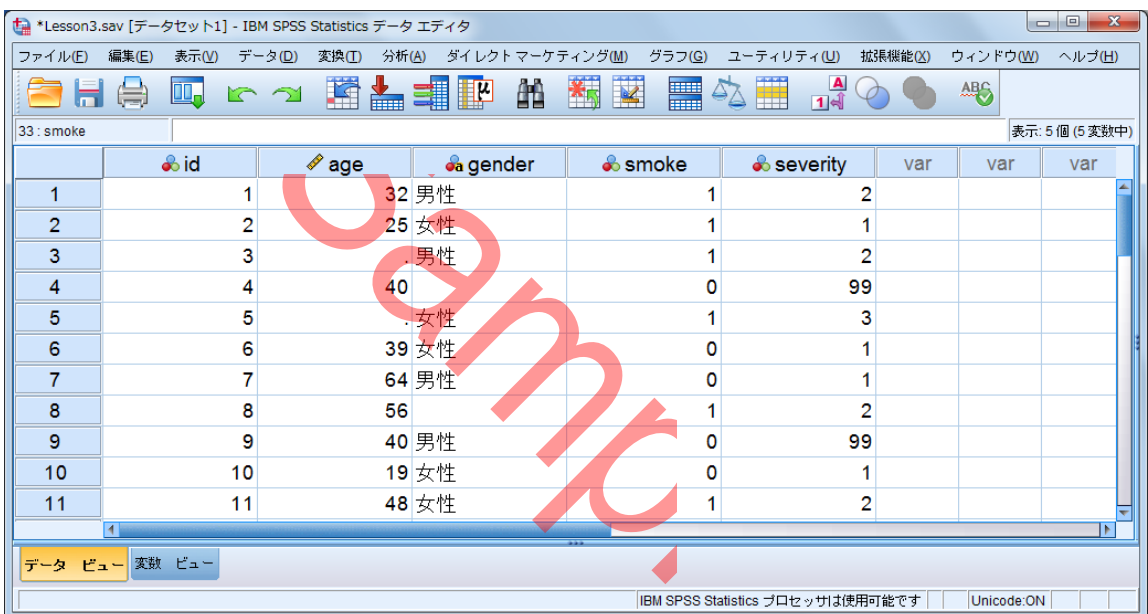


### §3.4.1 欠損値の概要

サンプルデータ **Lesson3.sav** について、欠損値の確認を行います。欠損値には、システム欠損値とユーザー欠損値の2種類があります。

**システム欠損値**は、数値型変数に含まれる値の未入力のセルを指します。IBM SPSS Statistics のデータビューではこの値をピリオド(.)として表示して、基本的に分析からは除外されることとなります。下記のデータファイルの例では、**age**にシステム欠損値が含まれていることがわかります。



	id	age	gender	smoke	severity	var	var	var
1	1	32	男性	1	2			
2	2	25	女性	1	1			
3	3		男性	1	2			
4	4	40		0	99			
5	5	.	女性	1	3			
6	6	39	女性	0	1			
7	7	64	男性	0	1			
8	8	56		1	2			
9	9	40	男性	0	99			
10	10	19	女性	0	1			
11	11	48	女性	1	2			

Figure3.4.1 Lesson3.savデータファイルの変数ビュー

なお、文字型変数の欠損は**空白**になりますが、セルには内部的に半角のスペースが入っており、システム欠損値にはなりません。上記のデータファイルの例では、**gender**に空白のセルが含まれていることがわかります。

#### POINT

数値型変数の未入力システム欠損値として認識されますが、文字型変数の未入力は空白として認識され、1つの有効な値として識別されます。

### §3.4.2 ユーザー欠損値の定義 —数値型変数—

サンプルデータ **Lesson3.sav** の数値型変数について、**ユーザー欠損値** の定義を行います。

ユーザー欠損値は、有効値を欠損値として定義するものです。 Lesson3.sav には、重症度をあらわす変数として **severity** が記録されており、値は、**1:軽症**、**2:中等症**、**3:重症** と定義されています。さらに、値 **99** が入力されており、これは、重症度が不明の場合に入力されるコードです。この値は、欠損値として入力されているため、分析に利用されないように欠損値として定義しておきます。

