女性; 前回の投票=はい; 運転頻度 >=30,000 マイル/年 に固定されます。

このテーブルは、参照カテゴリとして [10 ~ 14,999 マイル/年] を使用して、「運転頻度」の因子レベルの累積オッズ比を示したものです。「運転頻度」は交互作用項に含まれていないため、これらのオッズ比は指数化されたパラメータ推定値の比率になります。たとえば、[20 ~ 29,999 マイル/年] と [10 ~ 14,999 マイル/年] の累積オッズ比は  $0.101 / 0.444 = 0.227$  となります。

## 一般化累積モデル

図 21-13  
平行線の検定

df1	df2	調整済み Wald F	有意確率	Sidak の逐次有意確率
8.769	122.767	1.894	.061	.392

従属変数: 議会がガソリン税を制定するべきか (昇順)  
モデル: (しきい値) 年齢, 性別, 前回投票, 運転頻度  
リンク関数: ロジット

平行線の検定は、すべての応答カテゴリについてパラメータが同じであるという仮定が妥当かどうかを評価する場合に役立ちます。この検定では、すべてのカテゴリに対して係数の集合が 1 つある推定モデルと、カテゴリごとに異なる係数の集合がある一般化モデルとを比較します。

Wald F 検定は、平均線の仮定に関する対比行列のオムニバス検定です。この検定では、漸近的に正確な p 値が得られます。サンプルが小規模から中規模の場合は、調整済み Wald F 統計量が適しています。有意確率値は 0.05 に近い値になります。これは、一般化モデルによりモデルの適合度が向上する可能性があることを示します。ただし、調整済み逐次 Sidak 検定の結果は高い有意確率値 (0.392) を示しており、全体的には、平行線の仮定を棄却するための明確な証拠はありません。逐次 Sidak 検定では、最初に個別の対比 Wald 検定を行います。これにより、全体の p 値が

得られます。この結果は、オムニバス Wald 検定の結果と同等です。この例の場合、両者の結果に大きな違いが出たのはやや意外です。しかし、これは検定で対比が多かった点、計画の自由度が比較的小さかった点によるものと考えられます。

図 21-14

一般化累積モデルのパラメータ推定値（一部を抜粋）

議会在ガレソン税を制定するべきか	パラメータ	回帰係数	標準誤差	95%信頼区間	
				下段	上段
大差賛成	(しきい値)	-3.681	.221	-4.155	-3.207
	[年齢=1]	-.320	.096	-.525	-.115
	[年齢=2]	-.075	.071	-.227	.077
	[年齢=3]	-.022	.073	-.180	.135
	[年齢=4]	.000 <sup>a</sup>	.		
	[性別=0]	-.082	.054	-.197	.033
	[性別=1]	.000 <sup>a</sup>	.		
	[前回投票=0]	.008	.052	-.104	.120
	[前回投票=1]	.000 <sup>a</sup>	.		
	[運動頻度=1]	-4.096	.267	-4.669	-3.523
	[運動頻度=2]	-3.367	.237	-3.876	-2.857
	[運動頻度=3]	-2.678	.224	-3.158	-2.199
	[運動頻度=4]	-1.928	.213	-2.384	-1.471
	[運動頻度=5]	-1.015	.252	-1.555	-.476
[運動頻度=6]	.000 <sup>a</sup>	.			
賛成	(しきい値)	-1.963	.153	-2.291	-1.635
	[年齢=1]	-.385	.095	-.587	-.182
	[年齢=2]	-.130	.069	-.279	.018
	[年齢=3]	-.139	.101	-.356	.077
	[年齢=4]	.000 <sup>a</sup>	.		
	[性別=0]	-.004	.040	-.090	.082
	[性別=1]	.000 <sup>a</sup>	.		
	[前回投票=0]	.009	.059	-.117	.135
	[前回投票=1]	.000 <sup>a</sup>	.		
	[運動頻度=1]	-3.867	.318	-4.549	-3.185
	[運動頻度=2]	-3.005	.175	-3.380	-2.630
	[運動頻度=3]	-2.290	.187	-2.691	-1.888
	[運動頻度=4]	-1.633	.166	-1.988	-1.278
	[運動頻度=5]	-.909	.137	-1.204	-.615
[運動頻度=6]	.000 <sup>a</sup>	.			

また、一般化モデルの係数の推定値と平行線の仮定での推定値には大きな違いは見られません。

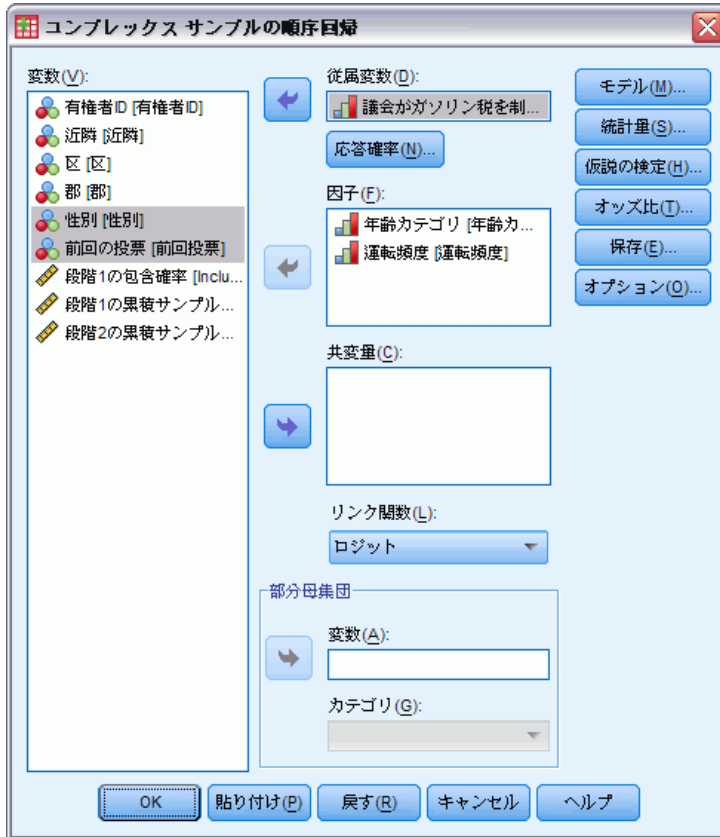
## 有意でない予測変数の削除

モデル効果の検定の結果、「性別」と「前回の投票」のモデル係数は0との差が十分に有意ではないことがわかりました。

- ▶ 縮小モデルを作成するために、[コンプレックス サンプルの順序回帰] ダイアログ ボックスを再表示します。

- ▶ [計画] ダイアログ ボックスで、[続行] をクリックします。

図 21-15  
[順序回帰] ダイアログ ボックス



- ▶ 因子として「性別」と「前回の投票」の選択を解除します。
- ▶ [オプション] をクリックします。



図 21-16  
[順序回帰: オプション] ダイアログ ボックス

- ▶ [反復の記述を表示] を選択します。

反復の記述は、推定アルゴリズムで発生した問題を診断する場合に役立ちます。

- ▶ [続行] をクリックします。
- ▶ [コンプレックス サンプルの順序回帰] ダイアログ ボックスで、[OK] をクリックします。

## 警告

図 21-17  
縮小モデルの警告

段階 2 分法の最大段階数に到達後、対数尤度値は増加できません。  
警告が表示されましたが、CS ORDINAL の手続は続行されます。この後の結果は、最後の反復に基づいたものとなります。モデルの適合度の有効性は保証できません。  
次のメッセージが一般化累積モデルに適用されます。  
段階 2 分法の最大段階数に到達後、対数尤度値は増加できません。

警告では、パラメータ推定値の現在値の変化（ステップ）に応じて対数尤度を増加できなかったために、パラメータ推定値が収束基準に達する前に縮小モデルの推定が終了したことを通知します。

図 21-18  
縮小モデルの警告

変数名	R 段階と分	階 -2 対 数尤度	しきい値			回帰分析								
			[ガンリン]係数]	[ガンリン]係数]	[ガンリン]係数]	[年齢]	[年齢]	[年齢]	[運動頻度]	[運動頻度]	[運動頻度]	[運動頻度]	[運動頻度]	[運動頻度]
0	0	303040.941	-1.359	-0.66	1.020	300	-000	000	000	000	000	000	000	000
1	0	303567.549	-0.242	-1.881	-704	-323	-137	-094	-3.841	-2.970	-2.248	-1.563	-858	
2	0	303336.336	-3.327	-1.897	-664	-326	-139	-095	-3.740	-2.908	-2.291	-1.568	-811	
3	0	303335.933	-3.333	-1.900	-664	-326	-139	-096	-3.750	-3.003	-2.295	-1.570	-812	
4	0	303335.933	-3.333	-1.900	-664	-326	-139	-096	-3.750	-3.003	-2.295	-1.570	-812	
5 <sup>a</sup>	0	303335.933	-3.333	-1.900	-664	-326	-139	-096	-3.750	-3.003	-2.295	-1.570	-812	

異なるパラメータは表示されません。これらのパラメータの値は、すべての段階で常に同じです。  
従属変数： 議会がガンリン税を制定するべきか (昇順)  
モデル： しきい値) 年齢 性別 前回投票 運動頻度  
リンク関数： ロジット

a. 段階と分法の最大段階数に到達後、対数尤度値は増加できません。  
b. パラメータの推定に Newton-Raphson 法が使用されました。

反復の記述を参照すると、直前の数回の反復におけるパラメータ推定値の変化はわずかで、警告メッセージについてそれほど重大視する必要がないことを確認できます。

## モデルの比較

図 21-19  
縮小モデルの擬 R<sup>2</sup> 乗値

Cox と Snell	.179
Nagelkerke	.191
McFadden	.071

従属変数： 議会がガンリン税を制定するべきか (昇順)

モデル： しきい値) 年齢 性別 前回投票 運動頻度  
リンク関数： ロジット

縮小モデルと元のモデルの R<sup>2</sup> 乗値は同じになります。これは、縮小モデルを肯定する根拠となります。

図 21-20  
縮小モデルの分類テーブル

観測	予測値				
	大変賛成	賛成	反対	大変反対	正答率
大変賛成	7067.567	12823.258	3183.380	2058.750	28.1%
賛成	4271.234	15684.090	6100.963	6205.137	48.6%
反対	2024.816	13157.809	5654.047	8640.746	19.2%
大変反対	889.869	9226.578	5889.053	15308.703	48.9%
全体のパーセント	12.1%	43.1%	17.6%	27.3%	37.0%

従属変数： 議会がガンリン税を制定するべきか (昇順)  
モデル： しきい値) 年齢 運動頻度  
リンク関数： ロジット

この分類テーブルには、問題をやや複雑にする面があります。縮小モデルの全体の分類率は 37.0% です。これは、元のモデルと同等であり、縮小モデルを肯定する根拠となります。しかし、縮小モデルでは、有権者

の 3.8% の予測応答が [反対] から [賛成] にシフトしてしまいます。シフトした有権者の半数以上が、[反対] または [大変反対] と回答したことが観測されています。これは、非常に重要な相違点で、縮小モデルを選択する前に慎重に考慮する必要があります。

## 集計 (報告書 データ列)

[コンプレックス サンプルの順序回帰] 手続きを使用して、有権者の人口統計に基づいた法案の支持水準の競合するモデルを作成しました。平行線の検定では、一般化累積モデルが必要でないことがわかりました。モデル効果の検定では、「性別」および「前回の投票」をモデルから削除できることがわかりました。また、縮小モデルが、元のモデルと比べても擬  $R^2$  乗値、および全体の分類率に関しては十分有効であることを確認できました。しかし、縮小モデルでは、[賛成] と [反対] の分割で元のモデルと比べて多くの有権者が誤って分類されます。こうした点を考慮し、議員は元のモデルを引き続き調査で採用することにしました。

## 関連手続き

[コンプレックス サンプルの順序回帰] 手続きは、コンプレックス サンプルリング法に従ってケースが抽出された場合に、順序変数をモデリングする上で便利なツールです。

- **コンプレックス サンプルのサンプリング ウィザード**は、コンプレックス サンプル計画仕様の指定およびサンプル抽出に使用されます。サンプリング ウィザードにより作成される抽出計画ファイルは、デフォルト分析計画を含んでおり、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- **コンプレックス サンプルの分析準備ウィザード**は、既存のコンプレックス サンプル用の分析仕様の指定に使用されます。サンプリング ウィザードにより作成される分析計画ファイルは、計画に基づいて抽出したサンプルを分析するときに [計画] ダイアログ ボックスで指定できます。
- **コンプレックス サンプルの一般線型モデル**手続きでは、尺度の応答をモデル化できます。
- **コンプレックス サンプルのロジスティック回帰**手続きでは、カテゴリ応答をモデル化できます。

# コンプレックス サンプルの Cox 回帰

コンプレックス サンプルの Cox 回帰手続きでは、コンプレックス サンプリ  
ング法によって抽出されたサンプルに対して、生存分析を実行します。

## 時間依存の予測変数の使用 (コンプレックス サンプルの Cox 回帰)

政府の法執行機関は、管轄地域での累犯率に注意を払っています。累犯率  
の測定方法の 1 つは、犯罪者の 2 回目の逮捕までの時間です。法執行機関  
は、コンプレックス サンプリング法によって抽出されたサンプルで Cox 回  
帰を使用して、再逮捕までの時間をモデル化したいと考えていますが、年齢  
カテゴリ全体にハザードの比例仮定が適用できないことを心配しています。

2003 年 6 月に最初の逮捕から釈放された者を抽出された部署から  
選択し、2006 年 6 月末までの彼らの犯罪歴を調査しました。サンプ  
ルは、recidivism\_cs\_sample.sav に収集されています。使用した抽  
出計画は、recidivism\_cs.csplan に保存されています。調査では確  
率比例 (PPS) 法を採用したため、結合選択確率を保持したファイル  
(recidivism\_cs\_jointprob.sav) も用意されています。詳細は、A 付録  
p. 281 [サンプル ファイル](#) を参照してください。コンプレックス サンプルの  
Cox 回帰を使用して、ハザードの比例仮定の有効性を評価し、適切な場合  
は時間依存の予測変数でモデルを当てはめてください。

## データの準備

データセットには、最初の逮捕の釈放日と 2 度目の逮捕日が含まれていま  
す。Cox 回帰では生存時間を分析するので、最初の逮捕の釈放日から 2 度  
目の逮捕日までの時間を計算する必要があります。

ただし、[2度目の逮捕日 [日付2]] には、日付変数の欠損値 10/03/1582  
を持つケースが含まれています。これらは、2 度目の違法行為がなかった  
人たちで、右側打ち切りのケースとしてモデルに含める必要があります。  
追跡期間の最終日は 2006 年 6 月 30 日なので、10/03/1582 を 06/30/2006  
に再割り当てします。

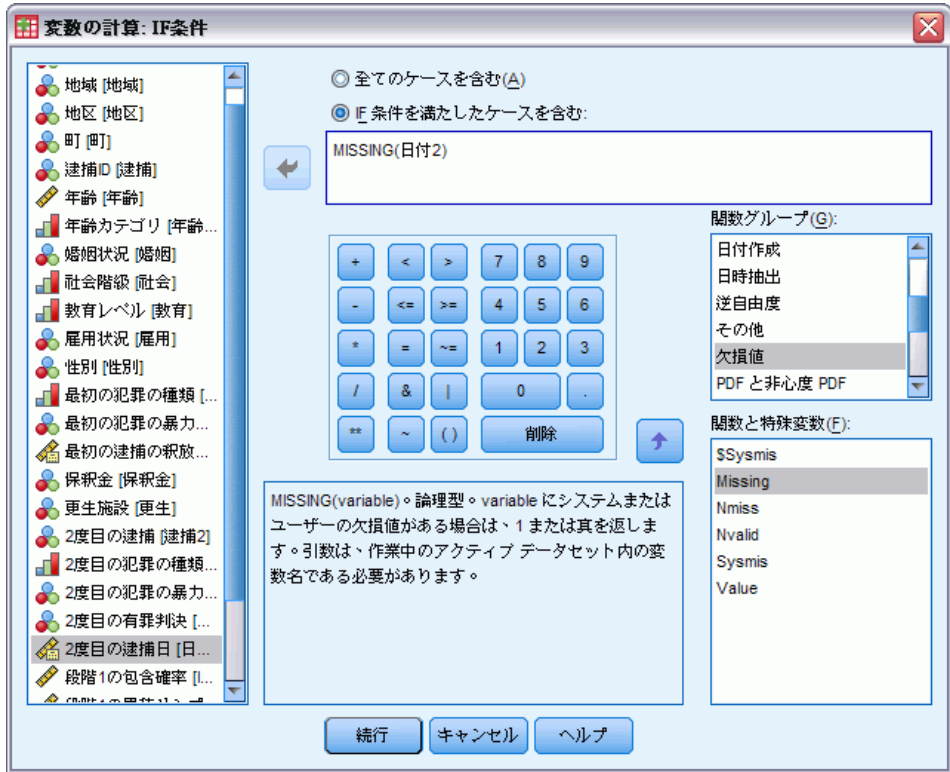
- ▶ 値を再割り当てするには、メニューから次の項目を選択します。  
変換 > 変数の計算...

図 22-1  
[変数の計算] ダイアログ ボックス



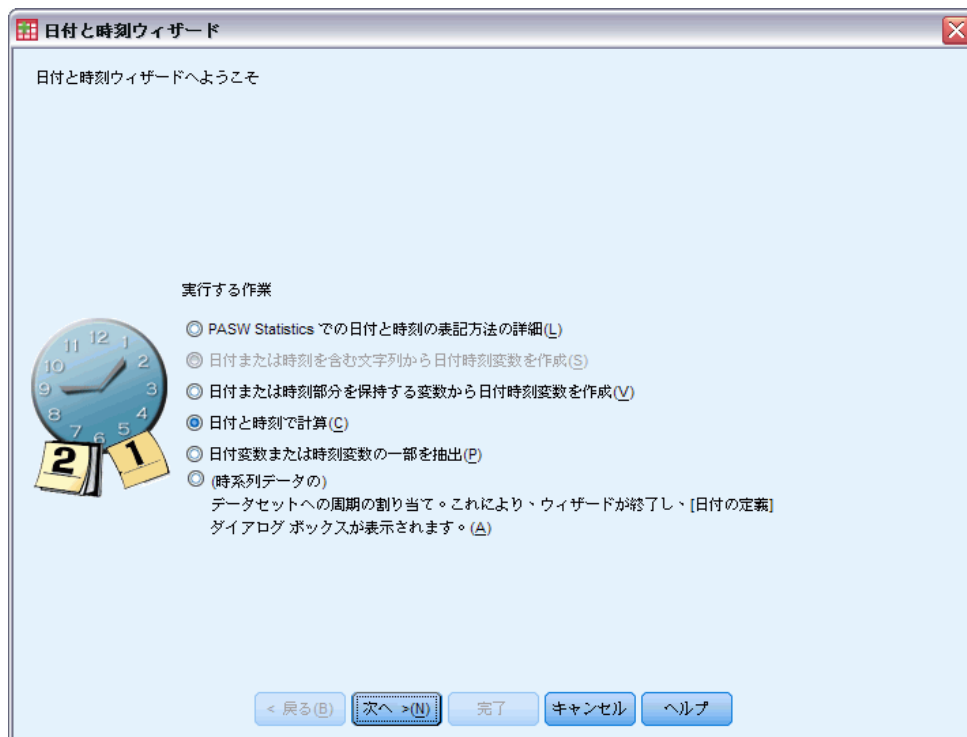
- ▶ 目標変数として「日付2」と入力します。
- ▶ 数式として「DATE.DMY(30,6,2006)」と入力します。
- ▶ [IF] をクリックします。

図 22-2  
[変数の計算: IF 条件] ダイアログ ボックス



- ▶ [If 条件を満たしたケースを含む] チェック ボックスをオンにします。
- ▶ 数式として「MISSING(日付2)」と入力します。
- ▶ [続行] をクリックします。
- ▶ [変数の計算] ダイアログ ボックスで [OK] をクリックします。
- ▶ 次に、最初の逮捕から 2 度目の逮捕までの時間を計算するには、メニューから次の項目を選択します。  
変換 > 日付と時刻ウィザード...

