Chapter 1	変数の計算	1-1
1.1.	1 変数の計算の概要	1-2
1.1.2	2 オプションの設定	1-3
1.1.3	3 変数の計算	1-6
1.2.	1 関数を利用した変数の計算	1-13
1.3.	1 ブール論理式を利用したフラグ変数の作成	1-18
1.4.	1 IF条件を含む変数の計算	1-22
Chapter 2	値の再割り当て	2-1
2.1.	1 値の再割り当ての概要	2-2
2.2.	1 ダミー変数	2-3
2.2.2	2 ダミー変数の作成ー他の変数への値の再割り当てー	2-4
2.3.	1 欠損値の置換-同一の変数への値の再割り当て	2-19
2.4.	1 年齢から年代の作成-連続変数のカテゴリ化-	2-25
2.5.	1 文字型を数値型へ変換ー連続数への再割り当て	2-33
Chapter 3	ケースの選択	3-1
3.1.	1 ケースの選択の概要	3-2
3.2.	1 IF条件によるケースの選択-分析から除外	3-3
3.2.	3 IF条件によっケースの選択-新しいデータセットにコピー-	3-10
3.3.	1 無作為抽出によるケースの選択	3-13
3.4.	1 フィルタ変数によるケースの選択	3-18
3.4.2	2 ケースの選択の解除	3-20
3.5.	1 ファイルの分割の概要	3-22
3.5.2	2 ファイルの分割の概要-別テーブルへの分割-	3-23
3.5.	3 ファイルの分割の概要-同一テーブル内での分割-	3-28
3.5.4	4 ファイルの分割の解除	3-31
Chapter 4	日付データの加工	4-1
4.1.	1 日付データの概要	4-2
4.1.2	2 日付と時刻ウィザード	4-3

4.2.	.1 日付書式の変更	4-5
4.3.	.1 生年月日から年齢を表す変数の作成	4-11
4.4.	.1 日付データの一部を抽出	4-17
4.5.	.1 平日を表す変数の作成	4-22
Chapter 5	ファイルの結合	5-1
5.1.	.1 ファイル結合の概要	5-2
5.2.	.1 ファイルの結合-ケースの追加の概要-	5-3
5.2.	.2 ファイルの結合-ケースの追加-	5-5
5.3.	.1 ファイルの結合 – 変数の追加の概要 –	5-17
5.3.	.2 両方のファイルがケースを提供する結合	5-19
5.3.	.3 検索テーブルによる結合	5-21
5.3.	.4 ファイルの結合-変数の追加-	5-23
Chapter 6	その他の加工機能	6-1
6.1.	.1 グループ集計の概要	6-2
6.1.		
6.1.	3 グループ集計とよる集計データファイルの作成	6-8
6.1.	4 集計データに基づく散布図の作成	6-11
6.2.	.1 ケースの並べ <mark>替え</mark>	6-14
6.3.	.1 ケースの重み付け	6-18

本書では、IBM SPSS Statistics for Windows 27 を使用しています。 IBM およびSPSSは、International Business Machines Corp.の登録商標です。

本書を無断で複写複製(コピー)することは、著作権法上の例外を除き、禁じられています。

2

値の再割り当て

値の再割り当て Recode variables は、変数の元の値から、別の値に置き換えるためのデータ加工の機能です。分析の目的によっては、多数のカテゴリを少数にカテゴリに分類する、目的のカテゴリとそれ以外のカテゴリに再分類する場合などがあります。また、年齢から年代を作成する例を典型として、量的変数(スケール)を質的変数(カテゴリ)へ分類する場合があります。

ここでは、IBM SPSS Statisticsを利用して、値を変換して同じ変数または別の変数へ再割り当てする方法、量的変数(スケール)を質的変数(カテゴリ)へ、文字型のカテゴリデータを数値コードへ変換する具体的な手順を確認します。

Contents

- 値の再割り当ての概要
- ダミー変数の作成 -他の変数への値の再割り当て-
- 欠損値の置換 一同一の変数への値の再割り当て一
- 年齢から年代の作成 / 連続変数のカテゴリ化-
- 文字型変数を数値に変換 -連続数への再割り当て-