#### 操作手順

1. **変数ビュー** を選択します。
2. **severity** の **欠損値** のセルを選択します。

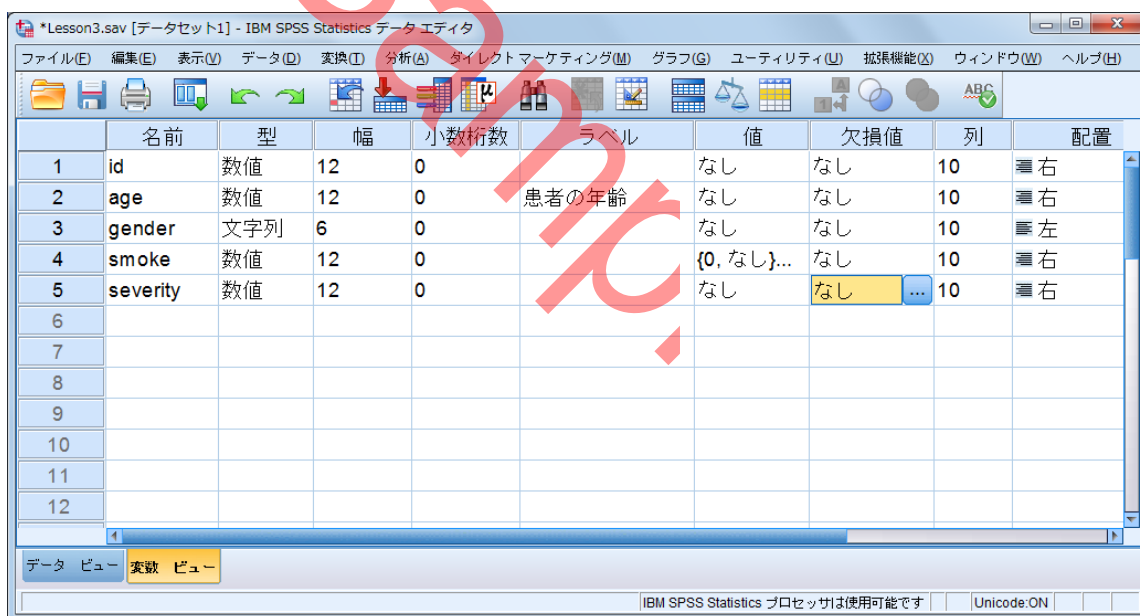


Figure3.4.2 変数ビューのユーザー欠損値の設定

#### 操作手順

3. 欠損値のセルに表示される…ボタンをクリックします。

## 操作手順

4. **個別の欠損値**ラジオボタンをクリックします。
5. **値**に**99**と入力します。

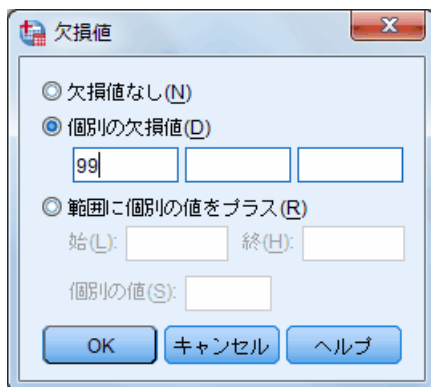


Figure3.4.3 欠損値ダイアログボックス

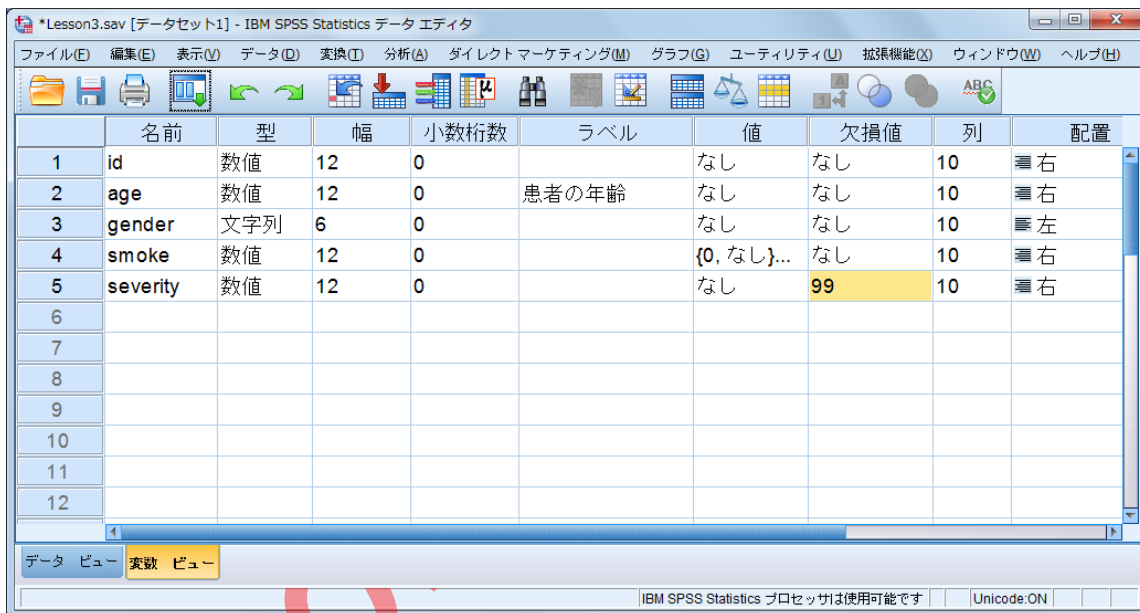
**個別の欠損値**は3つまで設定をすることができます。3つ以上のユーザー欠損値を指定する場合は、**範囲に個別の値をプラス**ラジオボタンを利用して設定を行うことができます。

## POINT

欠損値ダイアログボックスでの設定は、システム欠損値の設定ではなく、ユーザー欠損値の設定です。ここで指定される値がユーザー欠損値として認識されることになります。

## 操作手順

6. **OK**ボタンをクリックします。



	名前	型	幅	小数桁数	ラベル	値	欠損値	列	配置
1	id	数値	12	0		なし	なし	10	右
2	age	数値	12	0	患者の年齢	なし	なし	10	右
3	gender	文字列	6	0		なし	なし	10	左
4	smoke	数値	12	0		{0, なし}...	なし	10	右
5	severity	数値	12	0		なし	99	10	右
6									
7									
8									
9									
10									
11									
12									

Figure3.4.4 ユーザー欠損値が設定された変数ビュー

この設定によって、変数**severity**に含まれる**99**の値は欠損値として処理されることとなります。特定の値が入力されているため、システム欠損値と異なり値ラベルを設定することができます。たとえば、99に不明という値ラベルを設定することで、欠損値の意味がより分かりやすくなります。

### §3.4.3 ユーザー欠損値の定義 –文字型変数–

サンプルデータ **Lesson3.sav** の文字型変数について、**ユーザー欠損値** の定義を行います。

ユーザー欠損値は、有効値を欠損値として定義するものです。 Lesson3.sav には、性別をあらわす変数として **gender** が記録されており、値は文字列のデータとして **男性**、**女性** の2つの値が入力されています。しかし、値の未入力があり、文字型の変数の場合この値は**空白**として定義されています。これは、有効な値として認識されてしまいますので、空白をユーザー欠損値として定義しておきます。