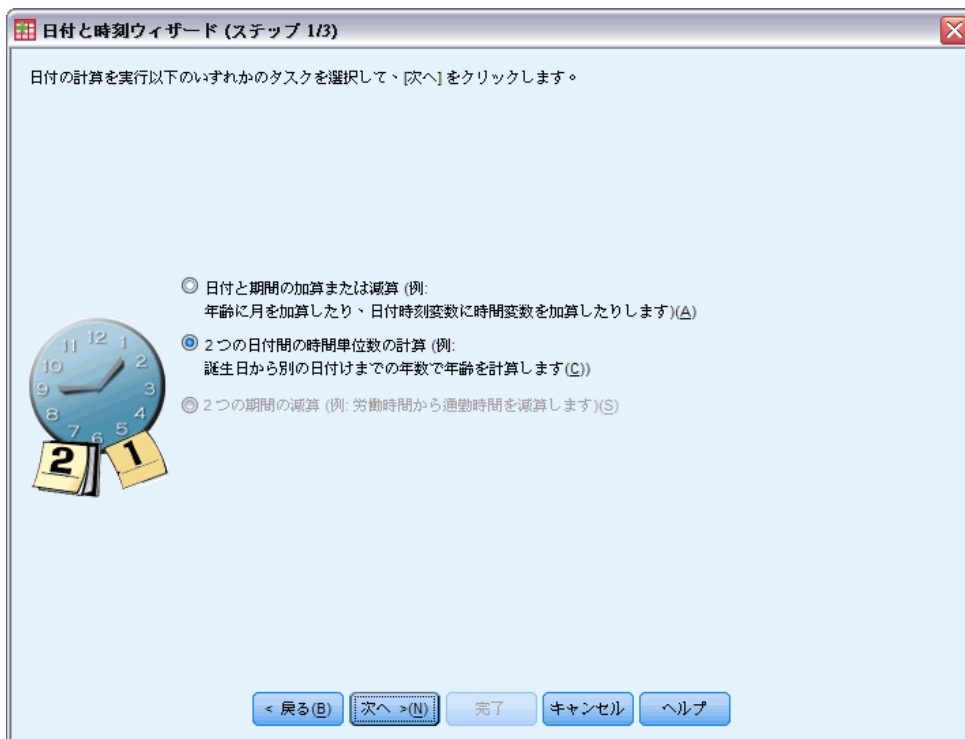
図 22-3  
日付と時刻ウィザードの [ようこそ] ステップ



▶ [日付と時刻で計算] を選択します。

▶ [次へ] をクリックします。

図 22-4  
日付と時刻ウィザードの [日付の計算を実行] ステップ

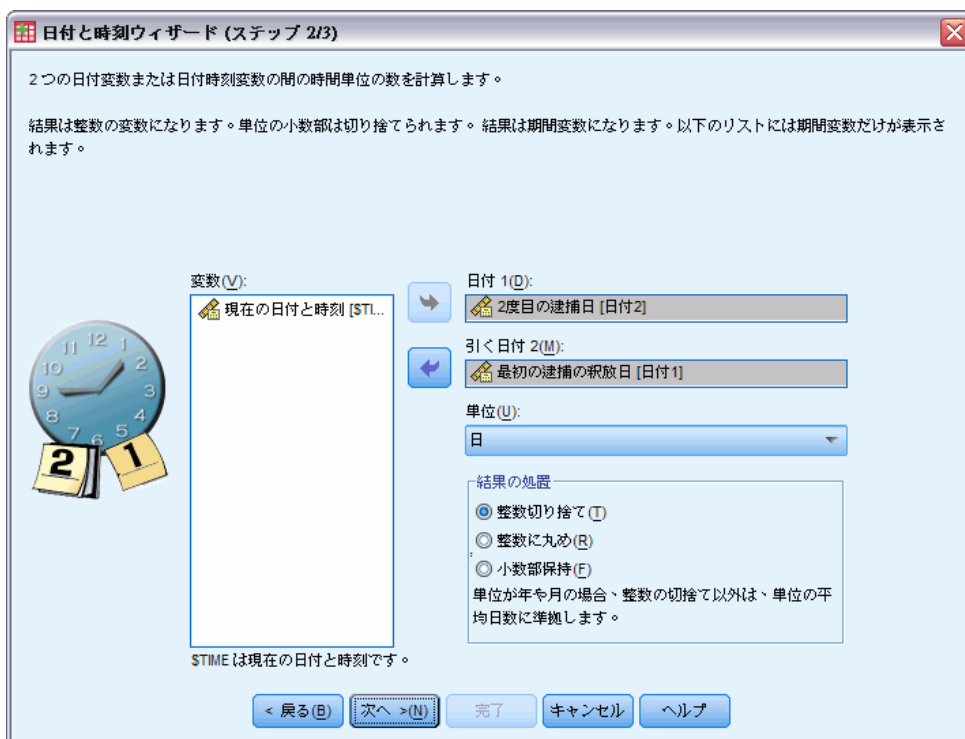


▶ [2つの日付間の時間単位数の計算] を選択します。

▶ [次へ] をクリックします。



図 22-5  
日付と時刻ウィザードの [2 つの日付間の時間単位数の計算] ステップ



日付と時刻ウィザード (ステップ 2/3)

2つの日付変数または日付時刻変数の間の時間単位数の数を計算します。

結果は整数の変数になります。単位の小数部は切り捨てられます。結果は期間変数になります。以下のリストには期間変数だけが表示されます。

変数 (V):  
現在の日付と時刻 [STL...]

日付 1 (D):  
2度目の逮捕日 [日付2]

引く日付 2 (M):  
最初の逮捕の釈放日 [日付1]

単位 (U):  
日

結果の処理

整数切り捨て (I)  
 整数に丸め (R)  
 小数部保持 (F)

単位が年や月の場合、整数の切捨て以外は、単位の平均日数に準拠します。

< 戻る (B) 次へ > (N) 完了 キャンセル ヘルプ

- ▶ 日付 1 として [2度目の逮捕日 [日付2]] を選択します。
- ▶ 日付 1 から引く日付として [最初の逮捕の釈放日 [日付1]] を選択します。
- ▶ 単位として [日] を選択します。
- ▶ [次へ] をクリックします。

図 22-6  
日付と時刻ウィザードの [計算] ステップ

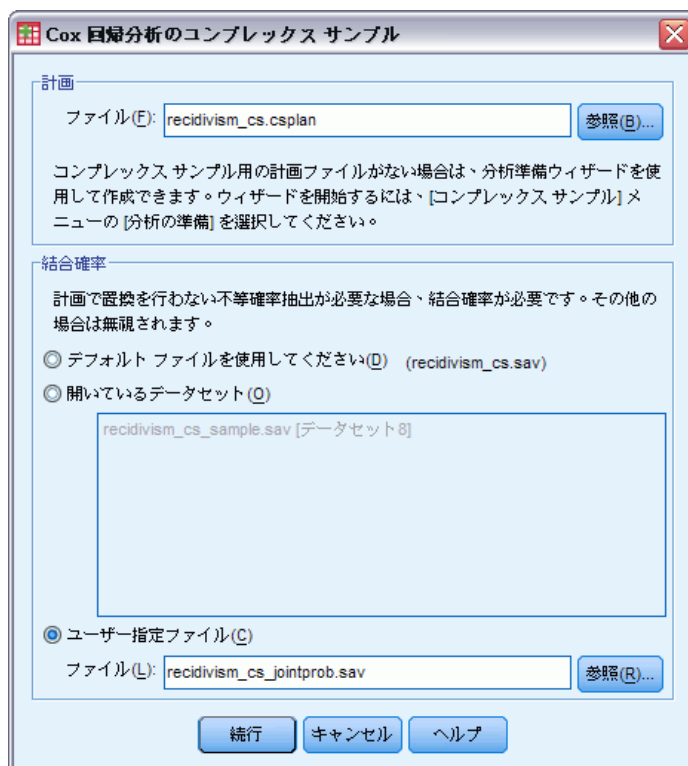


- ▶ 2つの日付間の時間を表す変数名として「イベントまでの時間」と入力します。
- ▶ 変数ラベルとして「2度目の逮捕までの時間」と入力します。
- ▶ [完了] をクリックします。

## 分析の実行

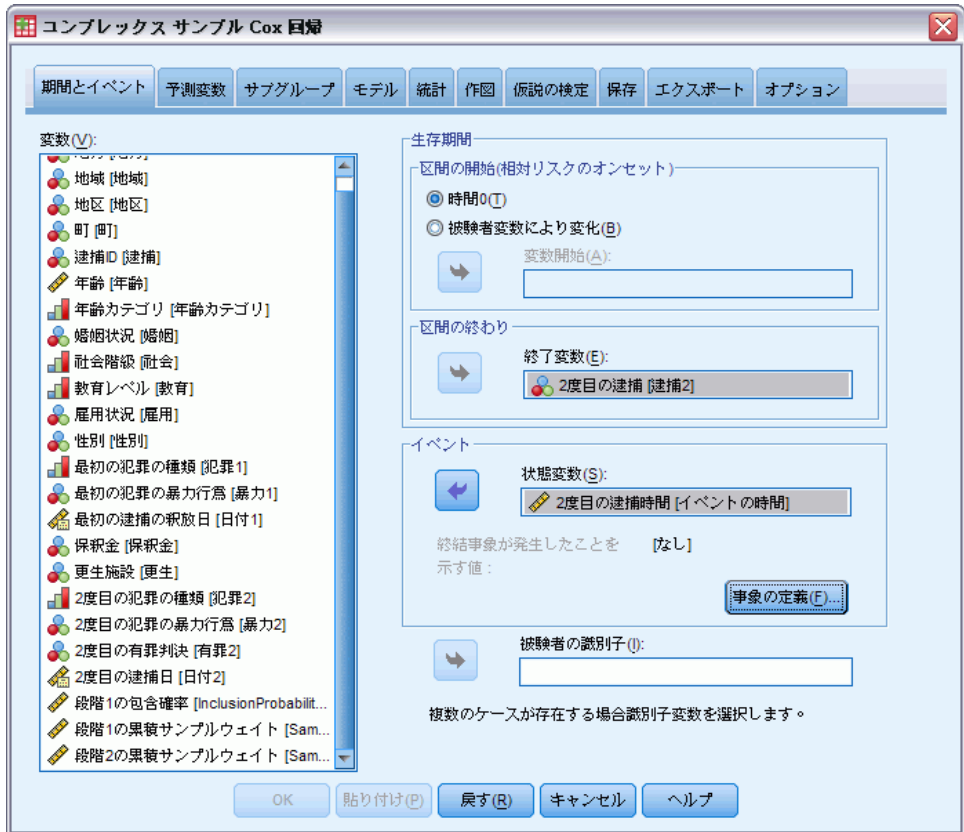
- ▶ コМПЛЕКС サンプルの Cox 回帰分析を実行するには、メニューから次の項目を選択します。  
分析 > КОМПЛЕКС サンプル > Cox 回帰分析...

図 22-7  
[Cox 回帰分析のコンプレックス サンプル] ダイアログ ボックス



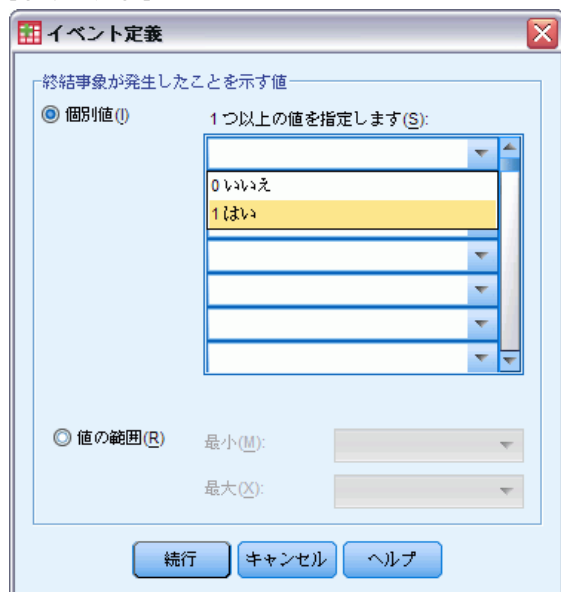
- ▶ 計画ファイルとして、サンプル ファイルのディレクトリを参照して recidivism\_cs.csplan を選択します。
- ▶ [結合確率] グループから [ユーザー指定ファイル] を選択し、サンプル ファイルのディレクトリを参照して recidivism\_cs\_jointprob.sav を選択します。
- ▶ [実行] をクリックします。

図 22-8  
[Cox 回帰分析] ダイアログ ボックスの [時刻およびイベント] タブ



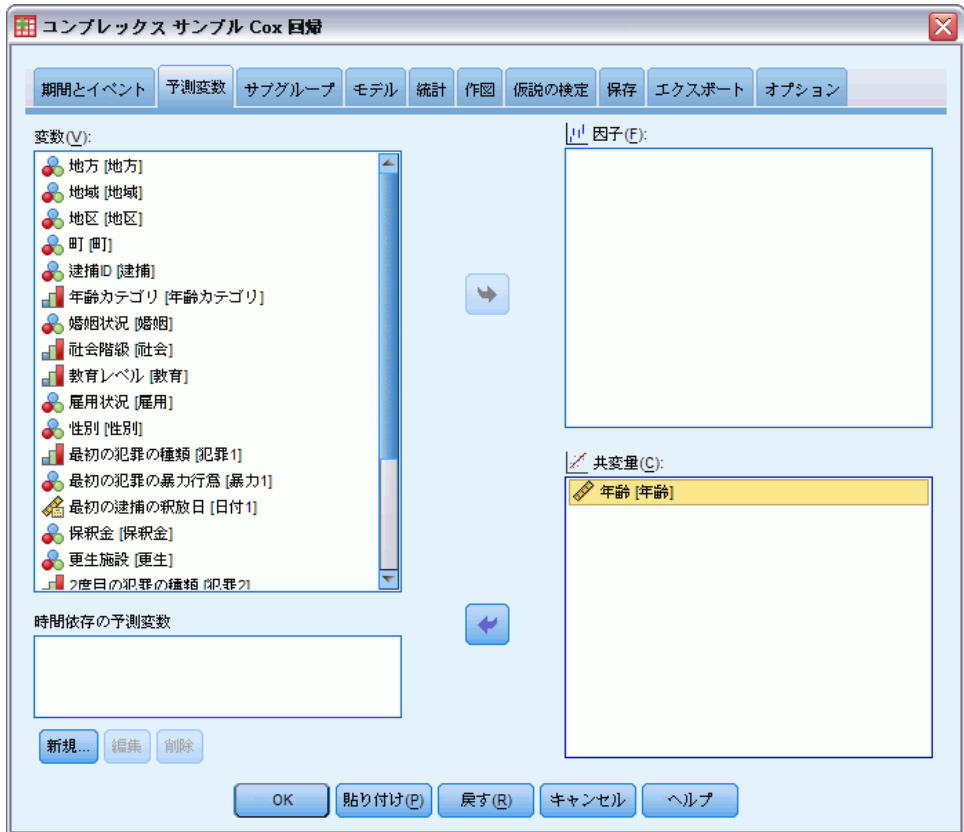
- ▶ 区間の終わり定義する変数として [2 度目の逮捕までの時間 [イベントまでの時間]] を選択します。
- ▶ イベントの発生を定義する変数として [2 度目の逮捕 [逮捕 2]] を選択します。
- ▶ [事象の定義] をクリックします。

図 22-9  
[事象の定義] ダイアログ ボックス



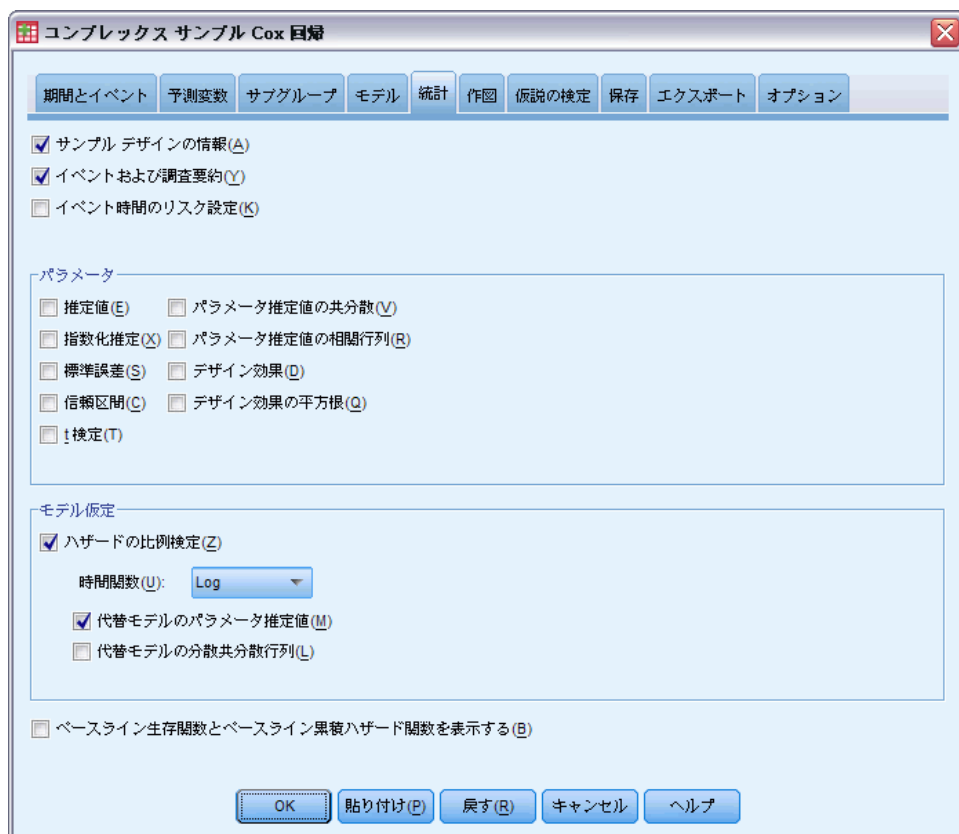
- ▶ 関心のある（再逮捕）のイベントが発生したことを示す値として、[1 はい] を選択します。
- ▶ [続行] をクリックします。
- ▶ [予測変数] タブをクリックします。

図 22-10  
[Cox 回帰分析] ダイアログ ボックスの [予測変数] タブ



- ▶ 共変量として [年齢 [年齢]] を選択します。
- ▶ [統計] タブをクリックします。

図 22-11  
[Cox 回帰分析] ダイアログ ボックスの [統計] タブ



- ▶ [モデル仮定] グループで [ハザードの比例検定] を選択してから、時刻関数として [ログ] を選択します。
- ▶ [代替モデルのパラメータ推定値] を選択します。
- ▶ [OK] をクリックします。

## サンプル デザインの情報

図 22-12  
サンプル デザインの情報

			度数
重み付けしない度数	有効数	被験者	5687
		ケース	5687
		無効なケース	0
		ケースの合計	5687
有効数		被験者母集団のサイズ	307583.898
第 1 段階	有効数	ストラータ	4
		単位	20
有効数		抽出計画の自由度	16

この表には、モデルの推定に関するサンプル デザインの情報が含まれています。

- 各被験者に 1 つのケースがあり、分析では 5,687 件すべてのケースが使用されています。
- サンプルは、推定母集団全体の 2% 未満に相当します。
- 計画の第 1 段階では、4 つのストラータとストラータ当たり 5 単位、合計 20 単位が要求されました。抽出計画の自由度は、 $20-4=16$  と推定されています。

## モデル効果の検定

図 22-13  
モデル効果の検定

ソース	df1	df2	Wald F	有意確率
年齢	1.000	16.000	504.787	1.58E-13

生存時間変数: 2度目の逮捕までの時間  
事象状態変数: 2度目の逮捕 = 1  
モデル: 年齢

ハザードの比例モデルでは、予測変数の「年齢」の有意確率が 0.05 未満であるため、モデルに影響すると思われます。

## ハザードの比例検定

図 22-14  
ハザードの比例に関する全体的な検定

df1	df2	Wald F	有意確率
1.000	16.000	29.924	5.14E-5

生存時間変数: 2度目の逮捕までの時間  
事象状態変数: 2度目の逮捕 = 1  
モデル: 年齢, 年齢\*\_TF



図 22-15  
代替モデルのパラメータ推定値。

パラメータ	回帰係数	標準誤差	90% 信頼区間	
			下段	上段
年齢	-.002	.014	-.025	.002
年齢*_TF <sup>a</sup>	-.012	.002	-.016	-.009

生存時間変数: 2度目の逮捕までの時間  
 事象状態変数: 2度目の逮捕 = 1  
 モデル: 年齢, 年齢\*\_TF

a. 時間関数: Log

ハザードの比例に関する全体的な検定の有意確率は 0.05 未満であり、ハザードの比例仮定がみたされていないことを示しています。代替のモデルにログの時刻関数が使用されているため、この時間依存の予測変数を簡単に複製することができます。