Keyword

値の再割り当て/連続変数のカテゴリ化/連続数への再割当て/システム欠損値/

§2.1.1 値の再割り当ての概要

値の再割り当て Recode variables は、特定の変数の値を新しい別の値に置き換えるための機能です。たとえば、多数の疾患名(名義尺度)を特定の疾患ごとに少数のグループに分類したり、体重(スケール)を一定の間隔で階級値(順序尺度)に分類したりすることができます。また、重症度の変数から重症(あり/なし)と軽症(あり/なし)のように2値に分類するダミー変数の作成を行う場合などに利用することも可能です。

IBM SPSS Statisticsでは、値の再割り当てを行うための主な**変換** Transform メニューとして、 以下の4つの機能を利用することができます。

メニュー	機能
同一変数への値の再割り当て	値の再割り当て(上書き)
他の変数への値の再割り当て	値の再割り当て(別名)
連続数への再割り当て	文字型変数を数値型変数へ変換
連続変数のカテゴリ化	スケール変数をカテゴリに分類

Table2.1.1 変換メニューに含まれる主な再割り当て機能

同一変数への値の再割り当て Recode into Same variables は、変数の値を新しい値に変換して同じ変数へ結果を上書きします。一方、他の変数への値の再割り当て Recode into Different variables は、変数の値を新しい値に変換して新しい変数へ結果を割り当てます。

連続数への再割り当て Automatic recode は、カテゴリのデータを数値に変換する場合に利用します。例えば、文字型で記録されている性別のカテゴリ(男性/女性)の値を(1(男性)/2(女性))に変換することができます。数値型変数しか扱えない統計手続きや、数値型のコードの方が扱いやすい場合にこのような変換が必要になります。

連続変数のカテゴリ化 Visual Binning は、量的変数 (スケール) をカテゴリに変換する機能です。例えば、体重の値を5kgごとに階級値に分類したり、年齢の値を10歳間隔で年代に変換したりする場合などに利用されます。

§2.2.1 ダミー変数

ダミー変数 dummy variable は、0と1でコード化された2値変数で、回帰分析やロジスティック回帰分析などの量的変数(スケール)を用いる分析手法において、質的変数(カテゴリ)を利用する場合に用いられる変数です。

たとえば、ロジスティック回帰分析では、独立変数として重症度(軽症/中等症/重症)のような質的変数を用いることはできません。この場合、以下の表のように質的変数を2値変数に変換することで、独立変数として利用することができます。

重症度	軽症	中等症
1 軽症	1	0
2 中等症	0	
3 重症	0	0

Table 2.2.1 重症度のダミーコード化

この例では、**軽症**のダミー変数は軽症かどうかを意味し、**中等症**のダミー変数は中等症かどうかを意味します。**重症**のカテゴリは解釈を行う際の<u>基準(参照)カテゴリ</u>として機能し、軽症と中等症の2つのダミー変数の値がともに0の場合、重症であることを意味します。

2値変数は、比率尺度の変数として量的変数(スケール)として分析することができるため、 元の質的変数(カテゴリ)をダミー変数に置き換えることで、ロジスティック回帰分析の独立変 数として利用することができます。

POINT

質的変数(カテゴリ)を、0と1のコードに置き換えることによって作成される変数を、ダミー変数と呼びます。ダミー変数は、回帰分析やロジスティック回帰分析などでよく利用されます。

§2.2.2 ダミー変数の作成 -他の変数への値の再割り当て-

値の再割り当ての機能を使用することで、変数の値を指定した新しい値に置き換えることができます。

ここでは、患者の属性と生活習慣や重症度を記録したサンプルデータファイルLesson2.sav を利用して、値の再割り当て機能によるダミー変数の手順を確認します。

操作手順

- 1. **ファイル**メニュー > **開く** > **データ**を選択します。
- 2. Lesson2.savを開きます。
- 3. ツールバーの**値ラベル**ボタンをクリックします。

