

Chapter 1 IBM SPSS Statisticsの起動とデータの入力	1-1
1.1.1 IBM SPSS Statisticsの起動.....	1-2
1.2.1 新規データの入力.....	1-4
1.3.1 データファイルの保存.....	1-9
Chapter 2 データの読み込み	2-1
2.1.1 Microsoft Excel形式のデータファイルの読み込み.....	2-2
2.2.1 CSV形式のデータファイルの読み込み.....	2-8
2.2.3 データベースの読み込み(概要).....	2-18
Chapter 3 変数の定義	3-1
3.1.1 変数の定義.....	3-2
3.1.2 データファイルの確認.....	3-3
3.2.1 変数ラベルの定義.....	3-4
3.2.2 変数ラベルの確認.....	3-6
3.3.1 値ラベルの定義.....	3-8
3.3.2 値ラベルの確認.....	3-11
3.4.1 欠損値の確認.....	3-13
3.4.2 ユーザー欠損値の定義 —数値型変数—.....	3-14
3.4.3 ユーザー欠損値の定義 —文字型変数—.....	3-17
3.4.4 設定されたユーザー欠損値の出力の確認.....	3-20
3.5.1 測定の尺度の定義.....	3-24
Chapter 4 1変量の要約	4-1
4.1.1 測定の尺度.....	4-2
4.1.2 要約統計量.....	4-4
4.1.3 データファイルの確認.....	4-6
4.1.4 オプションの設定.....	4-7
4.2.1 名義尺度の要約 —分析の実行—.....	4-10
4.2.2 名義尺度の要約 —出力の確認—.....	4-13
4.3.1 順序尺度の要約 —分析の実行—.....	4-18
4.3.2 順序尺度の要約 —出力の確認—.....	4-21
4.4.1 スケールの要約 —分析の実行—.....	4-23
4.4.2 スケールの要約 —出力の確認—.....	4-27
4.5.1 出力ファイルの保存.....	4-29

Chapter 5 グラフ作成と編集	5-1
5.1.1 グラフ作成の概要.....	5-2
5.1.2 データファイルの確認.....	5-3
5.2.1 積み上げ棒グラフの作成.....	5-4
5.2.2 積み上げ棒グラフの編集.....	5-11
5.3.1 パネル別の円グラフの作成.....	5-22
5.3.2 パネル別の円グラフの編集.....	5-26
Chapter 6 テーブルの編集	6-1
6.1.1 テーブルの編集の概要.....	6-2
6.1.2 出力ファイルの確認.....	6-3
6.2.1 テーブルの編集.....	6-4
6.2.2 データ値ラベルの移動.....	6-6
6.2.3 セルの書式設定.....	6-7
6.3.1 テーブルルックの使用.....	6-11
6.3.2 デフォルトテーブルルックの設定.....	6-13
6.4.1 ピボットトレイによる行変数と列変数の入れ替え.....	6-15
6.4.2 ピボットトレイによる層の配置.....	6-19
6.5.1 他のアプリケーションへのコピーと貼り付け.....	6-22
6.5.2 出力のエクスポート.....	6-26
Chapter 7 多重回答の入力と集計	7-1
7.1.1 多重回答とは.....	7-2
7.1.2 カテゴリ形式での入力.....	7-3
7.1.3 2分形式での入力.....	7-4
7.1.4 データファイルの確認.....	7-5
7.2.1 多重回答グループの定義 – テーブル集計の場合 –	7-6
7.2.2 多重回答の度数分布表 – 分析の実行 –	7-11
7.2.3 多重回答の度数分布表 – 出力の確認 –	7-13
7.2.4 多重回答のクロス集計表 – 分析の実行 –	7-15
7.2.5 多重回答のクロス集計表 – 出力の確認 –	7-19
7.3.1 多重回答グループの定義 – グラフ作成の場合 –	7-21
7.3.2 多重回答の棒グラフ.....	7-25

sample

本書では、IBM SPSS Statistics 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。

sample

2

データの読み込み

IBM SPSS Statisticsは、Microsoft Excel形式やCSV形式をはじめとする、さまざまなデータソースを読み込んで、分析やグラフ作成を行うことができます。分析のためのデータをIBM SPSS Statisticsに直接入力しておくこともできますが、ExcelやCSV形式で持っていれば、IBM SPSS Statistics以外のアプリケーションでも利用することができ、汎用性が高くなります。ここでは、Microsoft Excel形式とCSV形式のデータファイルをIBM SPSS Statisticsに読み込む手順と注意点を確認します。