## 時間依存の予測変数の追加

- ▶ [コンプレックス サンプル Cox 回帰] ダイアログ ボックスを再表示し、[予測変数] タブをクリックします。

- ▶ [新規] をクリックします。

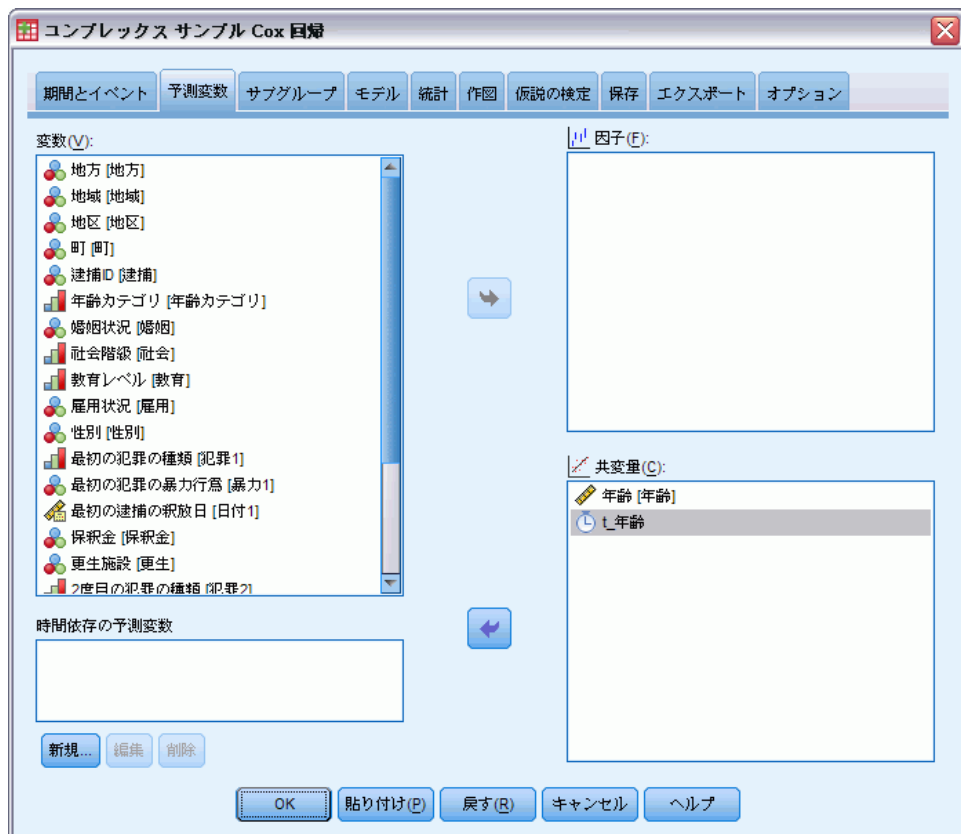
図 22-16

[Cox サンプル回帰分析: 時間依存の予測変数を定義] ダイアログ ボックス



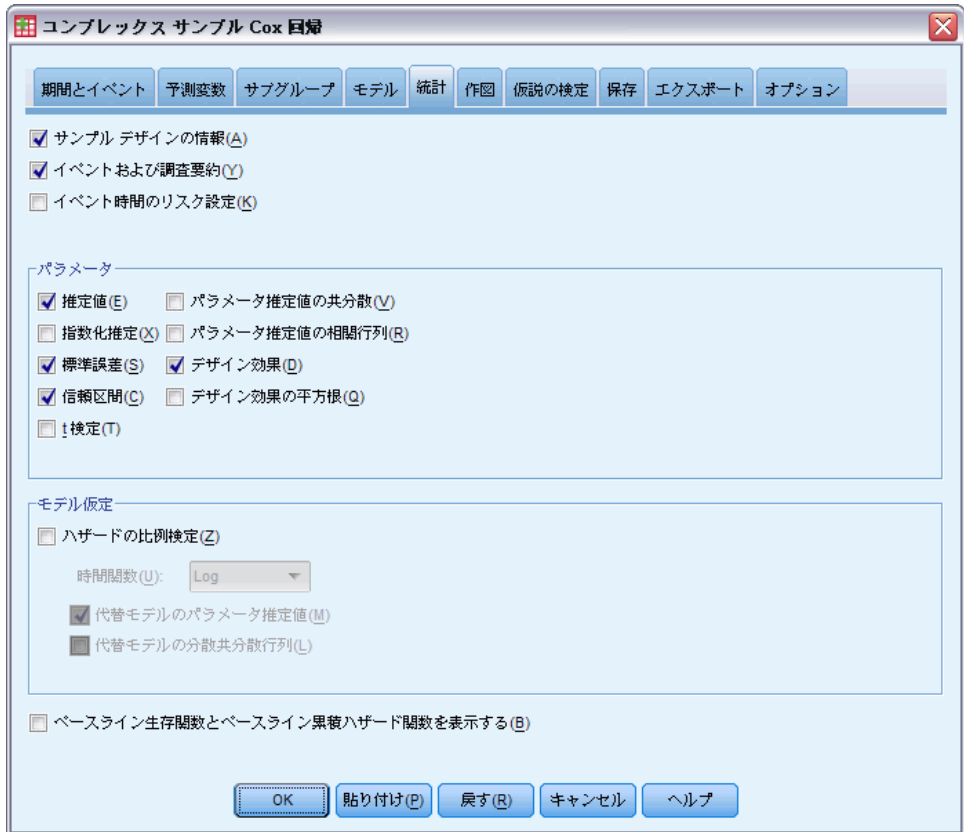
- ▶ 定義する時間依存の予測変数の名前として「t\_年齢」と入力します。
- ▶ 数式として「ln(T\_)\*年齢」と入力します。
- ▶ [続行] をクリックします。

図 22-17  
[Cox 回帰分析] ダイアログ ボックスの [予測変数] タブ



- ▶ 共変量として「t\_年齢」を選択します。
- ▶ [統計] タブをクリックします。

図 22-18  
[Cox 回帰分析] ダイアログ ボックスの [予測変数] タブ



- ▶ [パラメータ] グループで、[推定値]、[標準誤差]、[信頼区間]、[デザイン効果] を選択します。
- ▶ [モデル仮定] グループで [ハザードの比例検定] と [代替モデルのパラメータ推定値] の選択を解除します。
- ▶ [OK] をクリックします。

## モデル効果の検定

図 22-19  
モデル効果の検定

ソース	df1	df2	Wald F	有意確率
年齢	1.000	16.000	.015	.991
t_年齢	1.000	16.000	29.924	5.14E-5

生存時間変数: 2度目の逮捕までの時間  
 事象状態変数: 2度目の逮捕 = 1  
 モデル: 年齢, t\_年齢

時間依存の予測変数を追加したことにより、「年齢」の有意確率は 0.91 になり、モデルへの影響が「t\_年齢」の有意確率より優先されることを示しています。

## パラメータ推定値

図 22-20  
パラメータ推定値

パラメータ	回帰係数	標準誤差	95% 信頼区間		デザイン効果
			下段	上段	
年齢	-.002	.014	-.030	.027	.702
t_年齢	-.012	.002	-.017	-.008	.666

生存時間変数: 2度目の逮捕までの時間  
事象状態変数: 2度目の逮捕 = 1  
モデル: 年齢, t\_年齢

パラメータ推定値と標準誤差を見ると、ハザードの比例検定から代替モデルが複製されたことがわかります。モデルを明示的に指定することにより、追加のパラメータ統計量およびプロットを要求できます。ここでは、デザイン効果を要求しました。「t\_年齢」の値が 1 未満ということは、「t\_年齢」の標準誤差は、データセットが単純な無作為抽出であると仮定したときに得られる標準誤差よりも小さいことを示しています。この場合、「t\_年齢」の効果は統計的には有意であるけれども、信頼区間の幅が大きくなります。

## コンプレックス サンプルの Cox 回帰の被験者ごとに複数のケース

研究者は、虚血性脳卒中中で数回の困難に直面した後リハビリプログラムを終えた患者の生存時間について調査しています。

**被験者ごとに複数のケース。**患者の病歴変数を表す変数は、予測変数として有用です。時間の経過とともに、患者は病歴が変わる重大な事象を経験する可能性があります。このデータセットでは、心筋梗塞の発生、虚血性脳卒中、または出血性脳卒中が注意され、イベントの時間が記録されます。この手続きで計算可能な時間依存の共変量を作成し、その情報をモデルに含めることもできますが、被験者ごとに複数のケースを使用するほうが便利です。また、患者の病歴が複数の変数において記録されるように、変数はすでにコード化されています。したがって、データセットを再構成する必要があります。

**左側の切り捨て。**虚血性脳卒中中の発生時をリスクの起点とします。ただし、リハビリプログラムで生存した患者のみを含むサンプルの場合、観測される生存時間はリハビリの長さによって「膨張している」という意味で、サンプルは左側が切り捨てられます。調査に投入した時間としてリハビリを終えた時間を指定することでこれを説明できます。

**抽出計画なし。**データセットは、コンプレックス サンプル計画から収集されたものではないので、単純な無作為抽出と見なされます。コンプレックス サンプルの Cox 回帰を使用して、分析計画を作成する必要があります。

データセットは、stroke\_survival.sav に収集されています。詳細は、A 付録 p.281 [サンプル ファイル](#) を参照してください。再構成データ ウィザードで分析用のデータを準備してから、分析準備ウィザードで単純な無作為抽出計画を作成し、最後にコンプレックス サンプルの Cox 回帰を使用して生存時間のモデルを構築します。

## 分析用データの準備

データを再構成する前に、再構成で使用するための補助変数を 2 つ作成する必要があります。

- ▶ 新しい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...

図 22-21  
[変数の計算] ダイアログ ボックス



- ▶ 目標変数として「開始時間2」と入力します。
- ▶ 数式として「時間1」と入力します。
- ▶ [OK] をクリックします。

- ▶ もう一度、[変数の計算] ダイアログ ボックスを表示します。

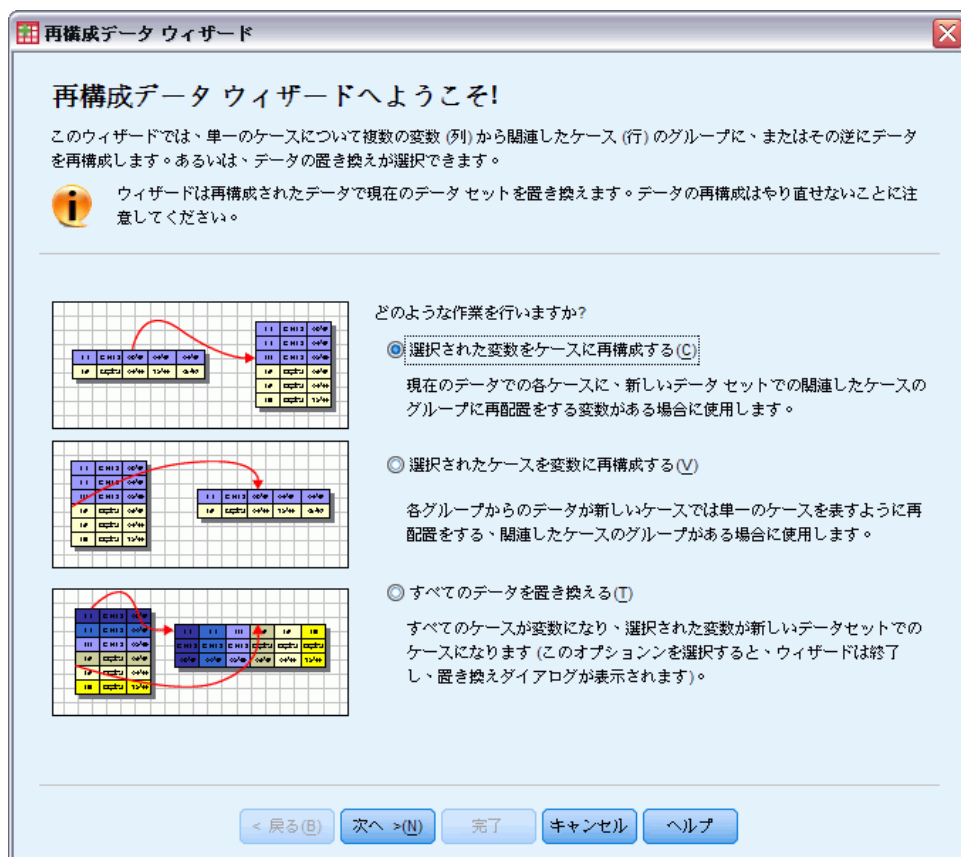
図 22-22  
[変数の計算] ダイアログ ボックス



- ▶ 目標変数として「開始時間3」と入力します。
- ▶ 数式として「時間3」と入力します。
- ▶ [OK] をクリックします。
- ▶ データを変数からケースに再構成するには、メニューから次の項目を選択します。  
データ > 再構成...

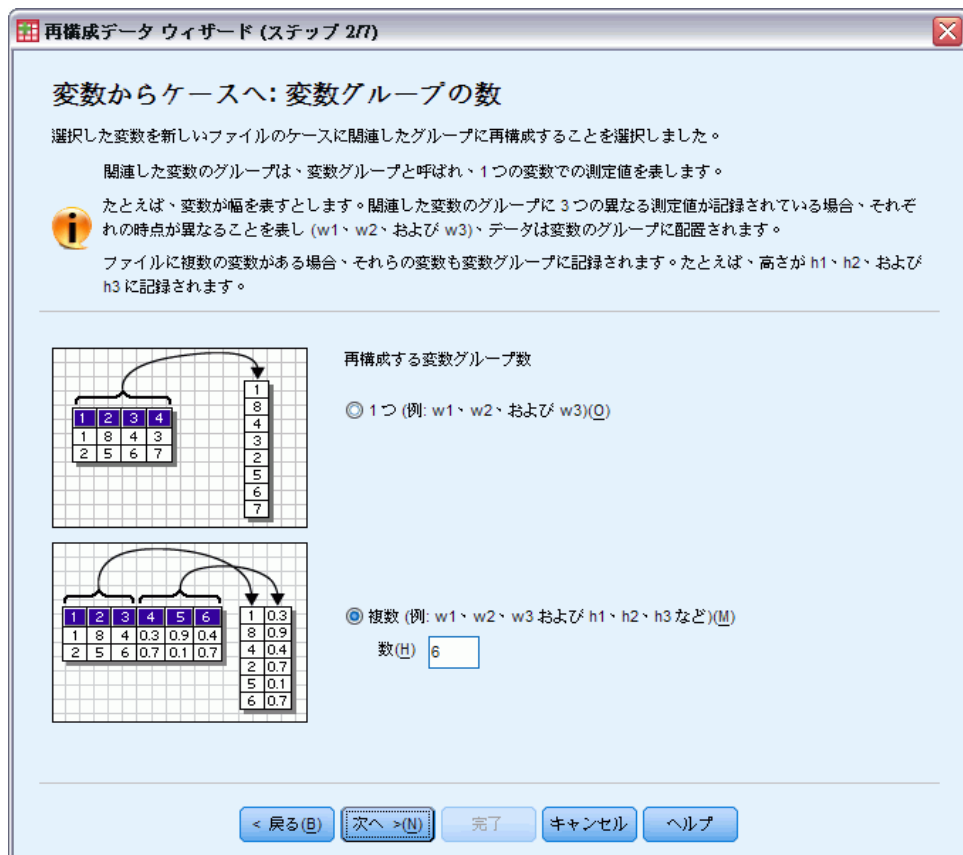


図 22-23  
再構成データ ウィザードの [よろこ] ステップ



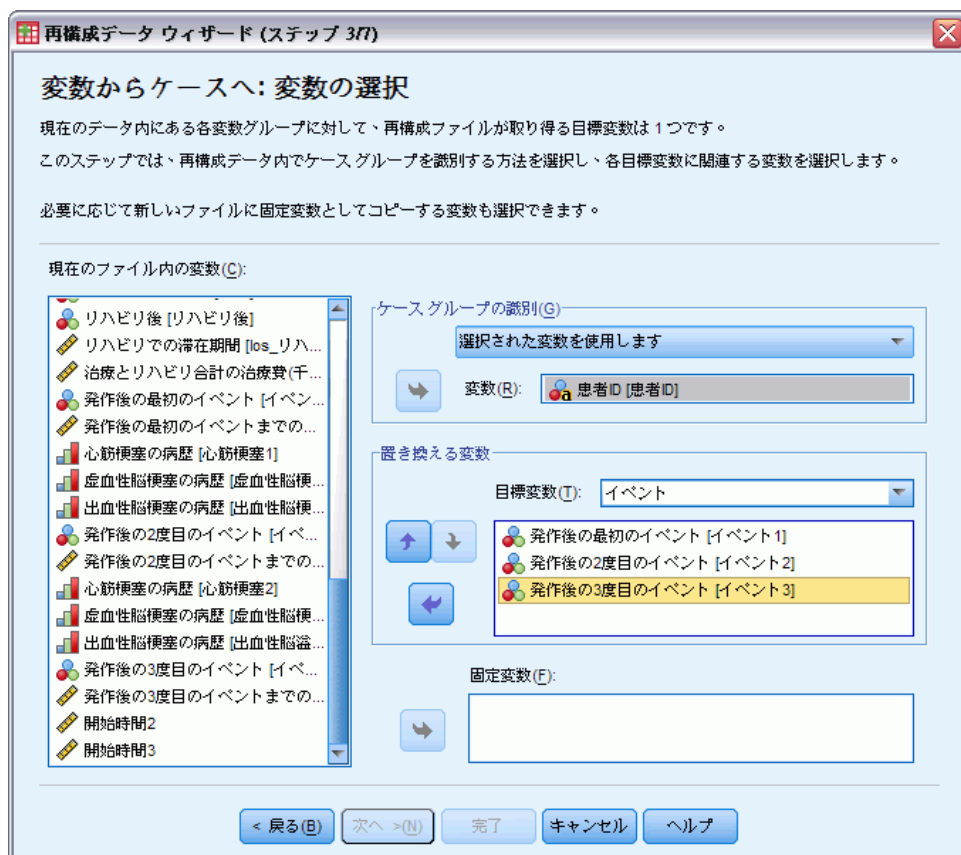
- ▶ [選択された変数をケースに再構成する] が選択されていることを確認します。
- ▶ [次へ] をクリックします。

図 22-24  
再構成データ ウィザードの [変数からケースへ: 変数グループの数] ステップ



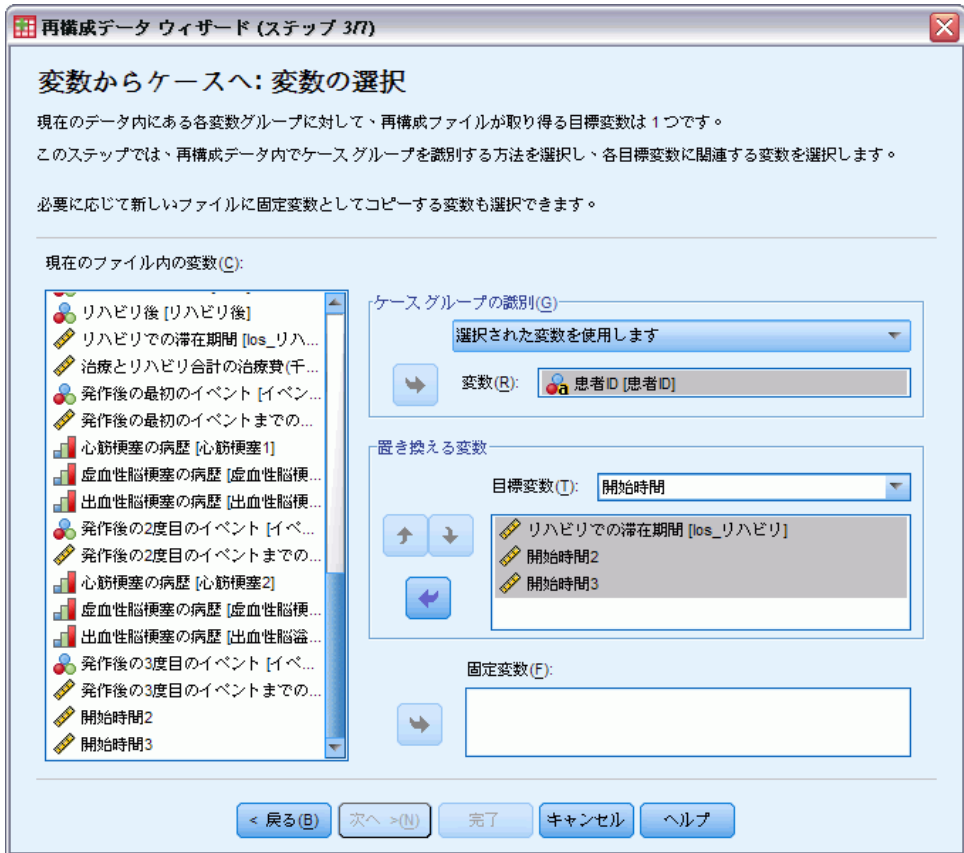
- ▶ 再構成する変数グループ数として [複数] を選択します。
- ▶ グループの数に「6」と入力します。
- ▶ [次へ] をクリックします。