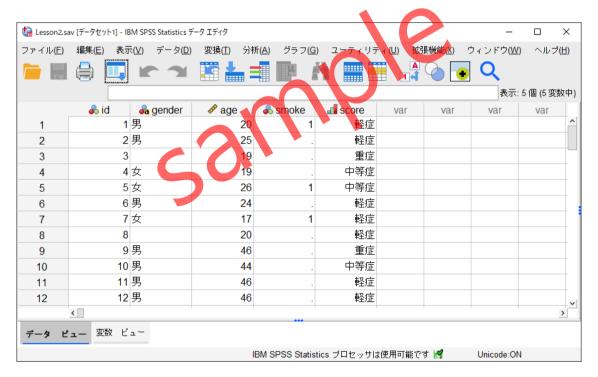


Figure 2.2.1 Lesson 2.sav データファイルのデータビュー(値ラベルを表示)

scoreには、**1**軽症、**2**中等症、**3**重症、の**3**つのカテゴリが記録されています。この変数に基づいて、**軽症**を表す変数と**中等症**を表す変数を作成します。

TIPS

カテゴリ変数に設定されている値ラベルを表示するためには、ツールバーの**値ラベル**ボタンをクリックします。

操作手順

4. 変換メニュー > **他の変数への値の再割り当て**を選択します。

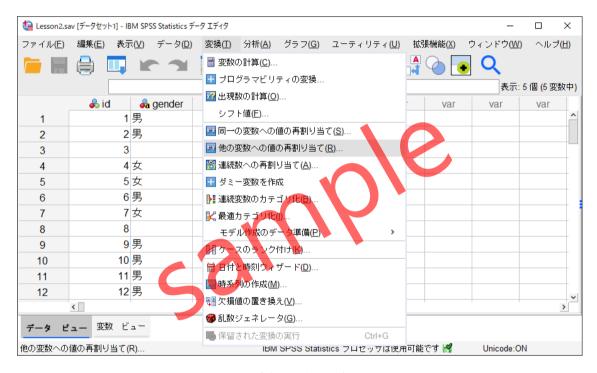


Figure 2.2.4 他の変数への値の再割り当てメニュー

TIPS

同一の変数への値の再割り当てを利用する場合、再割り当てされた値によって元の値が上書きされることになります。変換前の元の変数を残しておきたい場合は、他の変数への値の再割り当てを利用してください。

他の変数への値の再割り当てを行うためのダイアログボックスが表示されます。この画面で変換の元となる入力変数を選択して、変換後の新しい変数名を指定します。

操作手順

5. scoreを入力変数のリストに移動します。



ダイアログボックス右側の**変換先変数**のテキストボックスがアクティブになります。次に、新しく作成する変数の**名前**(必須)と**ラベル**(任意)を定義します。変換先変数の名前の命名規則は、一般的な変数名の規則に準拠します。

TIPS

値の再割り当てには、数値型変数だけではなく文字型変数を直接使用することもできます。

操作手順

- 6. 変換先変数の名前に軽症と入力します。
- 7. 変更ボタンをクリックします。



Figure 2.2.6 変換先変数の名前の設定

重症度の変数**score**に基づいて、**軽症**を表す新しい変数を作成するための名前の設定が完了しました。

POINT

変換先変数の名前を入力した後、必ず**変更**ボタンをクリックする必要があります。変更ボタンをクリックすることで入力内容がダイアログボックスに反映されます。

操作手順

8. 今までの値と新しい値ボタンをクリックします。

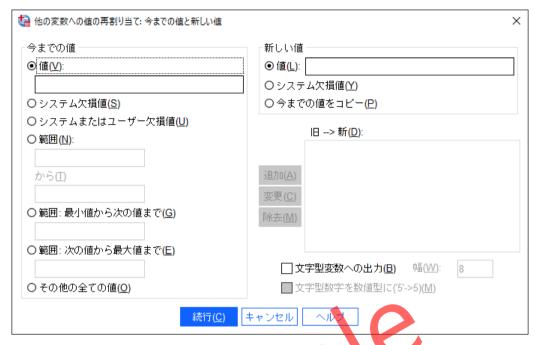


Figure2.2.7 今までの値と新しい値の設定ダイアログボックス

今までの値と新しい値の指定を行うための画面が表示されます。画面の左側の領域で、変換する前の元の値を指定し、右側の領域で変換後の新しい値の指定を行います。今までの値の指定では、個別の値のほか、値の範囲、システム欠損値、システムまたはユーザー欠損値などを指定することができます。