Contents

- Microsoft Excel形式のデータファイルの読み込み
- CSV形式のデータファイルの読み込み
- データベースの読み込み(概要)

Keyword

Microsoft Excel / カンマ区切り / CSV / タブ区切り / データベースの読み込み / ODBC /

§2.1.1 Microsoft Excel形式のデータファイルの読み込み

Microsoft Excelは、さまざまな分野で広く利用されている表計算ソフトです。IBM SPSS Statisticsは、Excel形式(.xls、.xlsx)のデータファイルを読み込んで集計や分析に利用することができます。

ここでは、以下のExcel形式のデータファイル**birthwt.xlsx**をIBM SPSS Statisticsに読み込みます。このデータファイルは、低体重児出生に関する母親の特徴と体重を記録したデータです。変数は11個、189行のデータが含まれています。

	A	B	C	D	E	F	G	H	I	J	K	L
1	id	age	lwt	race	smoke	ptl	ht	ui	lty	bwt	low	
2	180	17	120	その他	あり	なし	なし	なし	なし	3572	2500g以上	
3	15	25	85	その他	なし	なし	なし	あり	なし	1474	2500g未満	
4	89	18	107	白人	あり	なし	なし	あり	なし	2600	2500g以上	
5	93	17	103	その他	なし	なし	なし	なし	1回	2637	2500g以上	
6	100	18	100	白人	あり	なし	なし	なし	なし	2769	2500g以上	
7	101	18	100	白人	あり	なし	なし	なし	なし	2769	2500g以上	
8	102	15	98	黒人	なし	なし	なし	なし	なし	2778	2500g以上	
9	113	17	122	白人	あり	なし	なし	なし	なし	2906	2500g以上	
10	116	17	113	黒人	なし	なし	なし	なし	1回	2920	2500g以上	
11	117	17	113	黒人	なし	なし	なし	なし	1回	2920	2500g以上	
12	132	18	90	白人	あり	なし	なし	あり	なし	3076	2500g以上	
13	133	18	90	白人	あり	なし	なし	あり	なし	3076	2500g以上	
14	143	16	110	その他	なし	なし	なし	なし	なし	3175	2500g以上	
15	147	17	119	その他	なし	なし	なし	なし	なし	3225	2500g以上	

Figure2.1.1 Microsoft Excelで開いたbirthwt.xlsxファイル

ExcelのデータをIBM SPSS Statisticsに正しく読み込むためには、いくつかの注意点があります。まず、データの入力方法はIBM SPSS Statisticsと同じ形式にしておく必要があります。つまり、列に変数、行にケースを入力します。また、基本的にExcelの1行目には変数名を入力しておき、分析用のデータ値は2行目以降に入力しておきます。1列1変数、1行1ケースの形式になっていない場合、IBM SPSS Statisticsでは正しくデータを読み込むことができません。