図 22-25  
再構成データ ウィザードの [変数からケースへ: 変数の選択] ステップ



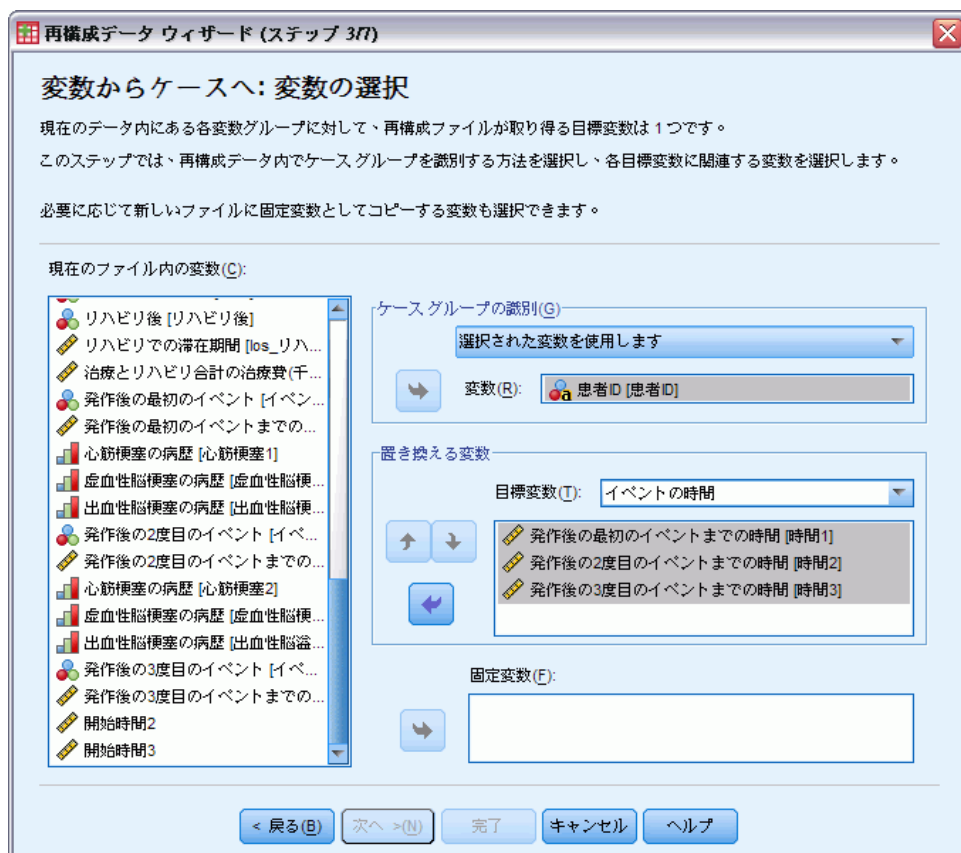
- ▶ [ケース グループの識別] グループで、[選択された変数を使用します] を選択し、被験者変数として [患者 ID [患者 ID]] を選択します。
- ▶ 1 番目の目標変数として「イベント」と入力します。
- ▶ 置き換える変数として、[発作後の最初のイベント [イベント 1]]、[発作後の 2 度目のイベント [イベント 2]]、[発作後の 3 度目のイベント [イベント 3]] を選択します。
- ▶ 目標変数リストから「trans2」を選択します。

図 22-26  
再構成データ ウィザードの [変数からケースへ: 変数の選択] ステップ



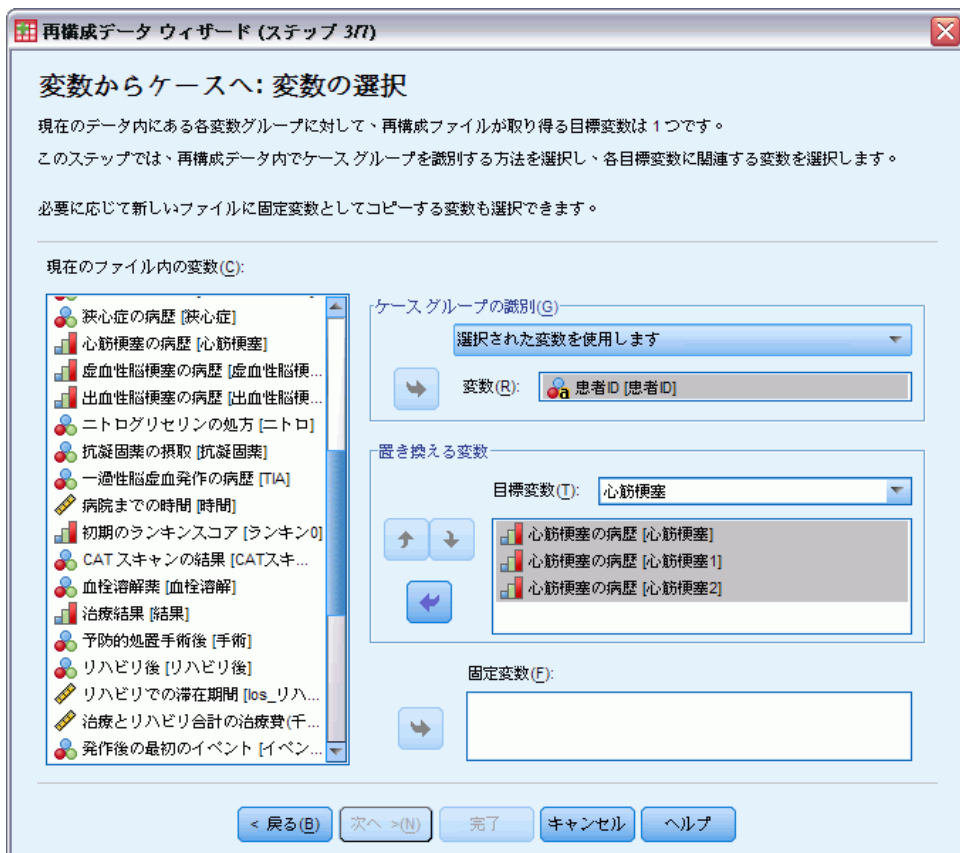
- ▶ 目標変数として「開始時間」と入力します。
- ▶ 置き換える変数として [リハビリでの滞在期間 [los\_リハビリ]]、「開始時間 2」、「開始時間 3」を選択します。[発作後の最初のイベントまでの時間 [時間 1] と [発作後の 2 度目のイベントまでの時間 [時間 2]] を使用して、終了時間を作成しますが、各変数は置き換える変数の 1 つのリストにしか表示されないため、「開始時間 2」と「開始時間 3」は必要ありませんでした。
- ▶ 目標変数リストから「trans3」を選択します。

図 22-27  
再構成データ ウィザードの [変数からケースへ: 変数の選択] ステップ



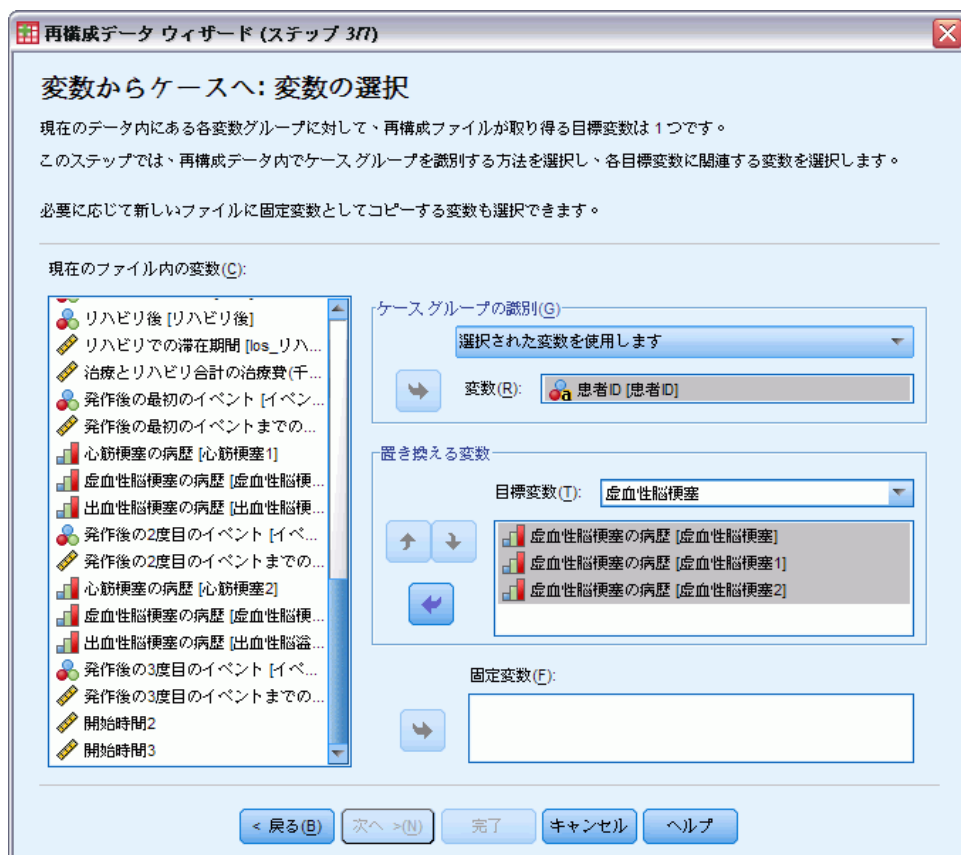
- ▶ 目標変数として「イベントまでの時間」と入力します。
- ▶ 置き換える変数として、[発作後の最初のイベントまでの時間 [時間 1]]、[発作後の 2 度目のイベントまでの時間 [時間 2]]、[発作後の 3 度目のイベントまでの時間 [時間 3]] を選択します。
- ▶ 目標変数リストから「trans4」を選択します。

図 22-28  
再構成データ ウィザードの [変数からケースへ: 変数の選択] ステップ



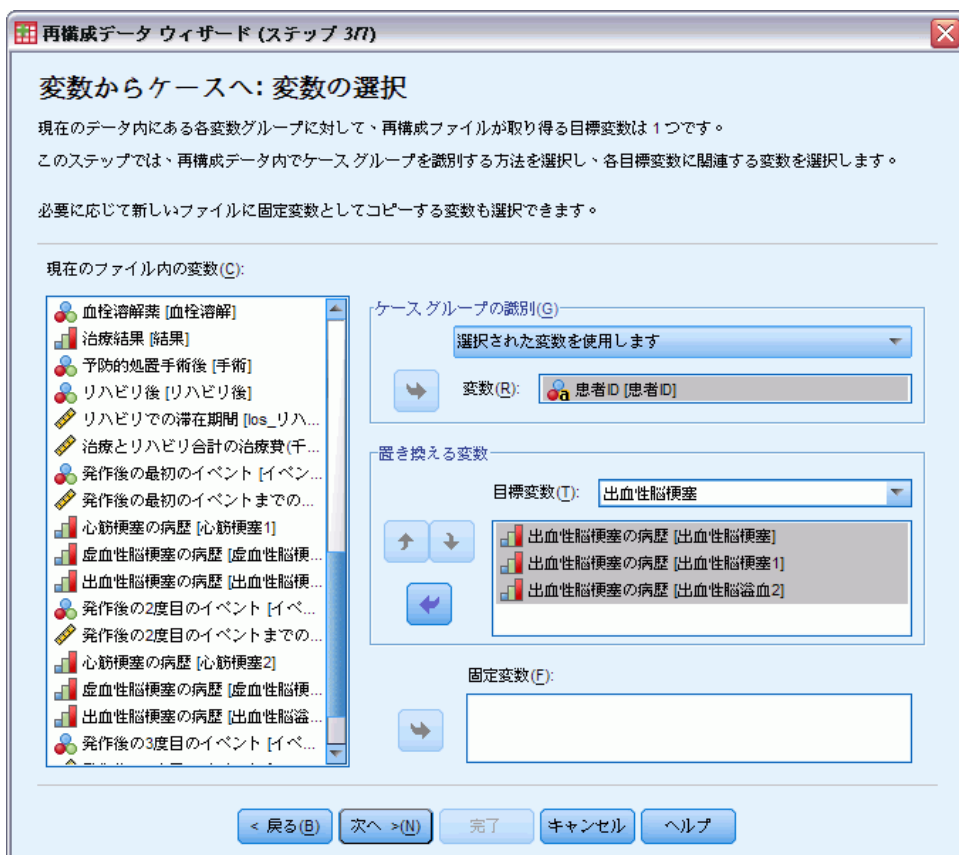
- ▶ 目標変数として「心筋梗塞」と入力します。
- ▶ 置き換える変数として、[心筋梗塞の病歴 [心筋梗塞]]、[心筋梗塞の病歴 [心筋梗塞 1]]、[心筋梗塞の病歴 [心筋梗塞 2]] を選択します。
- ▶ 目標変数リストから「trans5」を選択します。

図 22-29  
再構成データ ウィザードの [変数からケースへ: 変数の選択] ステップ



- ▶ 目標変数として「虚血性脳梗塞」と入力します。
- ▶ 置き換える変数として、[虚血性脳梗塞の病歴 [虚血性脳梗塞]]、[虚血性脳梗塞の病歴 [虚血性脳梗塞 1]]、[虚血性脳梗塞の病歴 [虚血性脳梗塞 2]] を選択します。
- ▶ 目標変数リストから「trans6」を選択します。

図 22-30  
再構成データ ウィザードの [変数からケースへ: 変数の選択] ステップ



- ▶ 目標変数として「出血性脳梗塞」と入力します。
- ▶ 置き換える変数として、[出血性脳梗塞の病歴 [出血性脳梗塞]]、[出血性脳梗塞の病歴 [出血性脳梗塞 1]]、[出血性脳梗塞の病歴 [出血性脳梗塞 2]] を選択します。
- ▶ [次へ] をクリックし、[インデックス変数の作成] ステップで [次へ] をクリックします。



図 22-31  
再構成データ ウィザードの [変数からケースへ: 1つのインデックス変数の作成] ステップ

再構成データ ウィザード (ステップ 5 of 7)

### 変数からケースへ: 1つのインデックス変数の作成

1つのインデックス変数を作成することを選択しました。変数値は、連続値かグループ内の変数名です。  
テーブル内でインデックス変数の名前とラベルを指定します。

インデックス値の種類

連続値(S)  
インデックス値(D): 1, 2, 3

変数名(A)  
インデックス値(D): イベント1, イベント2, イベント3

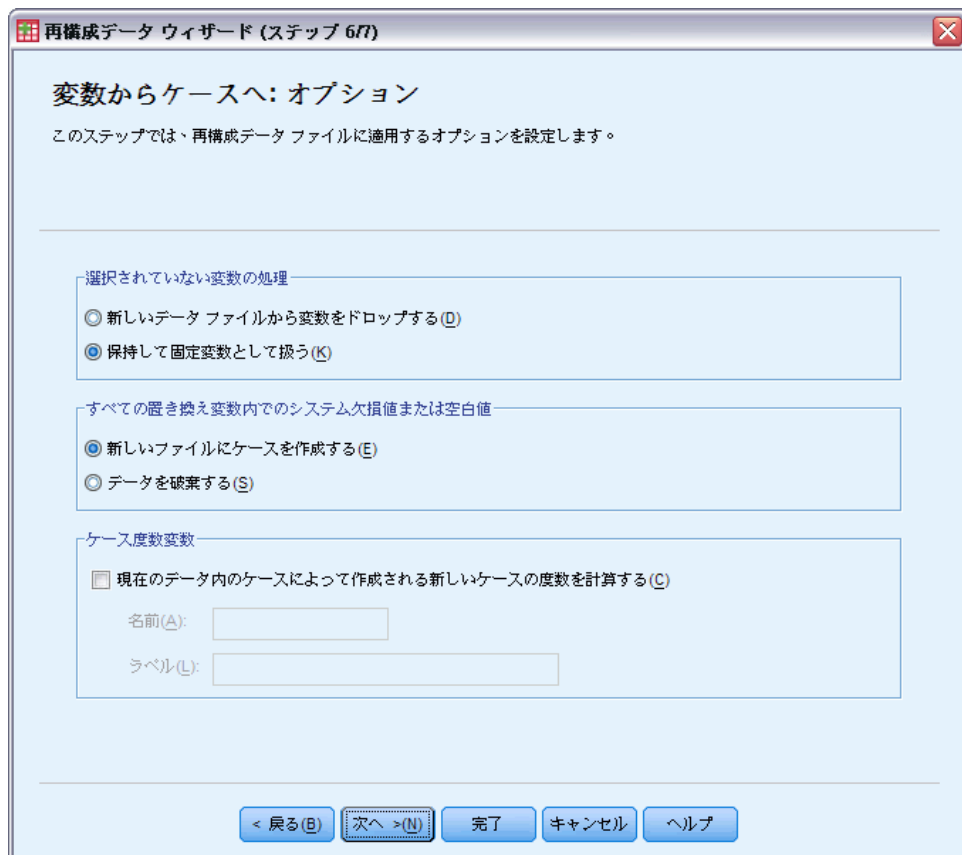
インデックス値とラベルを編集する(X):

	名前	ラベル	レベル	インデックス値
1	インデックス1		3	1, 2, 3

< 戻る(B) 次へ >(N) 完了 キャンセル ヘルプ

- ▶ インデックス変数の名前として「イベントインデックス」、変数ラベルとして「イベントインデックス」と入力します。
- ▶ [次へ] をクリックします。

図 22-32  
再構成データ ウィザードの [変数からケースへ: 1 つのインデックス変数の作成] ステップ



- ▶ [保持して固定変数として扱う] が選択されていることを確認します。
- ▶ [完了] をクリックします。

図 22-33  
再構成データ

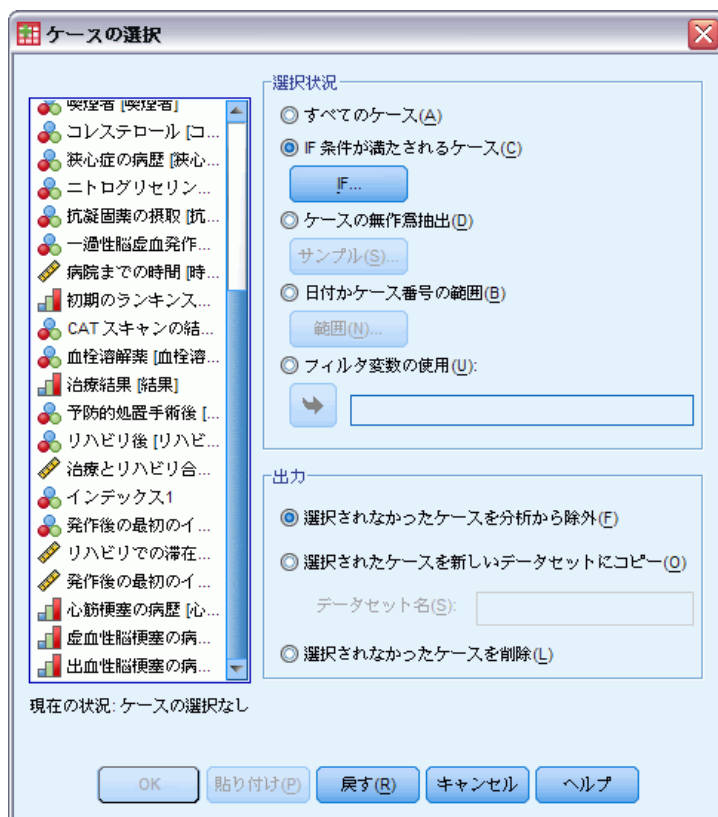
イベントイン デックス	イベント	開始時間	イベントまで の時間	心筋梗塞	虚血性脳梗塞	出血性脳梗塞
1	0	3	1500	0	4	0
2	-4	1500	-4	-4	-4	-4
3	-4	.	-4	-4	-4	-4
1	1	33	1311	0	11	0
2	4	1311	1325	1	1	0
3	-3	1325	-3	-3	-3	-3
1	4	12	1098	1	1	0
2	-3	1098	-3	-3	-3	-3
3	-3	.	-3	-3	-3	-3
1	4	4	1356	0	1	0
2	-3	1345	-3	-3	-3	-3
3	-3	.	-3	-3	-3	-3

再構成データには、患者ごとに3つのケースが含まれていますが、イベントを3回経験した患者は多くありません。したがって、「イベント」の値が負（欠損値）のケースが多数あります。これらは、フィルタにかけてデータセットから除外することができます。

- ▶ これらのケースをフィルタにかけるには、メニューから次の項目を選択します。

データ > ケースの選択...