#### 操作手順

1. **gender** の欠損値のセルを選択します。

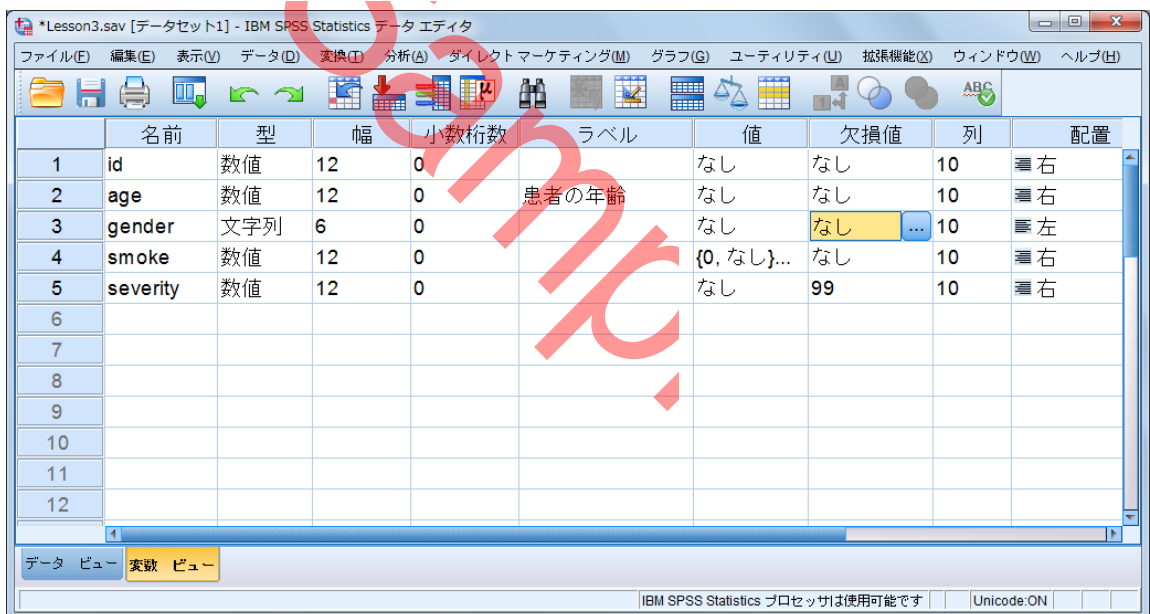


Figure3.4.5 変数ビューのユーザー欠損値の設定

#### 操作手順

2. ...ボタンをクリックします。

## 操作手順

3. **個別の欠損値**ラジオボタンをクリックします。
4. **値に半角スペース**を入力します。

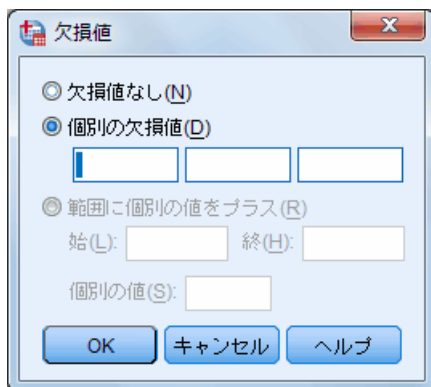


Figure3.4.6 欠損値ダイアログボックス

文字列の変数の空白をユーザー欠損値として定義するためには、個別の欠損値に**半角スペース**を定義します。

## 操作手順

5. **OK**ボタンをクリックします。

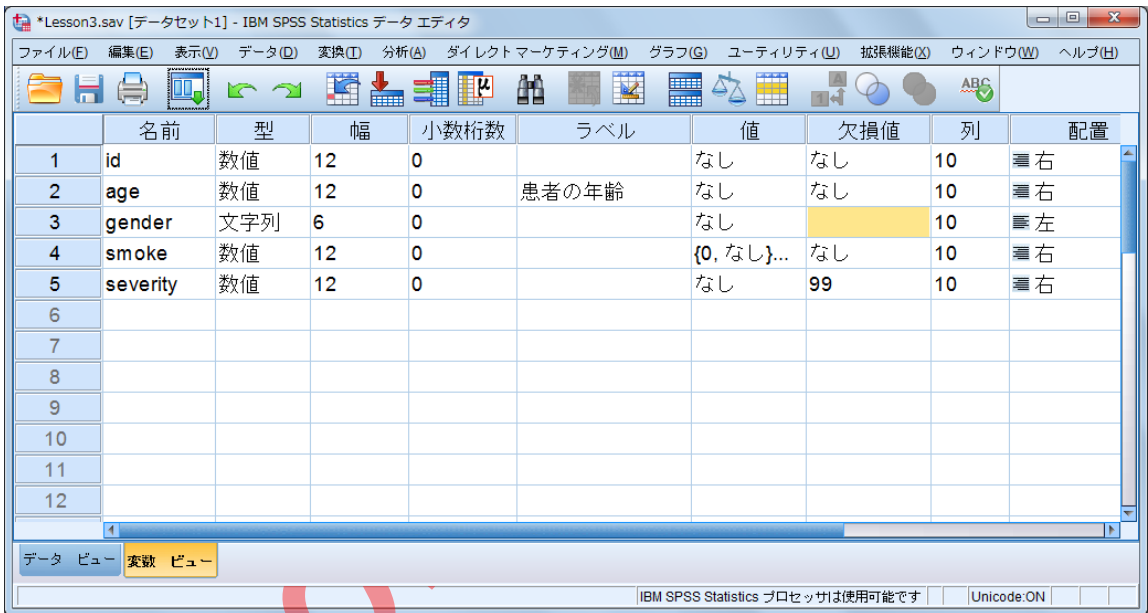


Figure3.4.7 文字型変数に対するユーザー欠損値の設定

この設定によって、変数**gender**に含まれる**空白**の値は欠損値として処理されることとなります。特定の値が入力されているため、システム欠損値と異なり値ラベルを設定することができます。たとえば、半角スペースに不明という値ラベルを設定することで、欠損値の意味がより分かりやすくなります。