また、このExcelファイルには、シートが1枚だけ含まれており、birthwtと名前が定義されています。

POINT

Excelには、列に変数、行にケースを入力しておきます。この入力方法は分析ソフトウェアによらず一般的な入力方法です。

操作手順

1. **ファイル**メニュー > **開く** > **データ**を選択します。

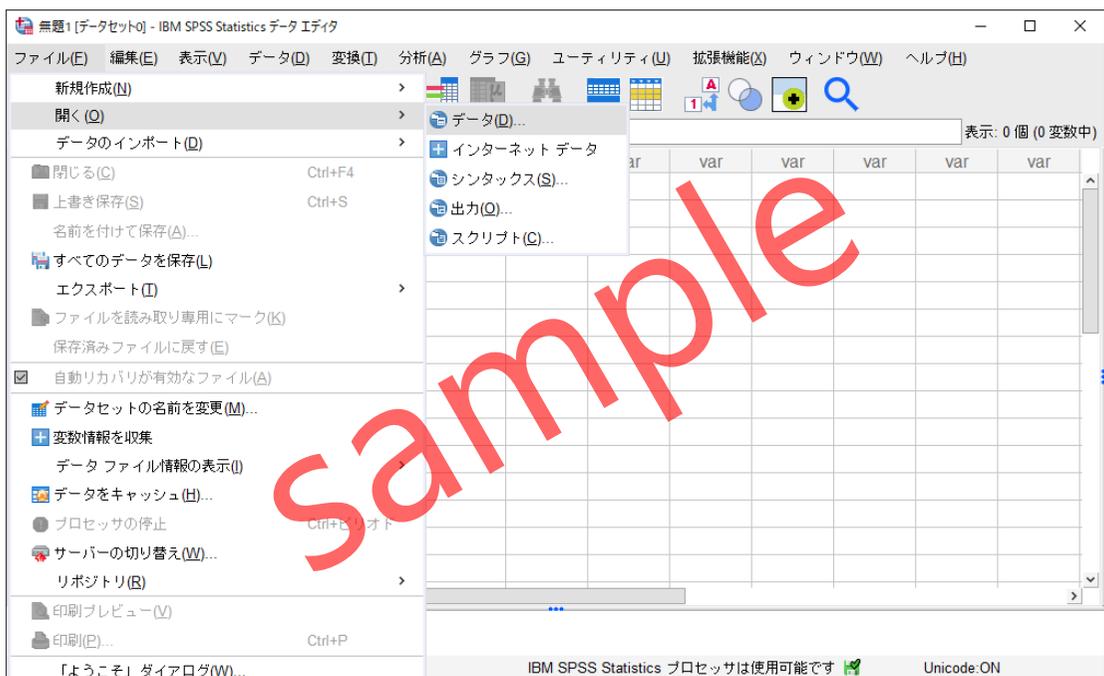


Figure2.1.2 IBM SPSS Statisticsのデータファイルを開くメニュー

操作手順

2. ファイルの種類から**Excel(*.xls,*.xlsx,*.xlsm)**を選択します。
3. ファイルの一覧から**birthwt.xlsx**を選択します。

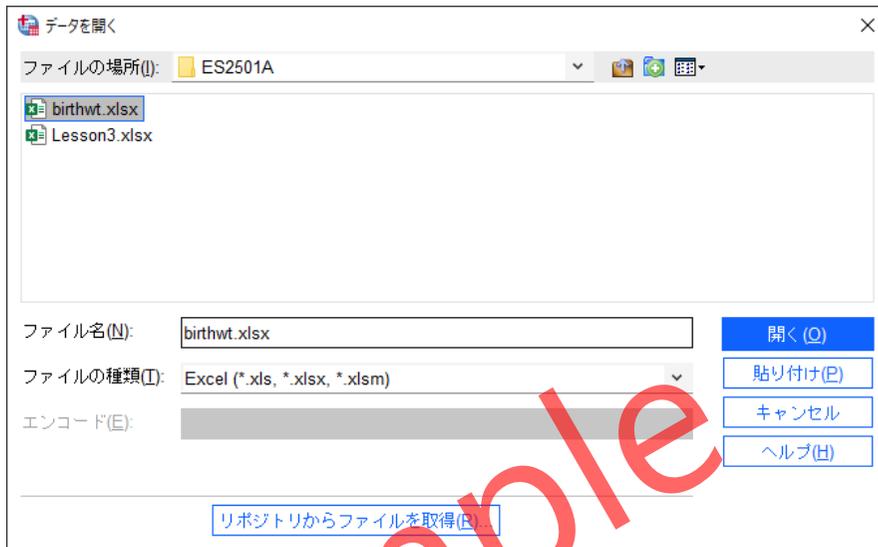


Figure2.1.3 データを開くダイアログボックス

IBM SPSS Statistics形式以外のデータファイルを読み込むためには、ファイルの種類を変更する必要があります。ここでは、Excel形式のデータファイルを読み込みます。