図 22-34  
[ケースの選択] ダイアログ ボックス



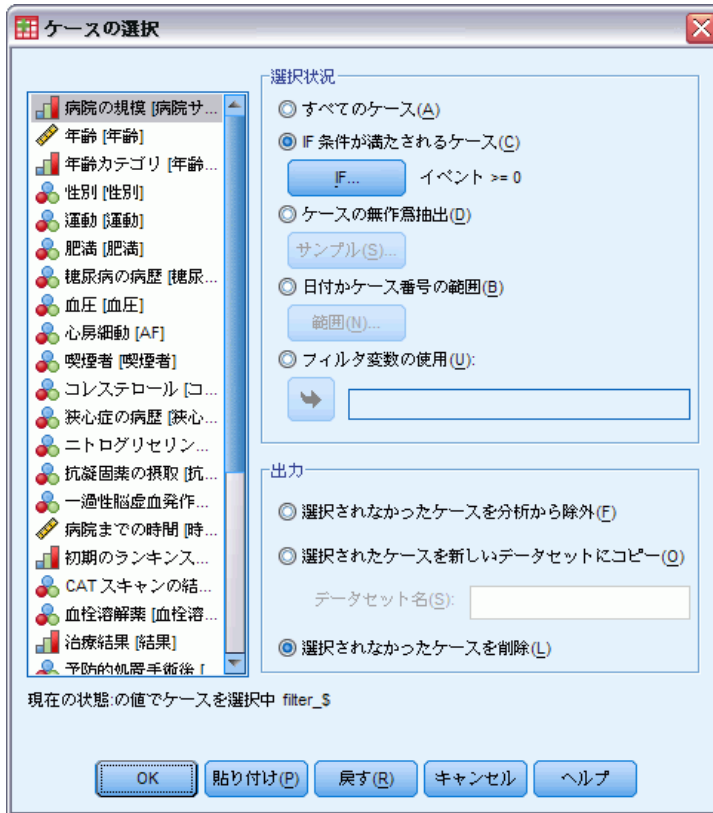
- ▶ [IF 条件が満たされるケース] を選択します。
- ▶ [IF] をクリックします。

図 22-35  
[ケースの選択: IF 条件の定義] ダイアログ ボックス



- ▶ 条件式として「イベント >= 0」と入力します。
- ▶ [続行] をクリックします。

図 22-36  
[ケースの選択] ダイアログ ボックス



- ▶ [選択されなかったケースを削除] を選択します。
- ▶ [OK] をクリックします。

## 単純な無作為抽出分析計画の作成

これで、単純な無作為抽出分析計画を作成する準備が整いました。

- ▶ まず、サンプルの重み付け変数を作成する必要があります。メニューから次の項目を選択します。  
変換 > 変数の計算...

図 22-37  
[Cox 回帰] ダイアログ ボックス



- ▶ 目標変数として「サンプル重み付け」と入力します。
- ▶ 数式として「1」と入力します。
- ▶ [OK] をクリックします。

これで、分析計画を作成する準備が整いました。

**注:** 以下の手順を省略し、データの分析に進みたい場合は、サンプル ファイルのディレクトリに既存の計画ファイル srs.csaplan があります。

- ▶ 分析計画を作成するには、メニューから次の項目を選択します。  
分析 > コンプレックス サンプル > 分析の準備...

図 22-38  
分析準備ウィザードの [ようこそ] ステップ



- ▶ [計画ファイルの作成] を選択し、ファイルの名前として「srs.csaplan」と入力します。または、参照して保存する場所を指定します。
- ▶ [次へ] をクリックします。

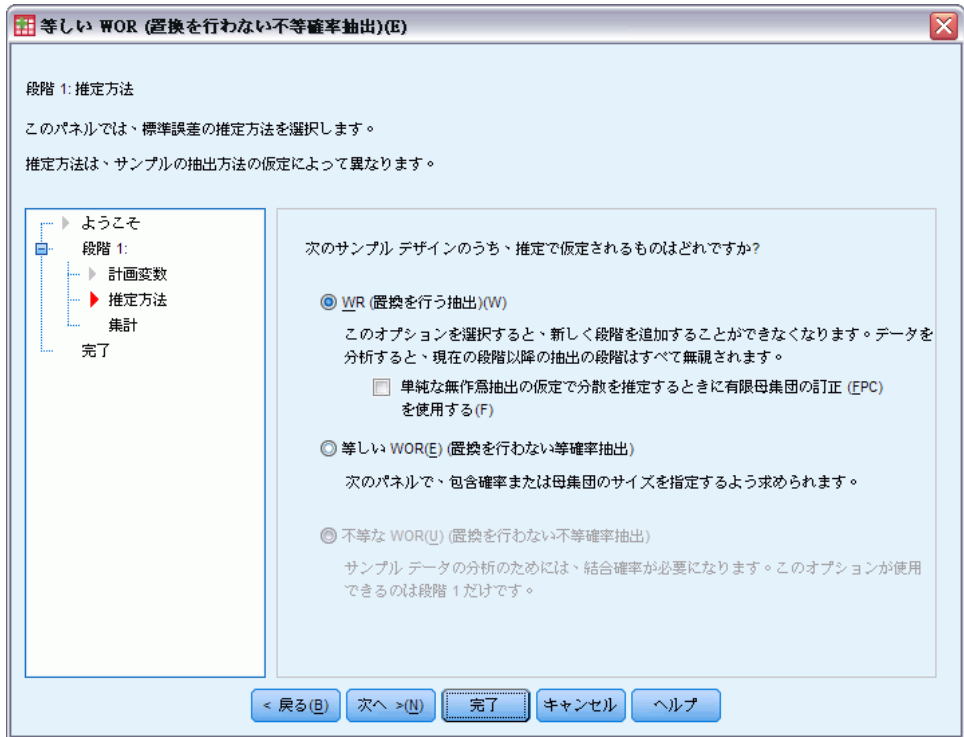


図 22-39  
分析準備ウィザードの [計画変数] ステップ



- ▶ 「サンプル重み付け」をサンプル重み付け変数として選択します。
- ▶ [次へ] をクリックします。

図 22-40  
分析準備ウィザードの [推定方法] ステップ



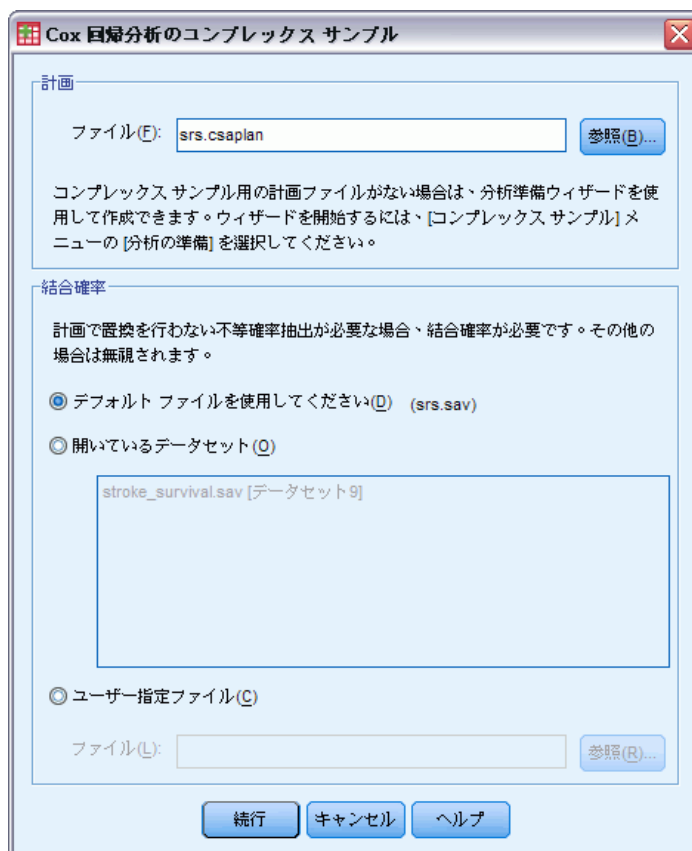
- ▶ [単純な無作為抽出の仮定で分散を推定するときに有限母集団の訂正 (FPC) を使用する] の選択を解除します。
- ▶ [完了] をクリックします。

これで、分析を実行する準備が整いました。

## 分析の実行

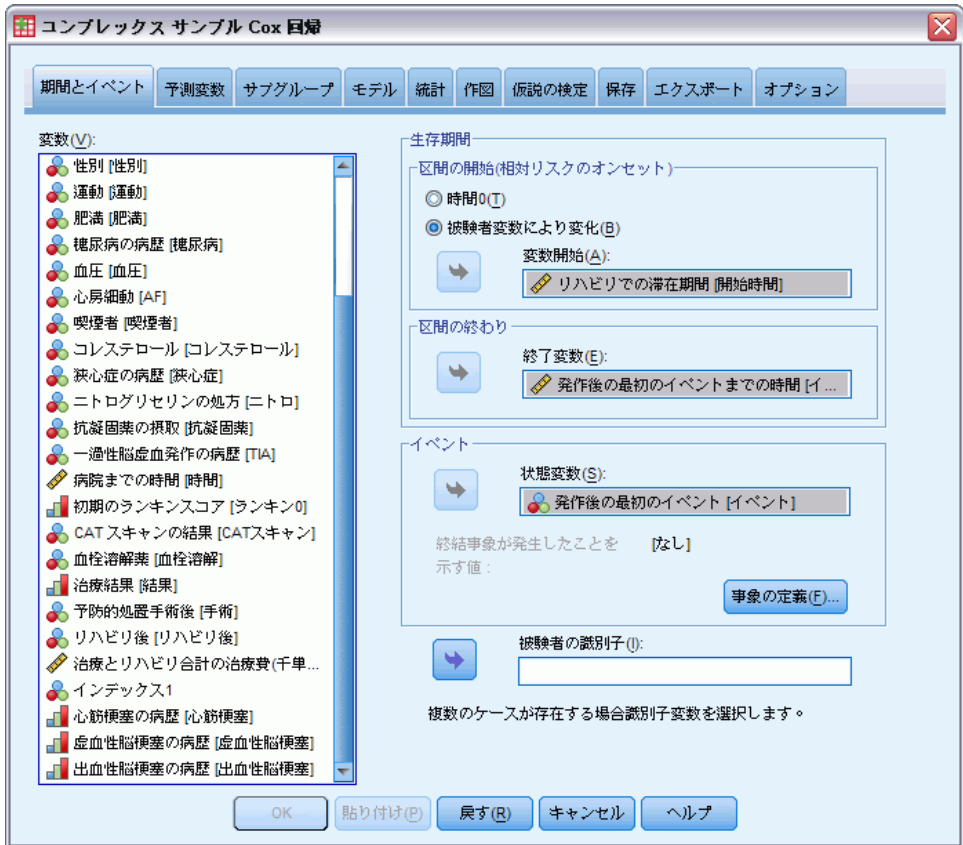
- ▶ コМПЛЕКС サンプルの Cox 回帰分析を実行するには、メニューから次の項目を選択します。  
分析 > コМПЛЕКС サンプル > Cox 回帰分析...

図 22-41  
[Cox 回帰分析のコンプレックス サンプル] ダイアログ ボックス



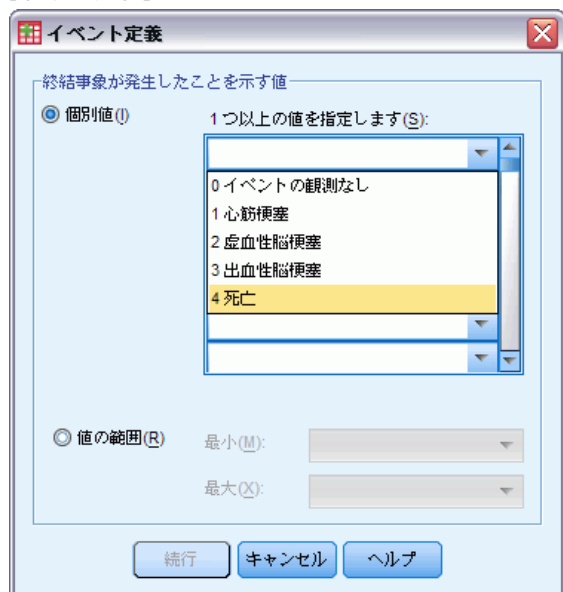
- ▶ 単純な無作為抽出計画プランを保存した場所、またはサンプル ファイルのディレクトリを参照し、srs.csaplan を選択します。
- ▶ [続行] をクリックします。

図 22-42  
[Cox 回帰分析] ダイアログ ボックスの [時刻およびイベント] タブ



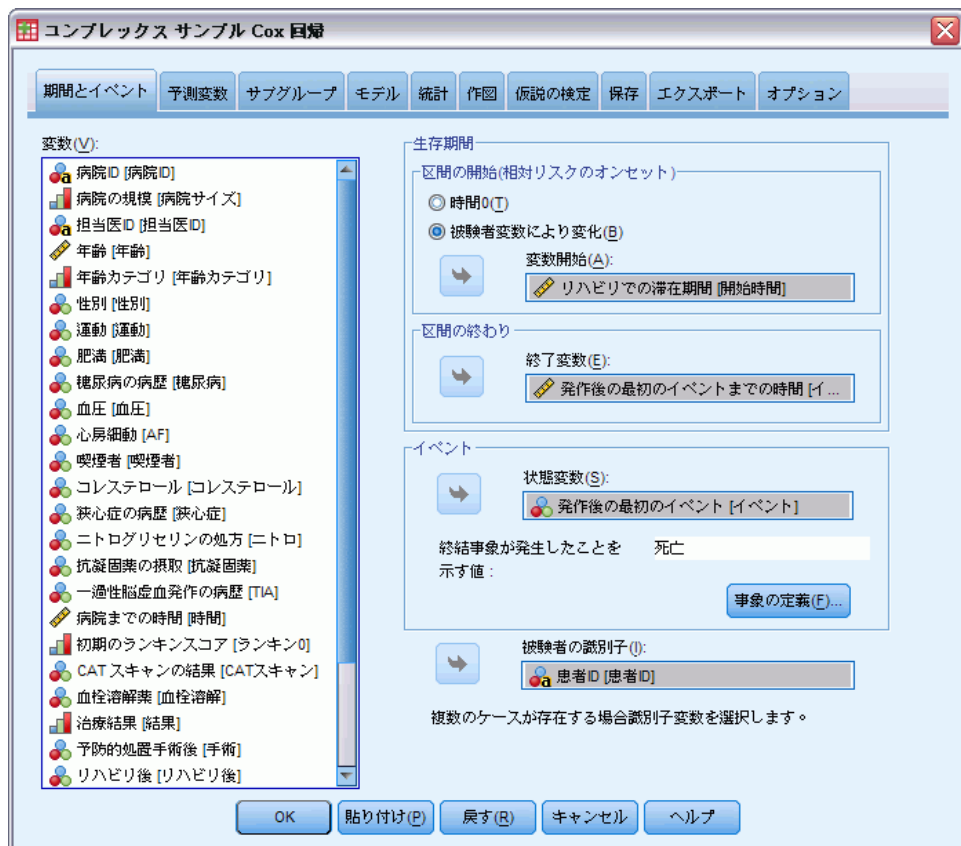
- ▶ [被験者数により変化] を選択し、変数開始として [リハビリでの滞在期間 [開始時間]] を選択します。再構成された変数では、構成で最初に使用された変数のラベルが使用されていることに注意してください。ただし、このラベルは構成された変数に必ずしも適しているとはかぎりません。
- ▶ 変数終了として、[発作後の最初のイベントまでの時間 [イベントまでの時間]] を選択します。
- ▶ 状態変数として、[発作後の最初のイベント [イベント]] を選択します。
- ▶ [事象の定義] をクリックします。

図 22-43  
[事象の定義] ダイアログ ボックス



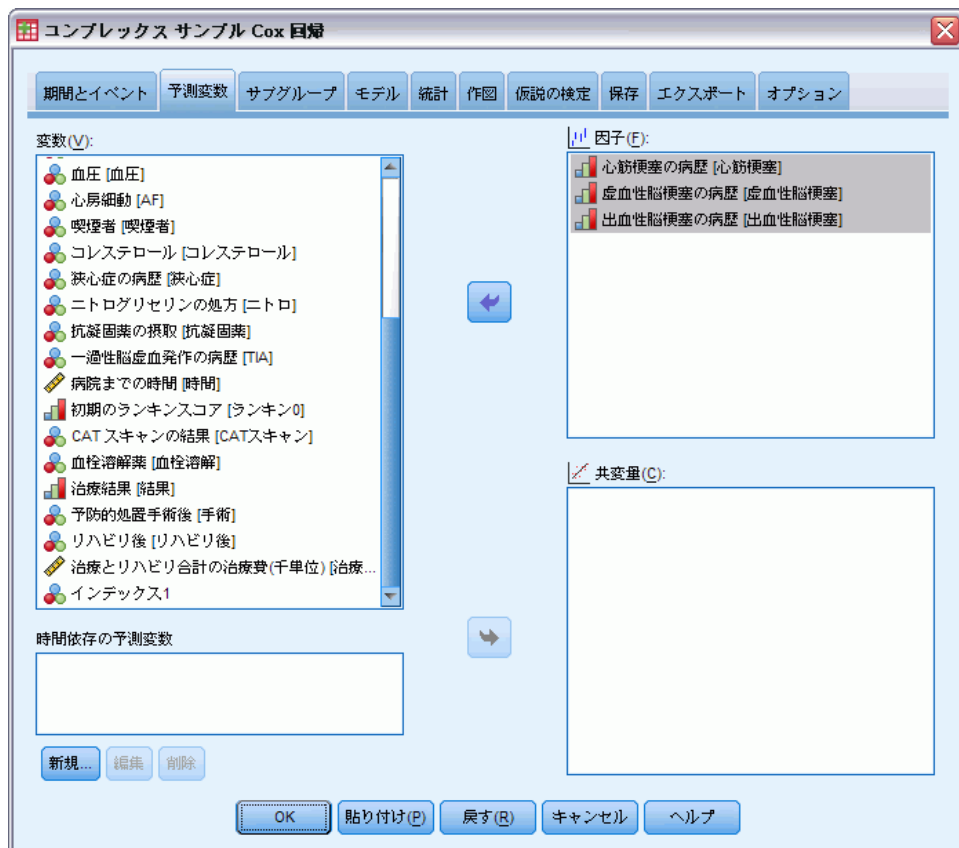
- ▶ 終結事象が発生したことを示す値として、[4 死亡] を選択します。
- ▶ [続行] をクリックします。

図 22-44  
[Cox 回帰分析] ダイアログ ボックスの [時刻およびイベント] タブ



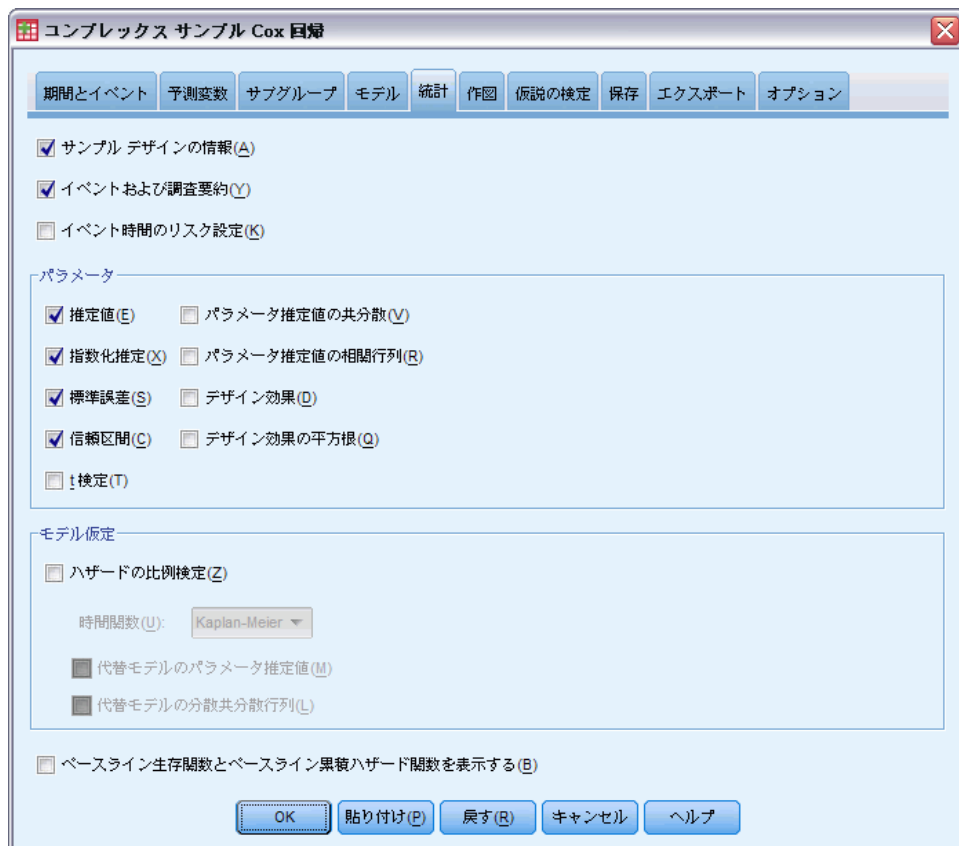
- ▶ 被験者の識別子として、[患者 ID [患者 ID]] を選択します。
- ▶ [予測変数] タブをクリックします。