### §3.4.4 設定されたユーザー欠損値の出力の確認

**severity**(重症度)と**gender**(性別)の2つの変数に設定したユーザー欠損値の出力を確認します。ここでは、データの数をカウントする度数分布表のメニューを利用します。

#### 操作手順

1. 分析メニュー > 記述統計 > 度数分布表を選択します。

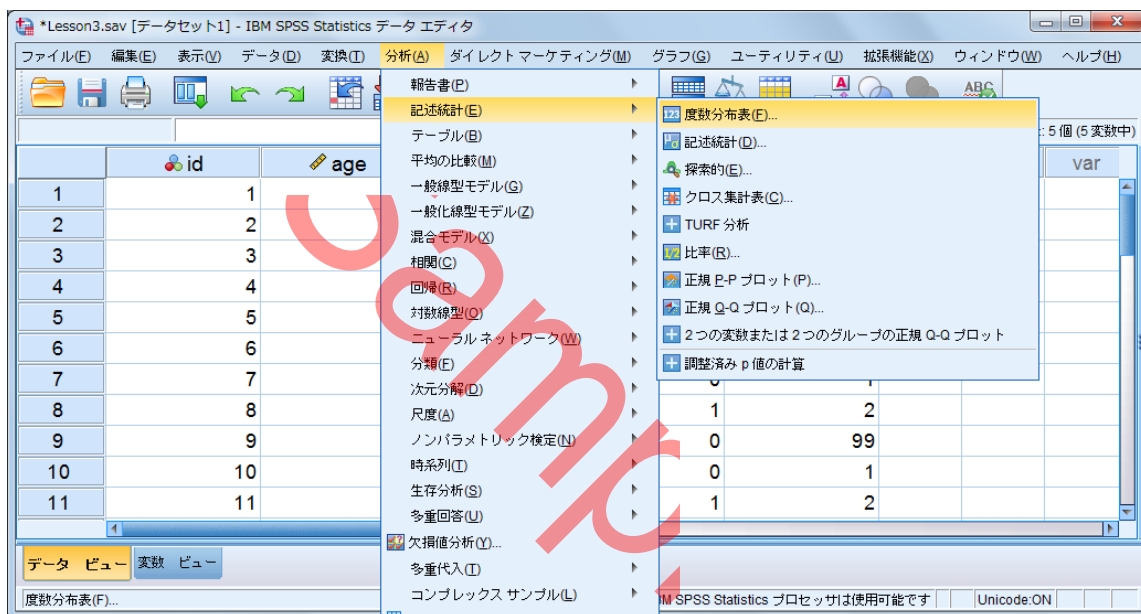


Figure3.4.8 度数分布表サブメニュー

#### POINT

統計解析ではデータの数を**度数 frequency**と呼びます。度数分布表はデータの数を数えて集計表を作成するための機能です。



操作手順

---

2. **gender**と**severity**を変数ボックスに移動します。



Figure3.4.9 変数の選択が完了した度数分布表ダイアログボックス

操作手順

---

3. **OK**ボタンをクリックします。

出力ビューアに度数分布表の結果が出力されます。まず、**統計**テーブルを確認します。

統計			
		gender	severity
度数	有効数	13	13
	欠損値	2	2

Figure3.4.10 度数分布表の統計テーブル

**gender**と**severity**のそれぞれに欠損値が2ケースずつ含まれていることがわかります。欠損値の度数には、システム欠損値とユーザー欠損値の両方が含まれます。

gender					
		度数	パーセント	有効パーセント	累積パーセント
有効数	女性	7	46.7	53.8	53.8
	男性	6	40.0	46.2	100.0
	合計	13	86.7	100.0	
欠損値		2	13.3		
合計		15	100.0		

severity					
		度数	パーセント	有効パーセント	累積パーセント
有効数	1	7	46.7	53.8	53.8
	2	4	26.7	30.8	84.6
	3	2	13.3	15.4	100.0
	合計	13	86.7	100.0	
欠損値		99	2	13.3	
合計		15	100.0		