TIPS

ユーザー欠損値とは、ユーザーが指定することで欠損値として扱われる値のことです。ユーザー欠損値の設定は、データセットの**変数ビュー**で指定することができます。

この例では、**重症度=1**(軽症)のケースを**1、重症度=2**(中等症)と**重症度=3**(重症)のケースを**0** に変換します。

操作手順

- 9. **今までの値**の**値**を選択し、テキストボックスに**1**と入力します。
- 10. **新しい値**の値を選択し、テキストボックスに**1**と入力します。



Figure 2.2.8 今までの値と新しい値の設定

操作手順

11. 追加ボタンをクリックします。

POINT

値の指定を行っても、**追加**ボタンをクリックして確定しないと操作内容が失われます。必 ず追加ボタンをクリックしてください。

操作手順

- **12. 今までの値**の**範囲**を選択し、1つ目のテキストボックスに**2**、2つ目のテキストボックスに**3**と入力します。
- 13. 新しい値の値を選択し、テキストボックスに**0**と入力します。



操作手順

14. 追加ボタンをクリックします。



Figure 2.2.10 値の再割り当ての指定

以上で、今までの値と新しい値の割り当ての定義は完了です。

TIPS

他の変数へ値の再割り当てを行う場合、指定されていない値は新しい変数においてシステム 欠損値となります。たとえば、重症度に4の値が含まれている場合、この値は変換後の変数 である軽症ではシステム欠損値が割り当てられます。

操作手順

- 15. 続行ボタンをクリックします。
- 16. **OK**ボタンをクリックします。

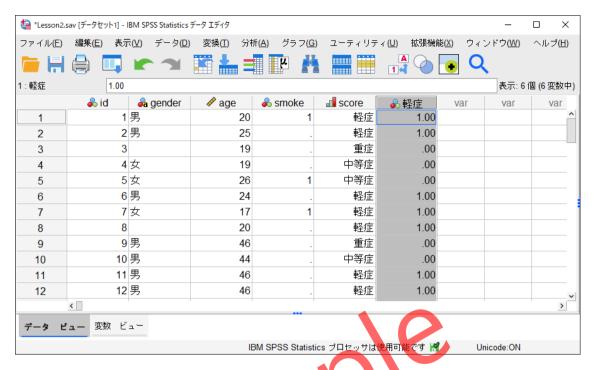


Figure 2.2.11 軽症を表す新しいダミー変数

元の変数である $\mathbf{score}(\mathbf{重症}\mathbf{E})$ に基づいて、**軽症**($\mathbf{score=1}$)のケースは新しく作成された変数である**軽症**において**1**に分類されています。また、中等症($\mathbf{score=2}$)と重症($\mathbf{score=3}$)については新しい変数で**0**に値が再割り当てされていることが分かります。

TIPS

新しく作成する変数は、データビューの右端に追加されます。変数の場所を移動する場合は、変数名(列ラベル)をクリックした後、再度変数名をクリックしてドラッグします。

TIPS

IBM SPSS Statisticsで新しく作成された数値型変数は、小数点以下2桁まで表示されます。この表示は変数ビューの**小数桁数**の設定を使用することで変更することができます。

5

ファイルの結合

ファイルの結合 Merge Files は、複数のファイルに含まれるケースや変数を統合するための機能です。ファイルの結合には、ケースの追加と変数の追加の2種類があります。たとえば、昨年度と今年度の測定結果が別々のデータファイルに含まれている場合、分析を行う前にデータファイルを1つに結合しておく必要があります。また、あるデータファイルに対象者の性別や年齢などの基本情報が含まれており、別のデータファイルに詳細な検査結果の変数が含まれている場合、これらを結合することで基本情報と検査結果を組み合わせた分析ができるようになります。

Contents

- ファイルの結合-ケースの追加-
- ファイルの結合-変数の追加-
- 両方のファイルがケースを提供する結合
- 検索テーブルによる結合

Keyword

ファイルの結合 ケースの追加 変数の追加 / キー変数 /

§5.1.1 ファイル結合の概要

ファイルの結合は、複数のファイルに含まれるケースや変数を統合するための機能です。 IBM SPSS Statisticsでは、1つのデータファイルのみを集計や分析の対象とするため、必要なデータが複数のファイルに別々に記録されている場合は、分析を実行する前にデータファイルを結合しておかなければなりません。