図 22-45  
[Cox 回帰分析] ダイアログ ボックスの [予測変数] タブ



- ▶ 因子として、[心筋梗塞の病歴 [心筋梗塞]] から [出血性脳梗塞の病歴 [出血性脳梗塞]] までを選択します。
- ▶ [統計] タブをクリックします。

図 22-46  
[Cox 回帰分析] ダイアログ ボックスの [統計] タブ



- ▶ [パラメータ] グループで、[推定値]、[指定化推定]、[標準誤差]、[信頼区間] を選択します。
- ▶ [作図] タブをクリックします。



図 22-47  
[Cox 回帰分析] ダイアログ ボックスの [統計] タブ



- ▶ [ログマイナスログ生存関数] を選択します。
- ▶ 「心筋梗塞の病歴」の[線の定義変数]チェックボックスをオンにします。
- ▶ 「虚血性脳梗塞の病歴」のレベルとして、[1.0]を選択します。
- ▶ 「出血性脳梗塞の病歴」のレベルとして、[0.0]を選択します。
- ▶ [オプション] タブをクリックします。

図 22-48  
[Cox 回帰分析] ダイアログ ボックスの [オプション] タブ

- ▶ [推定] グループの均衡を破る方法として、[Breslow] を選択します。
- ▶ [OK] をクリックします。

## サンプル デザインの情報

図 22-49  
サンプル デザインの情報

			度数
重み付けしない度数	有効数	被験者	2421
		ケース	3310
		無効なケース	0
		ケースの合計	3310
有効数		被験者母集団のサイズ	2421.000
第 1 段階	有効数	ストラータ	1
		単位	2421
有効数		抽出計画の自由度	2420

この表には、モデルの推定に関するサンプル デザインの情報が含まれています。

- 一部の被験者には複数のケースがあり、分析では 3,310 件すべてのケースが使用されています。
- 計画には、単一のストラータと 2,421 単位（各被験者に 1 つ）があります。抽出計画の自由度は、 $2421-1=2420$  と推定されています。

## モデル効果の検定

図 22-50  
モデル効果の検定

ソース	df1	df2	Wald F	有意確率
心筋梗塞	3.000	2418.000	452.873	.000
虚血性脳梗塞	2.000	2419.000	1604.936	.000
出血性脳梗塞	2.000	2419.000	739.197	.000

生存時間変数: リハビリでの滞在期間, 発作後の最初のイベントまでの時間  
 事象状態変数: 発作後の最初のイベント = 4  
 被験者 ID 変数: 患者ID  
 モデル: 心筋梗塞, 虚血性脳梗塞, 出血性脳梗塞

各効果の有意確率は 0 に近くなり、すべてがモデルに影響することを示しています。

## パラメータ推定値

図 22-51  
パラメータ推定値

パラメータ	回帰係数	標準誤差	95% 信頼区間		Exp (B)	Exp (B) の 95% 信頼区間	
			下段	上段		下段	上段
[心筋梗塞=0]	-6.381	.283	-6.935	-5.827	.002	.001	.003
[心筋梗塞=1]	-5.589	.284	-6.147	-5.032	.004	.002	.007
[心筋梗塞=2]	-2.119	.344	-2.794	-1.445	.120	.061	.236
[心筋梗塞=3]	.000 <sup>a</sup>	.	.	.	1.000	.	.
[虚血性脳梗塞=1]	-6.421	.202	-6.817	-6.024	.002	.001	.002
[虚血性脳梗塞=2]	-2.803	.222	-3.239	-2.366	.061	.039	.094
[虚血性脳梗塞=3]	-.000 <sup>a</sup>	.	.	.	1.000	.	.
[出血性脳梗塞=0]	-6.148	.355	-6.844	-5.453	0	.001	.004
[出血性脳梗塞=1]	-2.232	.373	-2.963	-1.502	.107	.052	.223
[出血性脳梗塞=2]	.000 <sup>a</sup>	.	.	.	1.000	.	.

生存時間変数: リハビリでの滞在期間, 発作後の最初のイベントまでの時間  
 事象状態変数: 発作後の最初のイベント = 4  
 被験者 ID 変数: 患者ID  
 モデル: 心筋梗塞, 虚血性脳梗塞, 出血性脳梗塞

- このパラメータは余分であるため、0 に設定されます。
- タイプブレイク方式: Breslow

この手続きでは、参照カテゴリとして各因子の最後のカテゴリを使用します。その他のカテゴリの効果は、参照カテゴリと比較されます。推定値は統計上の検定では有効ですが、参照カテゴリと相対的なハザードの予測変化として、指数化推定 (Exp(B)) のほうが簡単に解釈できます。

- 「心筋梗塞=0」の Exp(B) の値は、心筋梗塞の病歴がない患者が死亡するというハザードは、心筋梗塞 3 回という病歴を持つ患者の 0.002 倍であることを意味します。
- 「心筋梗塞=1」と「心筋梗塞=0」の信頼区間が重複しています。これは、心筋梗塞 1 回という病歴を持つ患者のハザードは、心筋梗塞の病歴を持たない患者と統計的には区別できないことを示しています。
- 「心筋梗塞=0」と「心筋梗塞=1」の信頼区間は、「心筋梗塞=2」の区間とは重複せず、これらにはすべて 0 が含まれていません。したがって、心筋梗塞の病歴が 1 回以下の患者のハザードは、2 回の患者と区別することができ、同様に 2 回の患者のハザードは 3 回の患者と区別することができます。

「虚血性脳梗塞」と「出血性脳梗塞」のレベルについても、過去の発生回数が多くなると死亡のハザードが大きくなるという同様の関係が存在します。

## パターン値

図 22-52  
パターン値

		生存時間の間隔				
		開始	終了	心筋梗塞の病歴	虚血性脳梗塞の病歴	出血性脳梗塞の病歴
参照パターン	1	.000	a	3回	3回	2回
パターン 1.1	1	.000	a	なし	1回	なし
パターン 1.2	1	.000	a	1回	1回	なし
パターン 1.3	1	.000	a	2回	1回	なし
パターン 1.4	1	.000	a	3回	1回	なし

未指定の予測値は、参照パターンでこの予測値の値を割り当てられます。  
各生存時間の間隔は、開始 < 生存時間 ≤ 終了と定義されます。  
モデル: 心筋梗塞, 虚血性脳梗塞, 出血性脳梗塞。

a. 無制限

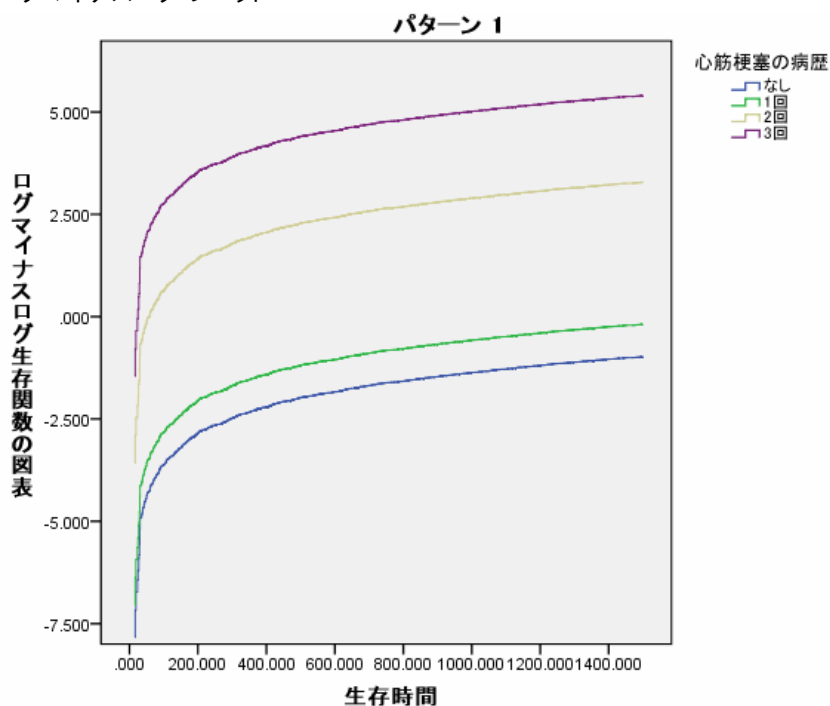
パターン値表には、各予測変数のパターンを定義する値が表示されます。モデルの予測変数のほかに、生存 j 時間間隔の開始時間と終了時間も表示されています。ダイアログから実行した分析の場合は、必ず開始時間と終了時間がそれぞれ 0 と無限になります。シンタックスを使用して、区的に一定な予測変数のパスを指定できます。

- 参照パターンは、各因子の参照カテゴリと各共変量の平均値で設定されます (このモデルには共変量がありません)。このデータセットでは、表示された参照モデルの因子の組み合わせは発生しないので、参照パターンのログマイナスログプロットは無視します。

- パターン 1.1 から 1.4 は、「心筋梗塞の病歴」の値のみが異なります。「心筋梗塞の病歴」の各値に対して、要求されたプロットで別のパターン（と別の線）が作成されます。

## ログマイナスログ プロット

図 22-53  
ログマイナスログ プロット



このプロットには、生存時間に対する生存関数のログマイナスログ  $\ln(-\ln(\text{生存}))$  が示されています。このプロットでは、「心筋梗塞の病歴」のカテゴリごとに曲線を表示し、「虚血性脳梗塞の病歴」を [1 回]、「出血性脳梗塞の病歴」を [なし] に固定して、生存関数への「心筋梗塞の病歴」の効果が有効に視覚化されています。パラメータ推測値テーブルにあるように、心筋梗塞の病歴が 1 回以下の患者の生存は、2 回の患者と区別することができ、同様に 2 回の患者は 3 回の患者と区別することができます。

## 集計（報告書 データ列）

脳卒中後の患者の病歴による効果を推定する、脳卒中後の生存の Cox 回帰モデルを当てはめました。他の予測変数をこのモデルに含めたいと研究者が考えるのは明らかなので、これはまだほんの手始めです。また、このデータセットをさらに分析するうえで、モデル構成に大きな変更を加えることも検討してみてください。たとえば、現在のモデルでは、患者の病歴を変える事象の効果は、倍数によってベースライン ハザードに定量化できると仮定しています。かわりに、ベースライン ハザードの形は、死亡以外の事象の発生によって変化すると仮定することもできるでしょう。これを行うには、「イベントインデックス」に基づいて分析を階層化することができます。

# サンプル ファイル

製品とともにインストールされるサンプル ファイルは、インストール ディレクトリの Samples サブディレクトリにあります。[サンプル] サブディレクトリ内に次の各言語の別のフォルダがあります。英語、フランス語、ドイツ語、イタリア語、日本語、韓国語、ポーランド語、ロシア語、簡体字中国語、スペイン語、そして繁体中国語です。

すべてのサンプル ファイルが、すべての言語で使用できるわけではありません。サンプル ファイルがある言語で使用できない場合、その言語のフォルダには、サンプル ファイルの英語バージョンが含まれています。

## 説明

以下は、このドキュメントのさまざまな例で使用されているサンプル ファイルの簡単な説明です。

- **accidents.sav**。与えられた地域での自動車事故の危険因子を年齢および性別ごとに調べている保険会社に関する架空のデータ ファイルです。各ケースが、年齢カテゴリと性別のクロス分類に対応します。
- **adl.sav**。脳卒中患者に提案される治療の効果を特定するための取り組みに関する架空のデータ ファイルです。医師団は、女性の脳卒中患者たちを、2つのグループのいずれかにランダムに割り当てました。一方のグループは標準的な理学療法を受け、もう一方のグループは感情面の治療も追加で受けました。治療の3か月後に、各患者が日常生活の一般的な行動をどの程度とることができるかを、順序変数として得点付けしました。
- **advert.sav**。広告費とその売上成果の関係を調べるための小売業者の取り組みに関する架空のデータ ファイルです。この小売業者は、そのために、過去の売上と、それに関する広告費のデータを収集しました。
- **aflatoxin.sav**。収穫物によって濃度が大きく異なる毒物であるアフラトキシンを、トウモロコシの収穫物に関して検定することに関する架空のデータ ファイルです。ある穀物加工業者は、8つそれぞれの収穫物から16のサンプルを受け取って、10億分の1単位でアフラトキシン レベルを測定しました。

- **aflatoxin20.sav.** このデータ ファイルには、aflatoxin.sav データ ファイルの収穫物 4 および 8 の 16 個のサンプルに関して、アフラトキシンの測定結果が含まれています。
- **anorectic.sav.** 拒食行動または過食行動の標準的な症状の特定を目指して、調査員 (Van der Ham, Meulman, Van Strien, および Van Engeland, 1997) が、摂食障害を持つ大人 55 人の調査を行いました。各患者が 4 年間で 4 回診察を受けたので、観測値は合計で 220 になりました。観測値ごとに、16 種類の症状に関して患者の得点が記録されました。患者 71 (2 回目)、患者 76 (2 回目)、患者 47 (3 回目) の症状の得点が見つからなかったため、残っている 217 回分の観測値が有効です。
- **autoaccidents.sav.** 運転手の年齢や性別も考慮したうえで運転手ごとの自動車事故の数をモデリングするための、ある保険アナリストの取り組みに関する架空のデータ ファイルです。各ケースが別々の運転手を表し、運転手の性別、年齢、最近 5 年間の自動車事故の回数が記録されています。
- **band.sav.** このデータ ファイルには、あるバンドの音楽 CD の架空の週間売上数が含まれています。3 つの予測変数のデータも含まれています。
- **bankloan.sav.** 債務不履行率を低減させるための銀行の取り組みに関する架空のデータ ファイルです。このファイルには、過去の顧客および見込み客 850 人に関する財務情報と人口統計情報が含まれています。最初の 700 ケースは、以前に貸付を行った顧客です。残りの 150 ケースは見込み顧客で、これらの顧客に関して銀行は信用リスクの良し悪しを分類する必要があります。
- **bankloan\_binning.sav.** 過去の顧客 5,000 人に関する財務情報と人口統計情報を含む架空のデータ ファイルです。
- **behavior.sav.** 52 人の学生に 15 の状況と 15 の行動の組み合わせについて、0 = 「非常に適切」から 9 = 「非常に不適切」までの 10 段階でランク付けするよう依頼した研究があります (Price および Bouffard, 1974)。個人間の平均を取ったため、値は非類似度としてみなされます。
- **behavior\_ini.sav.** このデータ ファイルには、behavior.sav の 2 次元の解の初期配置が含まれています。
- **brakes.sav.** 高性能自動車のディスク ブレーキを生産している工場での品質管理に関する架空のデータ ファイルです。このデータ ファイルには、8 台の機械で生産した 16 個のディスクの直径測定値が含まれています。ブレーキの目標の直径は 322 ミリメートルです。
- **breakfast.sav.** 21 人の Wharton School MBA の学生およびその配偶者に、15 種類の朝食を好みの順に (1 = 「最も好き」から 15 = 「最も嫌い」まで) ランク付けするよう依頼した研究があります (Green および Rao, 1972)。調査対象者の嗜好は、「すべて」から「スナックとドリンクのみ」まで、6 つの異なるシナリオに基づいて記録されました。
- **breakfast-overall.sav.** このデータ ファイルには、最初のシナリオ (「すべて」) のみの朝食の好みが含まれています。



- **broadband\_1.sav**。全国規模のブロードバンド サービスの地域ごとの契約者数を含む架空のデータ ファイルです。このデータ ファイルには、85 地域の月々の契約者数が 4 年間分含まれています。
- **broadband\_2.sav**。このデータ ファイルは `broadband_1.sav` と同じですが、データが 3 か月分追加されています。
- **car\_insurance\_claims.sav**。他の場所 (McCullagh および Nelder, 1989) で表示および分析される、自動車の損害請求に関するデータセットです。逆リンク関数を使用して従属変数の平均値を保険契約者の年齢、車種、製造年の線型結合と関連付けることにより、平均請求数はガンマ分布としてモデリングできます。申請された請求の数は、尺度重み付けとして使用できます。
- **car\_sales.sav**。このデータ ファイルには、自動車のさまざまな車種やモデルの架空の売上推定値、定価、仕様が含まれています。定価と仕様はそれぞれ、`edmunds.com` と製造元のサイトから入手しました。
- **car\_sales\_uprepared.sav**。変換したバージョンのフィールドを含まない `car_sales.sav` の修正したバージョンです。
- **carpet.sav**。一般的な例 (Green および Wind, 1973) としては、新しいカーペット専用洗剤を市販することに関心のある企業が消費者の嗜好に関する 5 種類の因子 (パッケージのデザイン、ブランド名、価格、サービスシール、料金の払い戻し) の影響について調べたい場合があります。パッケージのデザインには、3 つの因子レベルがあります。それぞれ塗布用ブラシの位置が異なります。また、3 つのブランド名 (K2R、Glory、および Bissell)、3 つの価格水準があり、最後の 2 つの因子のそれぞれに対しては 2 つのレベル (「なし」または「あり」) があります。10 人の消費者が、これらの因子により定義された 22 個のプロファイルに順位を付けます。変数「嗜好」には、各プロファイルの平均順位の序列が含まれています。順位が低いほど、嗜好度は高くなります。この変数には、各プロファイルの嗜好測定値がすべて反映されます。
- **carpet\_prefs.sav**。このデータ ファイルは `carpet.sav` と同じ例に基づいていますが、10 人の消費者それぞれから収集した実際のランキングが含まれています。消費者は、22 種類の製品プロファイルを、一番好きなものから一番嫌いなものまで順位付けすることを依頼されています。変数 `PREF1` から `PREF22` には、`carpet_plan.sav` で定義されている、関連するプロファイルの ID が含まれています。
- **catalog.sav**。このデータ ファイルには、あるカタログ会社が販売した 3 つの製品の、架空の月間売上高が含まれています。5 つの予測変数のデータも含まれています。
- **catalog\_seasfac.sav**。このデータ ファイルは `catalog.sav` と同じですが、季節性の分解手続きとそれに付随する日付変数から計算した一連の季節因子が追加されています。

- **cellular.sav**。解約率を削減するための携帯電話会社の取り組みに関する架空のデータ ファイルです。解約の傾向スコアは、0 ~ 100 の範囲でアカウントに適用されます。スコアリングが 50 以上のアカウントはプロバイダの変更を考えている場合があります。
- **ceramics.sav**。新しい上質の合金に標準的な合金より高い耐熱性があるかどうかを特定するための、ある製造業者の取り組みに関する架空のデータ ファイルです。各ケースが 1 つの合金の別々のテストを表し、軸受けの耐熱温度が記録されます。
- **cereal.sav**。880 人を対象に、朝食の好みについて、年齢、性別、婚姻状況、ライフスタイルが活動的かどうか（週 2 回以上運動するか）を含めて調査した、架空のデータ ファイルです。各ケースが別々の回答者を表します。
- **clothing\_defects.sav**。ある衣料品工場での品質管理工程に関する架空のデータ ファイルです。工場で生産される各ロットから、調査員が衣料品のサンプルを取り出し、不良品の数を数えます。
- **coffee.sav**。このデータ ファイルは、6 つのアイスコーヒー ブランド（Kennedy, Riquier, および Sharp, 1996）について受けた印象に関連しています。回答者は、アイス コーヒーに対する 23 の各印象属性に対して、その属性が言い表していると思われるすべてのブランドを選択しました。機密保持のため、6 つのブランドを AA、BB、CC、DD、EE、および FF で表しています。
- **contacts.sav**。企業のコンピュータ営業グループの担当者リストに関する架空のデータ ファイルです。各担当者は、所属する会社の部門および会社のランクによって分類されています。また、最新の販売金額、最後の販売以降の経過時間、担当者の会社の規模も記録されています。
- **creditpromo.sav**。最近のクレジット カード プロモーションの有効性を評価するための、あるデパートの取り組みに関する架空のデータ ファイルです。このために、500 人のカード所有者がランダムに選択されました。そのうち半分には、今後 3 か月間の買い物に関して利率を下げることをプロモーションする広告を送付しました。残り半分には、通常どおりの定期的な広告を送付しました。
- **customer\_dbase.sav**。自社のデータ ウェアハウスにある情報を使用して、反応がありそうな顧客に対して特典を提供するための、ある会社の取り組みに関する架空のデータ ファイルです。顧客ベースのサブセットをランダムに選択して特典を提供し、顧客の反応が記録されています。
- **customer\_information.sav**。名前や住所など、顧客の連絡先情報を含む架空のデータ ファイルです。
- **customer\_subset.sav**。customer\_dbase.sav の 80 件のケースのサブセット。
- **customers\_model.sav**。このファイルには、あるマーケティング キャンペーンの対象になった個人に関する架空のデータが含まれています。人口統計情報、購入暦の概要、各個人がキャンペーンに反応したかどうか、などのデータが含まれています。各ケースが別々の個人を表します。