操作手順

4. **開く** ボタンをクリックします。

Excel形式のデータファイルを読み込む場合、以下のダイアログボックスでオプション設定を指定する必要があります。

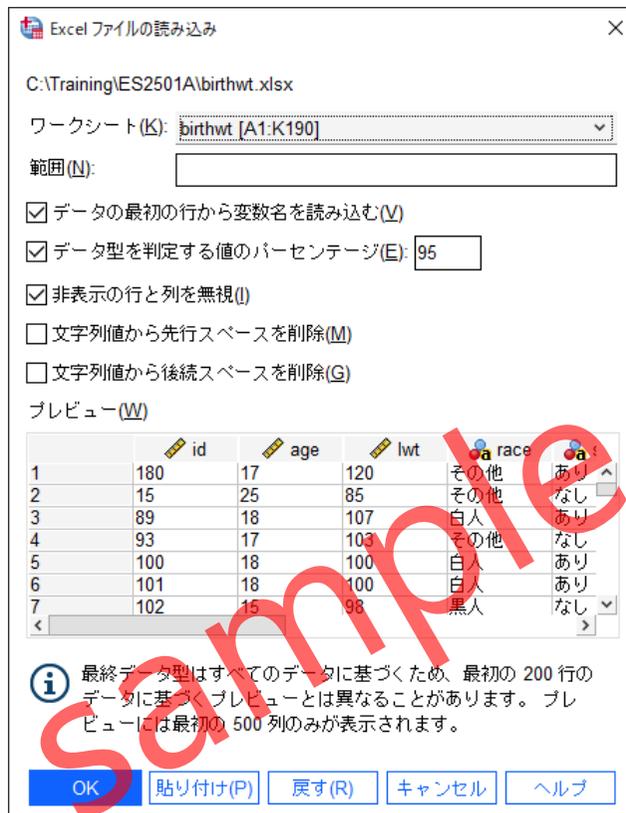


Figure2.1.4 Excelファイルの読み込みダイアログボックス

データの最初の行から変数名を読み込むチェックボックスを有効にすると、Excelデータファイルの1行目をIBM SPSS Statisticsの変数名として読み込むことができます。通常、Excelデータファイルの1行目には変数名が入力されていることが多いため、このオプションはデフォルトで有効になっています。

ワークシートドロップダウンリストでは、指定したExcelデータファイルに含まれるシートが表示されます。一度に読み込むことができるシートは1枚だけです。どのシートからデータを読み込むかを指定する必要がありますが、この例ではExcelデータファイルに含まれるシートが1枚のみであるため、このままの設定で次に進みます。

POINT

IBM SPSS Statisticsで読み込むことができるシートは1枚のみです。分析に必要なデータが複数のシートに入力されている場合は、事前にExcel側で1枚のシートに統合しておくか、IBM SPSS Statisticsで各シートを個別に読み込み、ファイル結合の機能を利用して統合する必要があります。

操作手順

4. **OK**ボタンをクリックします。

	id	age	lwt	race	smoke	pti	ht	ui
1	180	17	120	その他	あり	なし	なし	なし
2	15	25	85	その他	なし	なし	なし	あり
3	89	18	107	白人	あり	なし	なし	あり
4	93	17	103	その他	なし	なし	なし	なし
5	100	18	100	白人	あり	なし	なし	なし
6	101	18	100	白人	あり	なし	なし	なし
7	102	15	98	黒人	なし	なし	なし	なし
8	113	17	122	白人	あり	なし	なし	なし
9	116	17	113	黒人	なし	なし	なし	なし
10	117	17	113	黒人	なし	なし	なし	なし
11	132	18	90	白人	あり	なし	なし	あり
12	133	18	90	白人	あり	なし	なし	あり

Figure2.1.5 Excel形式のデータファイルから読み込んだデータ

Excel形式のデータファイルがIBM SPSS Statisticsに読み込まれました。Excel側で数値として入力されている変数は数値型として読み込まれ、文字として入力されている変数は文字型として読み込まれます。基本的に、数値型の変数は右寄せ、文字型の変数は左寄せで表示されます。