Figure3.4.11 度数分布表の統計テーブル

度数分布表のテーブルを確認すると、**gender**に含まれる**空白**が欠損値として認識されていることがわかります。また、**severity**に含まれる**99**の値が欠損値としてカウントされています。

このように、ユーザー欠損値を定義することで変数に含まれる特定の値を欠損値として処理することができます。

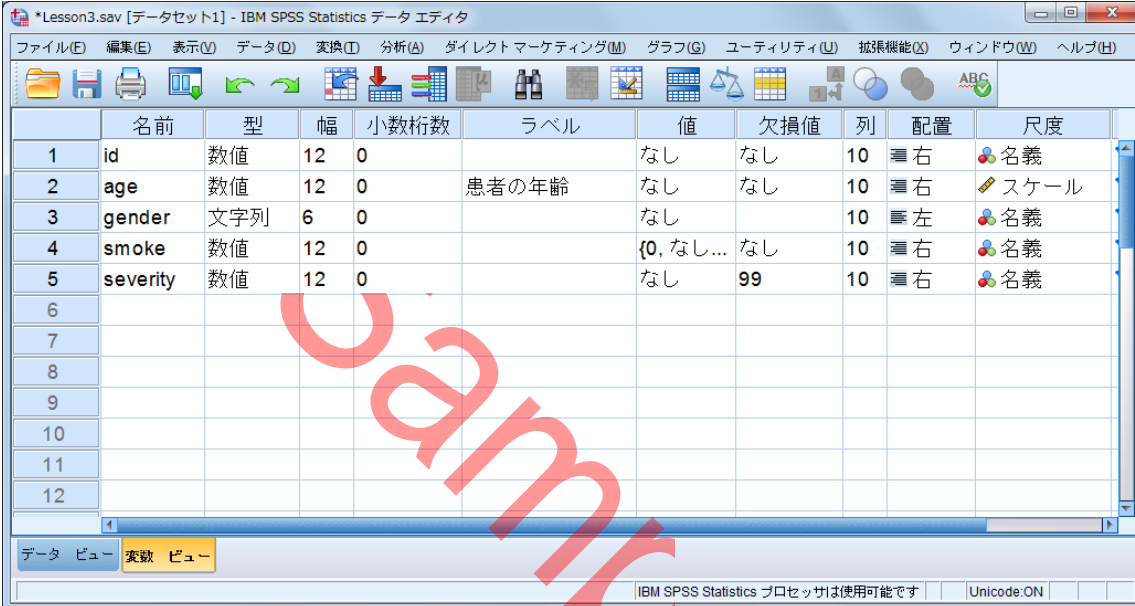
#### TIPS

ここでの度数分布表のテーブルルックが**Academic**に変更されています。テーブルのデフォルトテーブルルックは、**編集メニューのオプションサブメニュー**を選択し、**ピボットテーブルタブ**を使用すると変更することができます。

Cambridge

### §3.5.1 測定の尺度の定義

サンプルデータ **Lesson3.sav** について、測定の尺度を確認します。測定の尺度は、データの性質をあわらし、**名義** 尺度、**順序** 尺度、**スケール** の3種類が設定されます。これらの詳細は、Chapter4で確認します。



	名前	型	幅	小数桁数	ラベル	値	欠損値	列	配置	尺度
1	id	数値	12	0		なし	なし	10	右	名義
2	age	数値	12	0	患者の年齢	なし	なし	10	右	スケール
3	gender	文字列	6	0		なし		10	左	名義
4	smoke	数値	12	0		{0, なし...	なし	10	右	名義
5	severity	数値	12	0		なし	99	10	右	名義
6										
7										
8										
9										
10										
11										
12										

Figure3.5.1 変数ビューの尺度の設定

分析を実行する場合やグラフを作成する場合など、各変数の尺度を設定しておくとういでしょう。

#### POINT

測定の尺度のアイコンを変数の内容や分析の目的に応じて適切に設定しておくこと、分析やグラフ作成の際に分かりやすくなります。