IBM SPSS Statisticsのファイルの結合では、結合する情報によって、ケースの追加と変数の 追加の2つの機能を使い分けます。

	ケースの追加	変数の追加
結合される情報	ケース(行)	変数(列)
結合の基準	変数名の一致	キー変数(ID)の値の一致
	データ型の一致	

Table5.1.1 ケースの追加と変数の追加

ケースの追加は、異なるデータファイルのケースを追加する場合に使用します。たとえば、30人分のデータと20人分のデータを結合して、50人分のデータファイルを作成するような場合です。ケースの追加では、変数名とデータ型の一致が結合の条件となります。

変数の追加は、異なるデータファイルの変数を追加する場合に使用します。たとえば、あるファイルに対象者の性別や年齢、別のファイルに血糖値やコレステロールなどの検査結果が記録されており、これらを結合して1つのデータファイルを作成するような場合です。変数の追加では、キー変数(ID)の一致を結合の条件とするのが基本です。

POINT

ファイルの結合には、ケースの追加と変数の追加の2種類の方法があります。

§5.2.1 ファイルの結合 ケースの追加の概要

ケースの追加では、**変数名**と**データ型**の一致を条件として異なるファイルのケースが結合されます。例として、以下のファイルAとファイルBを、ケースの追加によって1つのデータファイルに統合する場合を考えてみます。

ファイルA

ID	年齢	性別	検査値	喫煙
1	25	男性	120	0
2	32	男性	150	1
3	52	女性	98	1
4	58	男性	115	0
5	71	女性	89	0

Table5.2.1 ケースの追加前のファイル

ファイルB

ID	年齢	性別	検査値	喫煙の有無
6	56	0	180	1
7	28	1	150	0
8	64	1	95	1
9	42	0	100	1
10	40	1	75	1

Table 5.2.2 ケースの追加前のファイルB

ケースの追加では、**変数名**と**データ型**の一致を条件としてケースが結合されますので、**ID**、**年齢、検査値**が結合対象となります。ファイルAの**喫煙**とファイルBの**喫煙の有無**は変数名が 異なっているため結合されません。この場合は、ファイル結合の前に変数名を一致させておくか、ファイル結合を実行する際に一致させる必要があります。

また、**データ型**も一致している必要があります。この例では、ファイルAの**性別**とファイル Bの**性別**は変数名が一致していますが、一方は**文字型**、もう一方が**数値型**とデータ型が異なる ため結合されません。この場合は、事前に変数のデータ型を一致させておく必要があります。 さらに、変数が文字型の場合は、文字数(析数)も一致させておく必要があります。

以上の前提に基づいて、ファイルAの性別を**値の再割り当て**によって数値型に変換し、ファイルBの喫煙の有無の変数名を変数ビューで変更(喫煙の有無→喫煙)して、ケースの結合を実行すると、以下のデータセットになります。

ケースの追加後のファイル

ID	年齢	性別	検査値	喫煙
1	25	0	120	0
2	32	0	150	1
3	52	1	98	1
4	58	0	115	0
5	71	1	89	0
6	56	0	180	1
7	28	1	150	0
8	64	1	95	1
9	42	0	100	1
10	40	1	75	1

Table 5.2.3 ケースの追加後のファイル

POINT

ケースの追加を行うための条件は、以下のとおりです。

- 1. 結合する変数の変数名の一致
- 2. 結合する変数のデータ型の一致
- 3. 結合する変数のデータ型が文字型の場合は析数の一致

なお、使用するデータファイルが、IBM SPSS Statistics形式(.sav)ではなく、Microsoft Excel など他の形式のデータファイルの場合は、あらかじめIBM SPSS Statisticsに読み込んでおく必要があります。

§5.2.2 ファイルの結合 ケースの追加

ケースの追加では、変数名とデータの型の一致を条件として、2つのファイルを結合します。

この例では、2つの治療法による生存時間を使用します。治療法1におけるデータを含む Lesson5-1.sav(n=315)と治療法2のデータを含むLesson5-2.sav(n=304)の2つのIBM SPSS Statistics形式のデータファイルを結合する手順を確認します。