必要なデータが読み込まれているため、このまま分析に進むこともできますが、より効率的に分析を行うことができるよう変数のラベルの設定などを行うとよいでしょう。変数定義の方法は次のレッスンで確認します。

TIPS

Excel側で、数式や関数によってセル内容が定義されている場合、IBM SPSS Statisticsには計算結果のみが読み込まれてきます。

最後に、読み込んだデータをIBM SPSS Statistics形式として保存しておきます。

操作手順

5. **ファイル**メニュー > **名前を付けて保存**を選択します。
6. ファイル名に**Lesson2**と入力します。
7. **保存**ボタンをクリックします。

POINT

IBM SPSS Statisticsのデータファイルは、.sav形式で保存されます。このファイルを開くためには、IBM SPSS Statisticsがインストールされている必要があります。

sample

§2.3.1 データベースの読み込み(概要)

IBM SPSS Statisticsでは、**ODBC**接続により、データベースのテーブルに含まれるデータを読み込むことができます。ODBCは汎用的なデータベースドライバであるため、ODBCに対応するデータベースであれば、IBM SPSS Statisticsに読み込むことが可能です。データベースの読み込みを行うためには、IBM SPSS Statisticsでの操作の前に、ODBCの定義を行っておく必要があります。

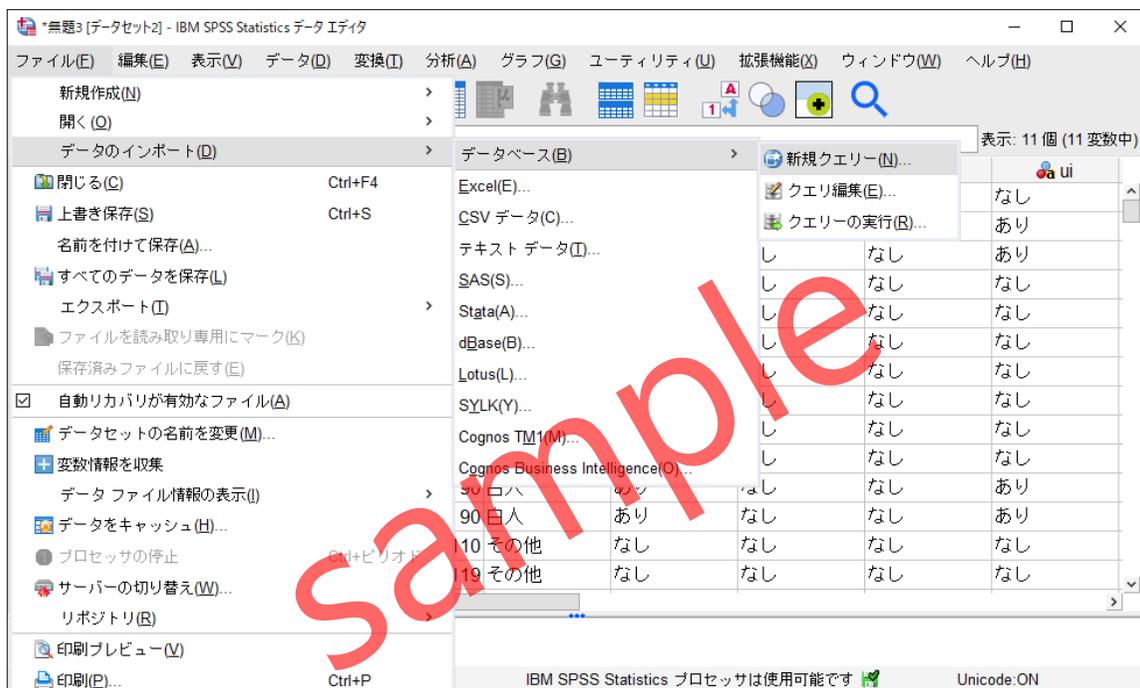


Figure2.3.1 データベースのテーブルを読み込むためのメニュー

データベースの読み込みのためには、ファイルメニューの**データのインポート**メニューを利用して、**新規クエリ**を実行します。テーブルを選択して読み込むだけでなく、直接SQLを実行することも可能です。これらのデータベースの読み込みは、このコースでは具体的な手順は行いません。