- **customers\_new.sav**。このファイルには、あるマーケティング キャンペーンの候補である個人に関する架空のデータが含まれています。各個人の人口統計情報および購入暦の概要のデータが含まれています。各ケースが別々の個人を表します。
- **debate.sav**。政治討論の出席者に対して行った調査の、討論の前後それぞれの回答に関する架空のデータ ファイルです。各ケースが別々の回答者に対応します。
- **debate\_aggregate.sav**。debate. sav 内の回答を集計する、架空のデータ ファイルです。各ケースが、討論前後の好みのクロス分類に対応しています。
- **demo.sav**。月々の特典を送付することを目的とした、購入顧客のデータベースに関する架空のデータ ファイルです。顧客が特典に反応したかどうか、さまざまな人口統計情報と共に記録されています。
- **demo\_cs\_1.sav**。調査情報のデータベースをコンパイルするための、ある会社の取り組みの最初のステップに関する架空のデータ ファイルです。各ケースが別々の都市に対応し、地域、地方、地区、および都市の ID が記録されています。
- **demo\_cs\_2.sav**。調査情報のデータベースをコンパイルするための、ある会社の取り組みの第 2 のステップに関する架空のデータ ファイルです。各ケースが、最初のステップで選択した都市の別々の世帯単位に対応し、地域、地方、地区、都市、区画、および単位の ID が記録されます。計画の最初の 2 つの段階からの抽出情報も含まれています。
- **demo\_cs.sav**。コンプレックス サンプル計画を使用して収集された調査情報を含む架空のデータ ファイルです。各ケースが別々の世帯単位に対応し、さまざまな人口統計情報および抽出情報が記録されています。
- **dmdata.sav**。これは、人口統計およびダイレクト マーケティングの購入情報のデータを含む架空のデータ ファイルです。dmdata2. sav には、テスト メールを受け取った連絡先のサブセットに関する情報が含まれ、dmdata3. sav には、テスト メールを受け取っていない残りの連絡先に関する情報が含まれています。
- **dietstudy.sav**。この架空のデータ ファイルには、” Stillman diet” (Rickman, Mitchell, Dingman, および Dalen, 1974) の研究結果が含まれています。各ケースが別々の被験者に対応し、被験者のダイエット前後の体重 (ポンド単位) と、トリグルセリド レベル (mg/100 ml 単位) が記録されています。
- **dvdplayer.sav**。新しい DVD プレーヤーの開発に関する架空のデータ ファイルです。プロトタイプを使用して、マーケティング チームはフォーカス グループ データを収集しました。各ケースが別々の調査対象ユーザーに対応し、ユーザーの人口統計情報と、プロトタイプに関する質問への回答が記録されています。

- **german\_credit.sav**。このデータ ファイルは、カリフォルニア大学アーバイン校の Repository of Machine Learning Databases (Blake および Merz, 1998) にある “German credit” データセットから取ったものです。
- **grocery\_1month.sav**。この架空のデータ ファイルは、grocery\_coupons.sav データ ファイルの週ごとの購入を「ロールアップ」して、各ケースが別々の顧客に対応するようにしたものです。その結果、週ごとに変わっていた変数の一部が表示されなくなり、買物の総額が、調査を行った 4 週間の買物額の合計になっています。
- **grocery\_coupons.sav**。顧客の購買習慣に関心を持っている食料雑貨店チェーンが収集した調査データを含む架空のデータ ファイルです。各顧客を 4 週間に渡って追跡し、各ケースが別々の顧客の週に対応しています。その週に食料品に費やした金額も含め、顧客がいつどこで買物をするかに関する情報が記録されています。
- **guttman.sav**。Bell (Bell, 1961) は、予想される社会グループを示す表を作成しました。Guttman (Guttman, 1968) は、この表の一部を使用しました。この表では、社会交相互作用、グループへの帰属感、メンバとの物理的な近接性、関係の形式化などを表す 5 個の変数が、理論上の 7 つの社会グループと交差しています。このグループには、観衆 (例、フットボールの試合の観戦者)、視聴者 (例、映画館または授業の参加者)、公衆 (例、新聞やテレビの視聴者)、暴徒 (観衆に似ているが、より強い交相互作用がある)、第一次集団 (親密な関係)、第二次集団 (自発的な集団)、および近代コミュニティ (物理的により密接した近接性と特化されたサービスの必要性によるゆるい同盟関係) があります。
- **health\_funding.sav**。医療用資金 (人口 100 人あたりの金額)、罹患率 (人口 10,000 人あたりの人数)、医療サービス機関への訪問率 (人口 10,000 人あたりの人数) のデータを含む、架空のデータ ファイルです。各ケースが別々の都市を表します。
- **hivassay.sav**。HIV 感染を発見する迅速な分析方法を開発するための、ある製薬研究所の取り組みに関する架空のデータ ファイルです。分析の結果は、8 段階の濃さの赤で表現され、色が濃いほど感染の可能性が高くなります。研究所では 2,000 件の血液サンプルに関して試験を行い、その半数が HIV に感染しており、半分は感染していませんでした。
- **hourlywagedata.sav**。管理職から現場担当まで、またさまざまな経験レベルの看護師の時給に関する架空のデータ ファイルです。
- **insurance\_claims.sav**。不正請求の恐れがある、疑いを区別するためにモデルを作成する必要がある保険会社の仮説データ ファイルです。各ケースがそれぞれの請求を表します。
- **insure.sav**。10 年満期の生命保険契約に対し、顧客が請求を行うかどうかを示す危険因子を調査している保険会社に関する架空のデータ ファイルです。データ ファイルの各ケースは、年齢と性別が一致する、請求を行った契約と行わなかった契約のペアを表します。

- **judges.sav.** 訓練を受けた審判（および 1 人のファン）が 300 件の体操の演技に対して付けた得点に関する架空のデータ ファイルです。各行が別々の演技を表し、審判たちは同じ演技を見ました。
- **kinship\_dat.sav.** Rosenberg と Kim (Rosenberg および Kim, 1975) は、15 種類の親族関係用語（祖父、祖母、父、母、叔父、叔母、兄弟、姉妹、いとこ、息子、娘、甥、姪、孫息子、孫娘）の分析を行いました。Rosenberg と Kim は、大学生の 4 つのグループ（女性 2 組、男性 2 組）に、類似性に基づいて上記の用語を並べ替えるよう依頼しました。2 つのグループ（女性 1 組、男性 1 組）には、1 回目と違う条件に基づいて、2 回目の並べ替えをするように頼みました。このようにして、合計で 6 つの「ソース」が取得できました。各ソースは、 $15 \times 15$  の近接行列に対応します。この近接行列のセルの数は、ソースの人数から、ソース内でオブジェクトを分割した回数を引いたものです。
- **kinship\_ini.sav.** このデータ ファイルには、kinship\_dat.sav の 3 次元の解の初期配置が含まれています。
- **kinship\_var.sav.** このデータ ファイルには、kinship\_dat.sav の解の次元の解釈に使用できる独立変数である性別、世代、および (ation), and 親等が含まれています。特に、解の空間をこれらの変数の線型結合に制限するために使用できます。
- **marketvalues.sav.** 1999 ~ 2000 年間の、イリノイ州アルゴンキンの新興住宅地での住宅売上に関するデータ ファイルです。個の売り上げは、公的レコードの一種です。
- **nhis2000\_subset.sav.** National Health Interview Survey (NHIS) は、米国民を対象とした人口ベースの大規模な調査です。全国の代表的な世帯サンプルについて対面式で調査が行われます。各世帯のメンバーに関して、人口統計情報、健康に関する行動および状態の観測値が得られます。このデータ ファイルには、2000 年の調査から得られた情報のサブセットが含まれています。National Center for Health Statistics. National Health Interview Survey, 2000。一般使用データおよびドキュメント。[ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Datasets/NHIS/2000/](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/)。2003 年にアクセス。
- **ozone.sav.** データには、残りの変数からオゾン濃度を予測するための、6 個の気象変数に対する 330 個の観測値が含まれています。それまでの研究者 (Breiman および Friedman (F), 1985)、(Hastie および Tibshirani, 1990) が、他の研究者と共に、これらの変数間に非線型性を確認しています。この場合、標準的な回帰アプローチは使用できません。
- **pain\_medication.sav.** この架空のデータ ファイルには、慢性関節炎を治療する抗炎症薬の臨床試験の結果が含まれています。特に興味深いことは、薬の効果が出るまでの時間と、既存の薬剤との比較です。

- **patient\_los.sav**。この架空のデータ ファイルには、心筋梗塞 (MI、または「心臓発作」) の疑いで入院した患者の治療記録が含まれています。各ケースが別々の患者に対応し、入院に関連する多くの変数が記録されています。
- **patlos\_sample.sav**。この架空のデータ ファイルには、心筋梗塞 (MI、または「心臓発作」) の治療中に血栓溶解剤を投薬された患者のサンプルの治療記録が含まれています。各ケースが別々の患者に対応し、入院に関連する多くの変数が記録されています。
- **polishing.sav**。これは、Data and Story Library の “Nambeware Polishing Times” データ ファイルです。生産スケジュールを立てるための、金属食器製造業者 (ニューメキシコ州サンタフェの Nambe Mills) の取り組みに関連しています。各ケースが、生産ラインの別々の製品を表します。各製品に関して、直径、研磨時間、価格、製品タイプが記録されています。
- **poll\_cs.sav**。市民の法案支持率を議会開会前に特定するための、世論調査員の取り組みに関する架空のデータ ファイルです。各ケースは登録有権者に対応しています。ケースごとに、有権者が居住している郡、町、区域が記録されています。
- **poll\_cs\_sample.sav**。この架空のデータ ファイルには、poll\_cs.sav の有権者のサンプルが含まれています。サンプルは、poll\_csplan 計画ファイルで指定されている計画に従って抽出され、このデータ ファイルには包含確率およびサンプル重み付けが記録されています。ただし、抽出計画では確率比例 (PPS) 法を使用するため、結合選択確率を含むファイル (poll\_jointprob.sav) もあります。サンプル抽出後、有権者の人口統計および法案に関する意見に対応する追加の変数が収集され、データ ファイルに追加されました。
- **property\_assess.sav**。限られたリソースで資産価値評価を最新に保つための、郡の評価担当者の取り組みに関する架空のデータ ファイルです。各ケースは、前年に郡内で売却された資産に対応します。データ ファイル内の各ケースでは、資産が存在する町、最後に訪問した評価担当者、その評価からの経過時間、当時行われた評価、および資産の売却価値が記録されています。
- **property\_assess\_cs.sav**。限られたリソースで資産価値評価を最新に保つための、州の評価担当者の取り組みに関する架空のデータ ファイルです。各ケースは州内の資産に対応します。データ ファイル内の各ケースでは、資産が存在する郡、町、および区域、最後の評価からの経過時間、および当時行われた評価が記録されています。
- **property\_assess\_cs\_sample.sav**。この架空のデータ ファイルには、property\_assess\_cs.sav の資産のサンプルが含まれています。サンプルは、property\_assess\_csplan 計画ファイルで指定されている計画に従って抽出され、このデータ ファイルには包含確率およびサンプル重み付けが記録されています。サンプル抽出後、現在の価値変数が収集され、データ ファイルに追加されました。

- **recidivism.sav.** 管轄地域での累犯率を把握するための、政府の法執行機関の取り組みに関する架空のデータ ファイルです。各ケースは元犯罪者に対応し、人口統計情報、最初の犯罪の詳細、初犯から 2 年以内の場合には 2 回目の逮捕までの期間が記録されています。
- **recidivism\_cs\_sample.sav.** 管轄地域での累犯率を把握するための、政府の法執行機関の取り組みに関する架空のデータ ファイルです。各ケースは 2003 年の 7 月に最初の逮捕から釈放された元犯罪者に対応し、人口統計情報、最初の犯罪の詳細、2006 年 7 月までの 2 回目の逮捕のデータが記録されています。犯罪者は `recidivism_cs.csplan` で指定された抽出計画に従って抽出された部門から選択されます。調査では確率比例 (PPS) 法を採用したため、結合選択確率を保持したファイル (`recidivism_cs_jointprob.sav`) も用意されています。
- **rfm\_transactions.sav.** 購入日、購入品目、各取引のマネタリー量など、購買取引データを含む架空のデータ ファイルです。
- **salesperformance.sav.** 2 つの新しい販売トレーニング コースの評価に関する架空のデータ ファイルです。60 人の従業員が 3 つのグループに分けられ、全員が標準のトレーニングを受けます。さらに、グループ 2 は技術トレーニングを、グループ 3 は実践的なチュートリアルを受けます。トレーニング コースの最後に各従業員がテストを受け、得点が記録されました。データ ファイルの各ケースは別々の訓練生を表し、割り当てられたグループと、テストの得点が記録されています。
- **satisf.sav.** ある小売業者が 4 箇所の店舗で行った満足度調査に関する架空のデータ ファイルです。合計で 582 人の顧客を調査し、各ケースは 1 人の顧客からの回答を表します。
- **screws.sav.** このデータ ファイルには、ねじ、ボルト、ナット、鋸 (びょう) (Hartigan, 1975) の特性に関する情報が含まれています。
- **shampoo\_ph.sav.** あるヘアケア製品工場での品質管理に関する架空のデータ ファイルです。定期的に、6 つの異なる製品が測定され、pH が記録されます。目標範囲は 4.5 ~ 5.5 です。
- **ships.sav.** 他の場所 (McCullagh など, 1989) で表示および分析される、波による貨物船への損害に関するデータセットです。件数は、船舶の種類、建造期間、およびサービス期間によって、ポワゾン率で発生するものとしてモデリングできます。因子のクロス分類によって形成されたテーブルの各セルのサービス月数の集計によって、危険にさらされる確率の値が得られます。
- **site.sav.** 業務拡大に向けて新たな用地を選択するための、ある会社の取り組みに関する架空のデータ ファイルです。2 人のコンサルタントを雇って、用地を別々に評価させました。広範囲のレポートに加えて、各用地を「良い」、「普通」、「悪い」のいずれかで集計しました。
- **smokers.sav.** このデータ ファイルは、1998 年の National Household Survey of Drug Abuse から抜粋したものであり、アメリカの世帯の確率サンプルです。 (<http://dx.doi.org/10.3886/ICPSR02934>) したがっ

て、このデータ ファイルを分析する場合は、まず人口の傾向を反映させてデータを重み付けする必要があります。

- **stroke\_clean.sav.** この架空のデータ ファイルには、[データの準備] オプションの手続きを使用して整理した後の、医療データベースの状態が含まれています。
- **stroke\_invalid.sav.** この架空のデータ ファイルには、医療データベースの初期状態が含まれており、データ入力にいくつかエラーがあります。
- **stroke\_survival.** この架空のデータ ファイルは、虚血性脳卒中で数回の困難に直面した後リハビリ プログラムを終えた患者の生存時間に関するものです。脳卒中後、心筋梗塞の発生、虚血性脳卒中、または出血性脳卒中が注意され、イベントの時間が記録されます。脳卒中後に実施されたリハビリ プログラムの最後まで生存した患者のみが含まれるため、サンプルは左側が切り捨てられます。
- **stroke\_valid.sav.** この架空のデータ ファイルには、[データの検証] 手続きを使用して確認した後の、医療データベースの状態が含まれています。異常である可能性のあるケースが含まれています。
- **survey\_sample.sav.** このデータ ファイルには、人口統計データおよびさまざまな態度指標などの調査データが含まれています。これは「1998 NORC General Social Survey」の変数のサブセットに基づいていますが、いくつかのデータ値が変更され、追加の架空変数がデモの目的で追加されています。
- **telco.sav.** 顧客ベースにおける解約率を削減するための電気通信会社の取り組みに関する架空のデータ ファイルです。各ケースが別々の顧客に対応し、人口統計やサービス利用状況などのさまざまな情報が記録されています。
- **telco\_extra.sav.** このデータ ファイルは telco.sav データ ファイルに似ていますが、「期間」および対数変換された顧客支出の属性が削除され、標準化された対数変換顧客支出の変数に置き換えられています。
- **telco\_missing.sav.** このデータ ファイルは telco.sav データ ファイルのサブセットですが、一部の人口統計データ値が欠損値に置き換えられています。
- **testmarket.sav.** この架空のデータ ファイルは、新しいメニューを追加しようというファーストフードチェーンの計画に関連しています。新製品をプロモーションするためのキャンペーンには3つの候補があるため、新メニューはいくつかのランダムに選択した市場のある場所で紹介されます。場所ごとに別々のプロモーションを使用し、最初の4週間の新メニューの週間売上高が記録されます。各ケースが場所と週に対応します。
- **testmarket\_1month.sav.** この架空のデータ ファイルは、testmarket.sav データ ファイルの週ごとの売上を「ロールアップ」して、各ケースが別々の場所に対応するようにしたものです。その結果、週ごとによって



ていた変数の一部が表示されなくなり、売上高が、調査を行った 4 週間の売上高の合計になっています。

- **tree\_car.sav**。これは、人口統計および自動車購入価格のデータを含む架空のデータ ファイルです。
- **tree\_credit.sav**。これは、人口統計および銀行ローン履歴のデータを含む架空のデータ ファイルです。
- **tree\_missing\_data.sav**。これは、人口統計および銀行ローン履歴のデータと、多数の欠損値を含む架空のデータ ファイルです。
- **tree\_score\_car.sav**。これは、人口統計および自動車購入価格のデータを含む架空のデータ ファイルです。
- **tree\_textdata.sav**。尺度および値ラベルを割り当てる前の、変数のデフォルトの状態を示すことを主な目的とする、変数を 2 つだけ含む単純なデータ ファイルです。
- **tv-survey.sav**。テレビスタジオで実施された、ヒットした番組の放送期間を延長するかどうかを検討する調査に関する架空のデータ ファイルです。906 人の回答者に、さまざまな条件下でこの番組を視聴するかどうかを質問しました。各行は別々の回答者を表し、各列は別々の条件を表します。
- **ulcer\_recurrence.sav**。このファイルには、潰瘍の再発を防ぐための 2 つの治療の有効性を比較するように計画された調査の情報の一部が含まれています。これは区間調査の良い例であり、他の場所 (Collett, 2003) で表示および分析されています。
- **ulcer\_recurrence\_recoded.sav**。このファイルでは、ulcer\_recurrence.sav の情報が、単に調査終了時のイベント確率ではなく調査の区間ごとのイベント確率をモデリングできるように再編成されています。これは他の場所 (Collett など, 2003) で表示および分析されています。
- **verd1985.sav**。このデータ ファイルは調査 (Verdegaal, 1985) に関連しています。8 つの変数に対する 15 人の被験者の回答を記録しました。対象となる変数が 3 つのグループに分類されます。グループ 1 には「年齢」と「婚姻」、グループ 2 には「ペット」と「新聞」、グループ 3 には「音楽」と「居住地域」がそれぞれ含まれます。「ペット」は多重名義として尺度化され、「年齢」は順序として尺度化されます。また、その他のすべての変数は単一名義として尺度化されます。
- **virus.sav**。自社のネットワーク上のウィルスの影響を特定するための、インターネット サービス プロバイダ (ISP) の取り組みに関する架空のデータ ファイルです。この ISP は、ネットワーク上の感染した E メールトラフィックの (およその) パーセンテージを、発見の瞬間から脅威が阻止されるまで追跡しました。
- **wheeze\_steubenville.sav**。これは、子供 (Ware, Dockery, Spiro III, Speizer, および Ferris Jr., 1984) に対する大気汚染の健康上の影響の長期調査から得られたサブセットです。このデータには、オハイオ州スビューベンビルの 7 歳、8 歳、9 歳、10 歳の子供を対象に行った、喘鳴

の状態の反復 2 値測定と、調査の初年に母親が喫煙していたかどうかの固定記録が含まれています。

- **workprog.sav**。体の不自由な人をより良い仕事に就かせようとする政府の事業プログラムに関する架空のデータ ファイルです。プログラムの参加者候補のサンプルが追跡されました。その中には、ランダムに選ばれてプログラムに登録された人と、そうでない人がいました。各ケースが別々のプログラム参加者を表します。

# Notices

Licensed Materials – Property of SPSS Inc., an IBM Company. © Copyright SPSS Inc. 1989, 2010.

Patent No. 7,023,453

**The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:** SPSS INC., AN IBM COMPANY, PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. SPSS Inc. may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-SPSS and non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this SPSS Inc. product and use of those Web sites is at your own risk.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Information concerning non-SPSS products was obtained from the suppliers of those products, their published announcements or other publicly available sources. SPSS has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-SPSS products. Questions on the capabilities of non-SPSS products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

#### COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to SPSS Inc., for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. SPSS Inc., therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided “AS IS”, without warranty of any kind. SPSS Inc. shall not be liable for any damages arising out of your use of the sample programs.

#### Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) are trademarks of IBM Corporation, registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>.

SPSS is a trademark of SPSS Inc., an IBM Company, registered in many jurisdictions worldwide.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

This product uses WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>.

Other product and service names might be trademarks of IBM, SPSS, or other companies.

Adobe product screenshot(s) reprinted with permission from Adobe Systems Incorporated.

Microsoft product screenshot(s) reprinted with permission from Microsoft Corporation.



---

# 参考文献

- Bell, E. H. 1961. Social foundations of human behavior: Introduction to the study of sociology. New York: Harper & Row.
- Blake, C. L., および C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- Breiman, L., および J. H. Friedman(F). 1985. Estimating optimal transformations for multiple regression and correlation. Journal of the American Statistical Association, 80, .
- Cochran, W. G. 1977. Sampling Techniques, 3rd ed. New York: John Wiley and Sons.
- Collett, D. 2003. Modelling survival data in medical research, 2 ed. Boca Raton: Chapman & Hall/CRC.
- Cox, D. R., および E. J. Snell. 1989. The Analysis of Binary Data, 2nd ed. London: Chapman and Hall.
- Green, P. E., および V. Rao. 1972. Applied multidimensional scaling. Hinsdale, Ill.: Dryden Press.
- Green, P. E., および Y. Wind. 1973. Multiattribute decisions in marketing: A measurement approach. Hinsdale, Ill.: Dryden Press.
- Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. Psychometrika, 33, .
- Hartigan, J. A. 1975. Clustering algorithms. New York: John Wiley and Sons.
- Hastie, T., および R. Tibshirani. 1990. Generalized additive models. London: Chapman and Hall.
- Kennedy, R., C. Riquier, および B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. Journal of Targeting, Measurement, and Analysis for Marketing, 5, .
- Kish, L. 1965. Survey Sampling. New York: John Wiley and Sons.
- Kish, L. 1987. Statistical Design for Research. New York: John Wiley and Sons.
- McCullagh, P., および J. A. Nelder. 1989. Generalized Linear Models, 2nd ed. London: Chapman & Hall.
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In: Frontiers in Economics, P. Zarembka, ed. New York: Academic Press.

- Murthy, M. N. 1967. *Sampling Theory and Methods*. Calcutta, India: Statistical Publishing Society.
- Nagelkerke, N. J. D. 1991. A note on the general definition of the coefficient of determination. *Biometrika*, 78:3, .
- Price, R. H., および D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, .
- Rickman, R., N. Mitchell, J. Dingman, および J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228, .
- Rosenberg, S., および M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, .
- Särndal, C., B. Swensson, および J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Van der Ham, T., J. J. Meulman, D. C. Van Strien, および H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, .
- Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens* (in Dutch). Leiden: Department of Data Theory, University of Leiden.
- Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, および B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, .

# 索引

## 分離

- コンプレックス サンプルの順序回帰, 80
- コンプレックス サンプルの一般線型モデル, 53-54
- コンプレックス サンプルのロジスティック回帰, 66

## 要約表

- 分析準備ウィザード, 155
- 分析準備ウィザード, 165
- コンプレックス サンプルの順序回帰, 74
- コンプレックス サンプルの Cox 回帰, 95
- コンプレックス サンプルの一般線型モデル, 54-55
- コンプレックス サンプルのロジスティック回帰, 62
- サンプリング ウィザード, 114, 147

## 応答確率

- コンプレックス サンプルの順序回帰, 71, 78, 80
- コンプレックス サンプルの一般線型モデル, 53, 55
- コンプレックス サンプルのロジスティック回帰, 65-66

## 部分母集団

- コンプレックス サンプルの Cox 回帰, 89

## 推定周辺平均

- コンプレックス サンプルの一般線型モデル, 53

## 最大対数尤度比残差

- コンプレックス サンプルの Cox 回帰, 96

## 欠損値

- コンプレックス サンプルの順序回帰, 80
- コンプレックス サンプルのロジスティック回帰, 66

## 最小有意差

- コンプレックス サンプルの Cox 回帰, 95

## 標準誤差

- コンプレックス サンプルの順序回帰, 74
- コンプレックス サンプルの一般線型モデル, 50
- コンプレックス サンプルのロジスティック回帰, 62

## 対数尤度収束

- コンプレックス サンプルの順序回帰, 80
- コンプレックス サンプルのロジスティック回帰, 66

## 集計残差

- コンプレックス サンプルの Cox 回帰, 96

## 段階 2 分

- コンプレックス サンプルの順序回帰, 80
- コンプレックス サンプルのロジスティック回帰, 66

## 逐次 Bonferroni 補正

- コンプレックス サンプルの Cox 回帰, 95

## 逆 Helmert 対比

- コンプレックス サンプルの一般線型モデル, 53

## 擬 R<sup>2</sup> 統計量

- コンプレックス サンプルの順序回帰, 74
- コンプレックス サンプルのロジスティック回帰, 62

## 逐次 Sidak 補正

- コンプレックス サンプルの Cox 回帰, 95

## Bonferroni (B)

- コンプレックス サンプルの Cox 回帰, 95

## Bonferroni の方法

- コンプレックス サンプル, 52, 63, 76

## Breslow の推定方法

- コンプレックス サンプルの Cox 回帰, 100

## Brewer 抽出法

- サンプリング ウィザード, 8

## Complex Samples

- オプション, 33, 37, 42, 46
- 仮説の検定, 52, 63, 76
- 欠損値, 32, 41

## Cox-Snell 残差

- コンプレックス サンプルの Cox 回帰, 96

## Efron の推定方法

- コンプレックス サンプルの Cox 回帰, 100

## F 統計量

- コンプレックス サンプルの Cox 回帰, 95

## Fisher スコア法 (F)

- コンプレックス サンプルの順序回帰, 80

## F 統計量

- コンプレックス サンプル, 52, 63, 76



- Helmert 対比  
 コМПレックス サンプルの一般線型モデル,  
 53
- legal notices, 293
- Murthy 抽出法  
 サンプルング ウィザード, 8
- Newton-Raphson 法  
 コМПレックス サンプルの順序回帰, 80
- PPS 抽出  
 サンプルング ウィザード, 8
- $R^2$  乗統計量  
 コМПレックス サンプルの一般線型モデル,  
 50
- Sampford 抽出法  
 サンプルング ウィザード, 8
- Schoenfeld の偏残差  
 コМПレックス サンプルの Cox 回帰, 96
- Sidak 補正  
 コМПレックス サンプルの Cox 回帰, 95
- Sidak 補正  
 コМПレックス サンプル, 52, 63, 76
- t 検定  
 コМПレックス サンプルの順序回帰, 74  
 コМПレックス サンプルの一般線型モデル,  
 50  
 コМПレックス  
 サンプルのロジスティック回帰,  
 62
- trademarks, 294
- 一般化線型モデル  
 コМПレックス サンプルの順序回帰, 224
- 一般使用データ  
 コМПレックス サンプルの記述統計, 172
- オッズ比  
 コМПレックス サンプルの順序回帰, 77  
 コМПレックス  
 サンプルのロジスティック回帰,  
 64
- オッズ比  
 コМПレックス サンプルのクロス集計表,  
 178  
 コМПレックス サンプルの順序回帰, 223  
 コМПレックス  
 サンプルのロジスティック回帰,  
 211
- オッズ比  
 コМПレックス サンプルのクロス集計表, 40
- 重み付けのない度数  
 コМПレックス サンプル記述統計, 36  
 コМПレックス サンプルのクロス集計表, 40  
 コМПレックス サンプルの度数分布表, 31
- コМПレックス サンプルの比率分析, 44
- カイ 2 乗  
 コМПレックス サンプルの Cox 回帰, 95
- カイ 2 乗  
 コМПレックス サンプル, 52, 63, 76
- 予測カテゴリ  
 コМПレックス サンプルの順序回帰, 78  
 コМПレックス サンプルの一般線型モデル,  
 53  
 コМПレックス  
 サンプルのロジスティック回帰,  
 59, 65
- 擬  $R^2$  統計量  
 コМПレックス サンプルの順序回帰, 219,  
 228  
 コМПレックス  
 サンプルのロジスティック回帰,  
 208
- 期待値  
 コМПレックス サンプルのクロス集計表, 40
- 行パーセント  
 コМПレックス サンプルのクロス集計表, 40
- 区分的に一定な時間依存の予測変数  
 コМПレックス サンプルの Cox 回帰, 247
- クラスタ  
 サンプルング ウィザード, 6  
 分析準備ウィザード, 22
- クロス表  
 コМПレックス サンプルのクロス集計表,  
 182
- 計画ファイル, 2
- 警告  
 コМПレックス サンプルの順序回帰, 227
- 系統的な抽出  
 サンプルング ウィザード, 8
- 欠損値  
 コМПレックス サンプル, 32, 41  
 コМПレックス サンプル記述統計, 37  
 コМПレックス サンプルの比率分析, 45
- 合計  
 コМПレックス サンプル記述統計, 36
- コМПレックス サンプル Cox 回帰  
 作図, 94  
 統計量, 92  
 予測変数, 86
- Kaplan-Meier 分析, 82

## 索引

- サブグループ, 89
- 日付変数と時刻変数, 82
- 事象の定義, 85, 95
- 時間依存の予測変数, 87
- モデル, 90
- コンプレックス サンプル Cox 回帰
  - オプション, 100
  - 変数の保存, 96
  - モデルをエクスポート, 98
- コンプレックス サンプルの順序回帰, 69
  - 統計量, 74
  - 応答確率, 71
  - モデル, 72
- コンプレックス サンプルの順序回帰
  - オッズ比, 77
  - オプション, 80
  - 変数の保存, 78
- コンプレックス
  - サンプルの分析準備ウィザード, 152
    - 要約表, 155, 165
    - 関連手続き, 166
    - 一般使用データ, 152
    - 抽出重み付けが使用できない, 155
- コンプレックス サンプルのサンプリング
  - ウィザード, 104
    - 要約表, 114, 147
    - PPS 抽出, 135
    - 関連手続き, 151
    - サンプリング枠、完全, 104, 116
- コンプレックス サンプルの一般線型モデル, 47
  - 統計量, 50
  - モデル, 49
- コンプレックス サンプルの一般線型モデル
  - 推定平均値, 53
  - オプション, 55
  - コマンドの追加機能, 56
  - 変数の保存, 54
- コンプレックス
  - サンプルのロジスティック回帰, 57
    - 統計量, 62
    - オッズ比, 64
    - オプション, 66
    - 参照カテゴリ, 59
    - コマンドの追加機能, 67
    - 変数の保存, 65
    - モデル, 60
- コンプレックス サンプル記述統計, 172
  - 一般使用データ, 172
  - 関連手続き, 176
  - 統計量, 175
  - 部分母集団による統計, 176
- コンプレックス サンプル記述統計, 34
  - 欠損値, 37
  - 統計量, 36
- コンプレックス サンプル度数分布表, 30
  - 統計量, 31
- コンプレックス サンプルの Cox 回帰, 230
  - 区分的に一定な時間依存の予測変数, 247
  - サンプルデザインの情報, 242, 276
  - 時間依存の予測変数, 230
  - ハザードの比例の検定, 242
  - パターン値, 278
  - パラメータ推定値, 247, 277
  - モデル効果の検定, 242, 246, 277
  - ログマイナスログ プロット, 279
- コンプレックス サンプルの一般線型モデル, 191
  - 関連手続き, 202
  - 周辺平均, 199
  - パラメータ推定値, 198
  - モデル効果の検定, 197
  - モデルの要約, 197
- コンプレックス サンプルのクロス集計表, 178
  - 関連手続き, 184
  - クロス表, 182
  - 相対リスク, 178, 182-183
- コンプレックス サンプルのクロス集計表, 38
  - 統計量, 40
- コンプレックス サンプルの順序回帰, 214
  - 一般化線型モデル, 224
  - オッズ比, 223
  - 関連手続き, 229
  - 擬 R<sup>2</sup> 統計量, 219, 228
  - 警告, 227
  - パラメータ推定値, 220
  - 分類表, 222
  - モデル効果の検定, 220
- コンプレックス サンプル度数分布表, 167
  - 関連手続き, 171
  - 度数分布表, 170
  - 部分母集団による度数分布表, 170
- コンプレックス
  - サンプルのロジスティック回帰, 204
    - オッズ比, 211
    - 関連手続き, 213
    - 擬 R<sup>2</sup> 統計量, 208
    - パラメータ推定値, 210
    - 分類表, 209
    - モデル効果の検定, 210
- コンプレックス サンプル比率, 185
  - 関連手続き, 190
  - 比率, 188
- コンプレックス サンプル比率, 43
  - 欠損値, 45

- 統計量, 44
- 最小有意差  
 コンプレックス サンプル, 52, 63, 76
- サイズの測定単位  
 サンプルング ウィザード, 8
- 残差  
 コンプレックス サンプルのクロス集計表, 40
- サンプルング枠、完全  
 サンプルング ウィザード, 104, 116
- サンプル デザインの情報  
 コンプレックス サンプルの Cox 回帰, 92
- サンプル ファイル  
 位置, 281
- サンプル重み付けの入力  
 サンプルング ウィザード, 6
- サンプル計画, 4
- サンプル サイズ  
 サンプルング ウィザード, 10, 12
- サンプル デザインの情報  
 コンプレックス サンプルの Cox 回帰, 242, 276
- サンプルの重み付け  
 サンプルング ウィザード, 12  
 分析準備ウィザード, 22
- サンプルの比率  
 サンプルング ウィザード, 12
- 時間依存の予測変数  
 コンプレックス サンプルの Cox 回帰, 230
- 自由度  
 コンプレックス サンプル, 52, 63, 76
- 周辺平均  
 GLM 1 変量, 199
- 信頼区間  
 コンプレックス サンプル記述統計, 36  
 コンプレックス サンプルの記述統計, 175–176  
 コンプレックス サンプルのクロス集計表, 40  
 コンプレックス サンプルの度数分布表, 170  
 コンプレックス サンプルの度数分布表, 31  
 コンプレックス サンプルの比率分析, 44
- スコア残差  
 コンプレックス サンプルの Cox 回帰, 96
- 層化 (層別)  
 サンプルング ウィザード, 6  
 分析準備ウィザード, 22
- 相対リスク  
 コンプレックス サンプルのクロス集計表, 178, 182–183
- コンプレックス サンプルのクロス集計表, 40
- 単純な無作為抽出  
 サンプルング ウィザード, 8
- 逐次 Bonferroni 補正  
 コンプレックス サンプル, 52, 63, 76
- 逐次 Sidak 補正  
 コンプレックス サンプル, 52, 63, 76
- 逐次抽出  
 サンプルング ウィザード, 8
- 抽出  
 コンプレックス デザイン, 4
- 抽出推定  
 分析準備ウィザード, 23
- 抽出方法  
 サンプルング ウィザード, 8
- 調整済み F 統計量  
 コンプレックス サンプル, 52, 63, 76
- 調整済みカイ 2 乗  
 コンプレックス サンプル, 52, 63, 76
- 調整済み残差  
 コンプレックス サンプルのクロス集計表, 40
- デザイン効果  
 コンプレックス サンプルの順序回帰, 74  
 コンプレックス サンプルの Cox 回帰, 92  
 コンプレックス サンプルの一般線型モデル, 50  
 コンプレックス  
 サンプルのロジスティック回帰, 62
- デザイン効果  
 コンプレックス サンプル記述統計, 36  
 コンプレックス サンプルのクロス集計表, 40  
 コンプレックス サンプルの度数分布表, 31  
 コンプレックス サンプルの比率分析, 44
- デザイン効果の平方根  
 コンプレックス サンプル記述統計, 36  
 コンプレックス サンプルのクロス集計表, 40  
 コンプレックス サンプルの度数分布表, 31  
 コンプレックス サンプルの比率分析, 44
- デザイン効果の平方根  
 コンプレックス サンプルの一般線型モデル, 50
- デザイン効果の平方根  
 コンプレックス サンプルの Cox 回帰, 92
- デザイン効果の平方根  
 コンプレックス サンプルの順序回帰, 74  
 コンプレックス  
 サンプルのロジスティック回帰, 62

## 索引

- 一般使用データ
  - 分析準備ウィザード, 152
- 全平均との対比
  - コンプレックス サンプルの一般線型モデル, 53
- R<sup>2</sup> 乗統計量
  - コンプレックス サンプルの一般線型モデル, 197
- 反復の記述
  - コンプレックス サンプルの順序回帰, 80
  - コンプレックス サンプルのロジスティック回帰, 66
- 多項式の対比
  - コンプレックス サンプルの一般線型モデル, 53
- 時間依存の予測変数
  - コンプレックス サンプルの Cox 回帰, 87
- 平行線の検定
  - コンプレックス サンプルの順序回帰, 74
- ハザードの比例の検定
  - コンプレックス サンプルの Cox 回帰, 242
- ハザードの比例の検定
  - コンプレックス サンプルの Cox 回帰, 92
- パラメータ推定値
  - コンプレックス サンプルの順序回帰, 74
  - コンプレックス サンプルの一般線型モデル, 50
  - コンプレックス サンプルのロジスティック回帰, 62
- パラメータ収束基準
  - コンプレックス サンプルのロジスティック回帰, 66
- パラメータ推定値
  - コンプレックス サンプルの Cox 回帰, 92
- パラメータ収束基準
  - コンプレックス サンプルの順序回帰, 80
- パラメータ推定値
  - コンプレックス サンプルの一般線型モデル, 198
  - コンプレックス サンプルの順序回帰, 220
  - コンプレックス サンプルのロジスティック回帰, 210
- パラメータ推定値の共分散
  - コンプレックス サンプルのロジスティック回帰, 62
- パラメータ推定値の相関
  - コンプレックス サンプルの一般線型モデル, 50
- パラメータ推定値の共分散
  - コンプレックス サンプルの順序回帰, 74
  - コンプレックス サンプルの一般線型モデル, 50
- パラメータ推定値の相関
  - コンプレックス サンプルのロジスティック回帰, 62
- パラメータ推定値の相関
  - コンプレックス サンプルの順序回帰, 74
- 標準誤差
  - コンプレックス サンプルの記述統計, 175-176
  - コンプレックス サンプルの度数分布表, 170
- 標準誤差
  - コンプレックス サンプル記述統計, 36
  - コンプレックス サンプルのクロス集計表, 40
  - コンプレックス サンプルの度数分布表, 31
  - コンプレックス サンプルの比率分析, 44
- 表パーセント
  - コンプレックス サンプルのクロス集計表, 40
  - コンプレックス サンプルの度数分布表, 170
  - コンプレックス サンプルの度数分布表, 31
- 比率
  - コンプレックス サンプルの比率分析, 188
- 複合抽出
  - サンプル計画, 4
  - 分析計画, 20
- 分析計画, 20
- 分類表
  - コンプレックス サンプルの順序回帰, 222
  - コンプレックス サンプルのロジスティック回帰, 209
- 平均値
  - コンプレックス サンプル記述統計, 36
  - コンプレックス サンプルの記述統計, 175-176
- 平行線の検定
  - コンプレックス サンプルの順序回帰, 224

## 変動係数 (COV)

- コンプレックス サンプル記述統計, 36
  - コンプレックス サンプルのクロス集計表, 40
  - コンプレックス サンプルの度数分布表, 31
  - コンプレックス サンプルの比率分析, 44
- ベースライン ストラータ
- コンプレックス サンプルの Cox 回帰, 89

## コンプレックス

- サンプルのロジスティック回帰, 66

## ログマイナスログ プロット

- コンプレックス サンプルの Cox 回帰, 279

## 包含確率

- サンプリング ウィザード, 12

## 母集団のサイズ

- コンプレックス サンプル記述統計, 36
- コンプレックス サンプルのクロス集計表, 40
- コンプレックス サンプルの度数分布表, 170
- コンプレックス サンプルの度数分布表, 31
- コンプレックス サンプルの比率分析, 44
- サンプリング ウィザード, 12

## マーチンゲール残差

- コンプレックス サンプルの Cox 回帰, 96

## 調整済み F 統計量

- コンプレックス サンプルの Cox 回帰, 95

## 調整済みカイ 2 乗

- コンプレックス サンプルの Cox 回帰, 95

## モデル効果の検定

- コンプレックス サンプルの Cox 回帰, 277
- コンプレックス サンプルの一般線型モデル, 197
- コンプレックス サンプルの順序回帰, 220
- コンプレックス サンプルのロジスティック回帰, 210

## 予測変数のパターン

- コンプレックス サンプルの Cox 回帰, 278

## リスク差

- コンプレックス サンプルのクロス集計表, 40

## 累計値

- コンプレックス サンプルの度数分布表, 31

## 列パーセント

- コンプレックス サンプルのクロス集計表, 40

## 確信度レベル

- コンプレックス サンプルの順序回帰, 80