操作手順

- 1. **ファイル**メニュー > **開く** > **データ**を選択します。
- 2. Lesson5-1.savを開きます。

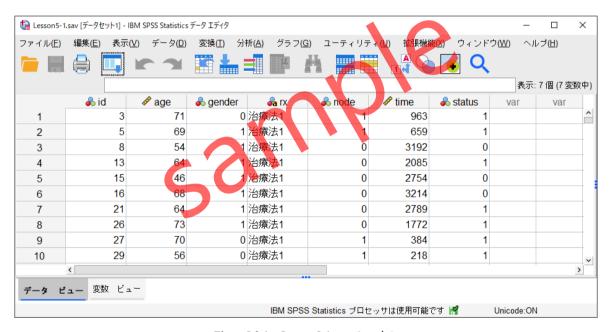


Figure 5.2.1 Lesson 5-1.savファイル

このデータファイルには、治療法1のデータとして**315**人分のケースが含まれています。**ID**、**age**(年齢)、**gender**(性別)、**rx**(治療法)など7個の変数を含んでいます。

操作手順

3. **変数ビュー**を選択します。

Lesson5-1	1.sav [データセット1]	- IBM SPSS Stat	istics データ エディ	9				- 🗆	×
ファイル <u>(E)</u>	編集(<u>E</u>) 表	示(⊻) データ	マ(<u>D</u>) 変換(<u>T</u>)) 分析(<u>A</u>)	グラフ(<u>G</u>) ユーティリティ(<u>U</u>)	拡張機能(<u>X</u>)	ウィンドウ(<u>W</u>)	ヘルブ <u>(H)</u>	
					μμμ	M 🔷 💽	Q		
	名前	型	幅	小数桁数	ラベル	値	欠損値	列	
1	id	数値	3	0		なし	なし	8	^
2	age	数値	2	0	年齢	なし	なし	8	
3	gender	数値	1	0	性別	{0, 女性}	なし	8	
4	rx	文字列	14	0	治療法	なし	なし	8	
5	node	数値	1	0	リンパ節転移	{0,なし}	なし	8	
6	time	数値	4	0	生存日数	なし	なし	8	
7	status	数値	1	0	状態	{0, 打ち切り	なし	8	
8									
9									
10									
11									T
10	<								>
データ ビ	ュー変数ピ)		_
		_			214 0200 01 A C	4 = B = Ar -z + 18			
				ll ll	BM SPSS Stati <mark>stic</mark> s ブロセッサ!	は使用可能です 🧗	¶ Unicod	de:ON	

Figure 5.2.2 Lesson 5-1.savファイルの変数ビュー

rxは文字列で定義されており、その幅は14に設定されています。文字型で定義されている変数をケースの追加で結合するためには2つのファイルにおいてそれぞれの幅が一致している必要があります。

POINT

ケースの追加を行う場合、変数名だけではなくデータの**型**が一致しているかどうかを確認する必要があります。また、文字型の変数の場合は**幅**が一致している必要があります。

次に、追加分のケースが含まれるLesson5-2.savファイルを開きます。

操作手順

- **4. ファイル**メニュー > **開く** >**データ**を選択します。
- 5. Lesson5-2.savを開きます。



Figure 5.2.3 Lesson 5-2.savファイル

このデータファイルには、**304**人分のケースが含まれています。変数の数は7個で先のデータファイルと同じですが、**状態**を記録している変数の名前が**event**になっていて相違があります。ケースの追加を行う場合、結合の対象となる変数名は一致している必要があります。

操作手順

6. **変数ビュー**を選択します。

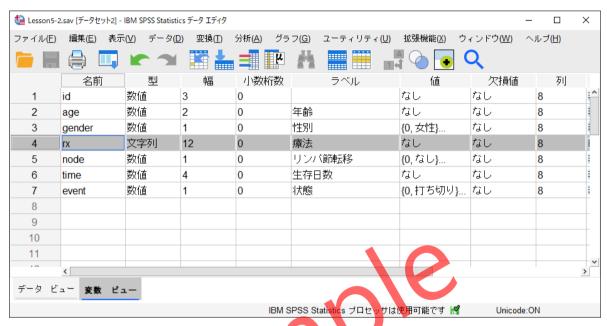


Figure 5.2.4 Lesson 5-2.savファイルの変数ビュー

rxは文字列で定義されており、その幅は12に設定されています。治療法1のファイルでの幅は14であったため、このままではrxは結合できません。幅の設定はファイルの結合を行う前に一致させておかなければなりません。