TIPS

データベースのテーブルに含まれるデータを、CSV形式のファイルとして出力しておいて、IBM SPSS Statisticsのテキストデータの読み込み手順を利用する方法もあります。

7

多重回答の入力と集計

1つの設問(項目)につき、複数の結果が記録されるデータを多重回答データと呼びます。アンケート調査によくみられ、回答者によって回答の個数が異なるのが特徴です。多重回答の場合、1つの項目に複数の結果を持つこととなりますので、通常のデータ入力方法では対応できません。ここでは、多重回答データの入力方法と分析方法を確認します。

Contents

- 多重回答
- カテゴリ形式の入力
- 2分形式の入力
- 多重回答の度数分布表
- 多重回答のクロス集計表
- 多重回答の棒グラフ

Keyword

多重回答 / マルチアンサー / MA項目 / カテゴリ形式 / 2分形式 / 集計 / 棒グラフ /

§7.1.1 多重回答とは

多重回答 multi answer とは、1つの設問に複数の回答が得られるデータです。マルチアンサーやMA項目とも呼ばれます。例えば、以下の設問は多重回答のデータになります。

[設問] 現在、治療中の病気があれば教えてください。（複数回答可）

[1] 高血圧	[2] 糖尿病	[3] 心臓病
[4] 脳梗塞	[5] 胃潰瘍	

Figure7.1.1 多重回答の設問

該当するものが複数ある場合、設問は1つにもかかわらずその回答は複数になります。そこで、多重回答の設問では、通常の変数とは少し異なる入力方法が必要になります。

例えば、以下のように複数の回答をカンマ区切りで入力してしまうと、統計ソフトは、セル内のデータ値を1つのデータと認識するため、適切な集計や分析を行うことができません。

id	治療中の病気
1	1,2,5
2	2,3
3	
4	5
5	1,3

Figure7.1.2 多重回答の間違ったデータ入力の場合

多重回答のデータの入力方法として、**カテゴリ形式**または**2分形式**のいずれかを使用します。

TIPS

実際の集計や解析業務では、Figure7.1.2のように多重回答のデータをカンマ区切りで入力することはあります。ただし、このままでは分析することはできないため、分析の前にデータ加工を行い、適切な形式にする必要があります。手動での加工は煩雑になるため、通常はシンタックスと呼ばれるコマンドを利用することになります。

§7.1.2 カテゴリ形式での入力

多重回答データの入力方法の1つが、**カテゴリ形式**です。この方法は、多重回答の設問によって選ばれる選択肢の数と同数の変数を用意して、回答者によって選択されたカテゴリ値を直接セルに入力します。例えば、Figure7.1.2の多重回答のデータは、以下のように入力します。

id	治療中の病気1	治療中の病気2	治療中の病気3	治療中の病気4	治療中の病気5
1	1	2	5	.	.
2	2	3	.	.	.
3
4	5
5	1	3	.	.	.

Figure7.1.3 多重回答の入力方法(カテゴリ形式)

カテゴリ形式の入力では、変数1つに対して1つのカテゴリ値が直接入力されることとなります。この入力方法であれば、1変数に1データ値の形式となり、集計や解析を行うことができます。

この例では、idが1の患者は、1と2と5が選択されており、高血圧と糖尿病と胃潰瘍の治療中であることが判別できます。また、idが3の患者は何も入力されていないため欠損値となり、治療中の病気がないことが分かります。

このように、1つの変数に1つの回答の形式になるように入力しておき、実際の集計や分析の際には、これらの5つの変数を1つの**多重回答グループ**として定義します。

TIPS

カテゴリ形式は、選択肢が選ばれていない場合に**欠損値**が発生することとなります。また、すべての変数を利用しないと統合された情報を提供しないため、カテゴリ形式で入力する特別な理由がない限り、2分形式での入力を検討するのが一般的です。

§7.1.3 2分形式での入力

カテゴリ形式に対して、もう1つの多重回答データの入力方法が**2分形式**です。この方法は、多重回答の設問において選択される選択肢の数と同数の変数を用意して、回答者によってそれぞれの設問に該当するか該当しないかを入力します。例えば、Figure7.1.2の多重回答のデータは、以下のように入力します。

id	高血圧	糖尿病	心臓病	脳梗塞	胃潰瘍
1	1	1	0	0	1
2	0	1	1	0	0
3	0	0	0	0	0
4	0	0	0	0	1
5	1	0	1	0	0

Figure7.1.4 多重回答の入力方法(2分形式)

選択肢と同数の変数を用意するところは、カテゴリ形式の場合と同じですが、2分形式の場合はカテゴリ値ではなく、**選択の有無**を入力します。多くの場合、選択されたら**1**、選択されていないならば**0**が入力されます。

このように、1つの変数に1つの回答の形式になるように入力しておき、実際の集計や分析の際には、これらの5つの変数を1つの**多重回答グループ**として定義します。

TIPS

2分形式で入力された多重回答は、**2値変数**として汎用性が高くなります。カテゴリとして集計することもできますし、量的変数として解析することもできます。特に、医療統計でよく利用されるロジスティック回帰分析などでも使用可能となります。そこで、多重回答の入力方法としては、カテゴリ形式より2分形式が推奨されます。

§7.1.4 データファイルの確認

ここでは、IBM SPSS Statistics形式で保存されているデータファイル**Lesson7.sav**を利用します。このデータファイルには、多重回答の設問が1つ含まれています。

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. **Lesson7.sav**を開きます。

	id	age	gender	d_1	d_2	d_3	d_4	d_5
1	1	27	1	0	1	0	0	0
2	2	37	1	0	1	0	0	1
3	3	34	0	0	0	0	0	0
4	4	33	1	1	0	0	0	0
5	5	15	0	1	0	1	0	0
6	6	16	0	0	1	0	0	1
7	7	34	1	0	0	0	0	0
8	8	57	0	1	1	0	0	0
9	9	35	0	0	0	0	0	0
10	10	34	0	1	0	0	0	0
11	11	33	1	0	0	0	0	0
12	12	24	0	0	0	0	0	0

Figure7.1.5 Lesson7.savデータファイル

このデータファイルには、**id**(患者番号)、**age**(患者の年齢)、**gender**(患者の性別)のほか、多重回答の設問が1つ含まれています。

多重回答の設問は、**d_1**(高血圧)、**d_2**(糖尿病)、**d_3**(心臓病)、**d_4**(脳梗塞)、**d_5**(胃潰瘍)の5つの変数に記録されており、データの入力方法は**2分形式**です。

データの入力は完了しているため、はじめに多重回答のグループ定義から開始します。具体的には、**d_1~d_5**までの変数を1つの回答グループとして設定します。

§7.3.1 多重回答のグループ定義 – グラフ作成の場合 –

はじめに、多重回答による棒グラフを作成するための、変数グループの定義を行います。IBM SPSS Custom Tablesオプションがインストールされていない場合、テーブル集計を行う場合は**多重回答**メニューを利用しますが、グラフ作成を行う場合は**データ**メニューを利用します。この設定で作成される多重回答グループは、**図表ビルダー**の画面で利用することができます。