操作手順

7. rxの幅を14に変更します。

POINT

文字列の変数は、ケースの追加を行う前に2つのファイルで**幅**の設定を一致させておく必要があります。幅の設定は変数ビューで変更することができますが、入力されている文字数より短くしてしまうとデータ値が途中で切れてしまう危険性もあるため、設定を変更する場合は注意してください。

次に、2つのファイルのケースを追加して1つのデータセットを作成します。この例では、最初に開いたLesson5-1.savファイルの最終行の下に、Lesson5-2.savファイルのケースを追加します。

操作手順

- 8. Lesson5-1.savデータセットをアクティブ(画面の最前面)にします。
- 9. \vec{r} \vec{r}

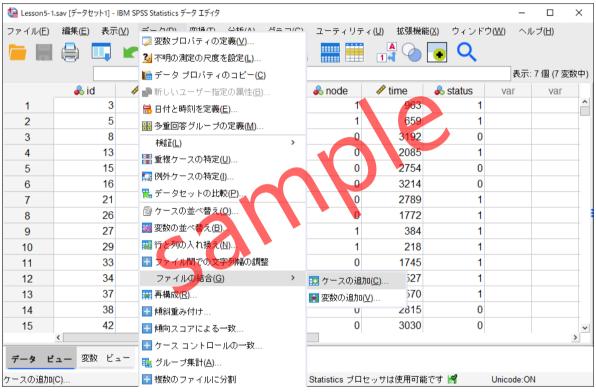


Figure 5.2.5 ファイルの結合のケースの追加サブメニュー

追加するケースを含むデータファイルを1つ選択します。すでに開いているデータセットのほか、IBM SPSS Statistics形式のデータファイルであれば、あらかじめ開かれていない外部データファイルを結合のファイルとして指定することができます。

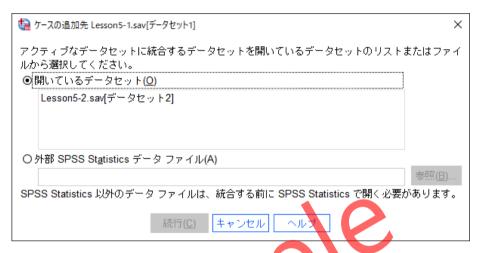


Figure 5.2.6 追加するケースを含むデータファイルの選択

なお、Microsoft Excelなど、他の形式のデータファイルを結合ファイルとして利用する場合は、 あらかじめIBM SPSS Statisticsとデータを読み込んでおかなければなりません。

POINT

Microsoft Excel形式のデータ(.xls、.xlsx)やカンマ区切りデータファイル(.csv)、データベースのテーブルは直接結合することができません。これらのデータを結合する場合は、事前にIBM SPSS Statisticsで読み込んでおく必要があります。

操作手順

- 10. 開いているデータセットのLesson5-2.savを選択します。
- 11. 続行ボタンをクリックします。

ケースの追加ダイアログボックスが表示されます。右側のリストが、結合後の新しいデータ セットに残る変数です。左側のリストは、一方のファイルにしか含まれないため対応のない変 数であり、このまま手順を進めると結合後のファイルから除外されます。



Figure 5.2.7 ケースの追加ダイアログボックス

この例では、対応のない変数としてアクティブデータセット(Lesson5-1.sav)に含まれる **status**と、データセット2(Lesson5-2.sav)に含まれる**event**が表示されています。

2つの変数は、変数名が一致していないだけですので、結合後の新しいデータセットに含める 設定を行います。この設定は、変数の名前を変更する方法と、変数をペアにする方法のいずれ かを利用することができます。

変数の名前を変更する場合は、対応のない変数の一覧で該当する変数を選択し、**名前の変更** ボタンを利用して変数名を一致させます。

変数をペアにする場合は、該当する変数を2つ選択し(CTRLキーを利用)、ペアボタンを利用 して、新しいアクティブなデータセットの変数に追加します。