操作手順

1. **データ**メニュー > **多重回答グループの定義**を選択します。

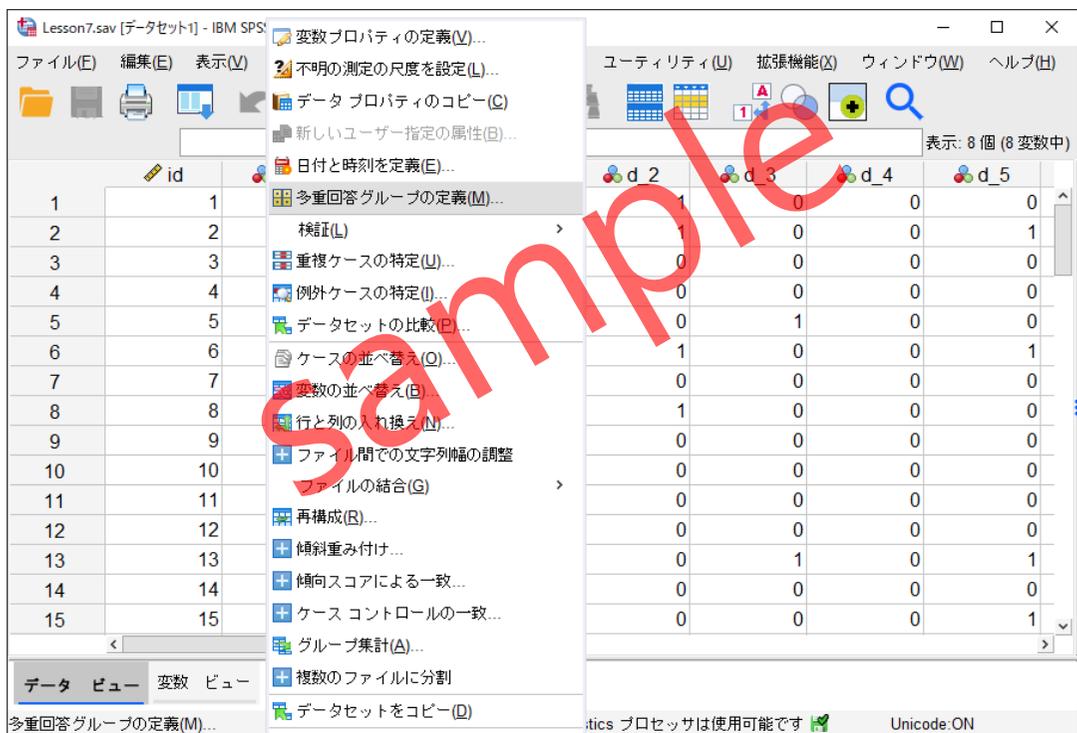


Figure7.3.1 データメニューの多重回答グループの定義

多重回答グループを定義ダイアログボックスが表示されます。この画面で、多重回答の変数のグループ定義と変数のコード化方法の指定を行い、グループに名前をつけます。

操作手順

2. 変数の一覧から、**d_1、d_2、d_3、d_4、d_5**を**CTRL**キーを利用して複数選択します。
3. **変数グループ内の変数**に移動します。
4. 変数のコード化様式に**2分**を選択します。
5. **集計値**に**1**と入力します。



Figure 7.3.2 多重回答グループを定義(変数の選択と入力形式の指定)

ここで選択した5つの変数は、2分形式で入力されているため、コード化様式として**2分**を指定します。変数は、選択された場合に1が入力されているため、集計値に**1**を指定します。

操作手順

6. **グループ名**にd_allと入力します。
7. **グループラベル**に**治療中の病気**と入力します。



Figure7.3.3 多重回答グループを定義(名前とラベルの指定)

操作手順

8. **追加**ボタンをクリックします。

POINT

追加ボタンをクリックしないと、設定内容が確定しません。必ず追加ボタンや変更ボタンで操作内容をダイアログに反映させるようにしましょう。

追加された多重回答グループには自動的に\$マークが付けられます。この変数は、データファイルに列として追加されることはなく、データファイルに定義情報として含まれています。



Figure7.3.4 多重回答グループの定義

一度、定義した多重回答グループの定義を変更する場合は、多重回答グループの一覧で対象を選択し、内容を変更してから変更ボタンをクリックして確定します。このメニューで作成した多重回答グループは、グラフ作成のメニューである**図表ビルダー**の画面で利用することができます。

操作手順

9. **OK**ボタンをクリックします。

§7.3.2 多重回答の棒グラフ

データメニューで定義した多重回答グループは、図表ビルダーで利用することができます。ここでは、多重回答グループである治療中の病気の棒グラフを作成します。

操作手順

1. **グラフメニュー** > **図表ビルダー**を選択します。

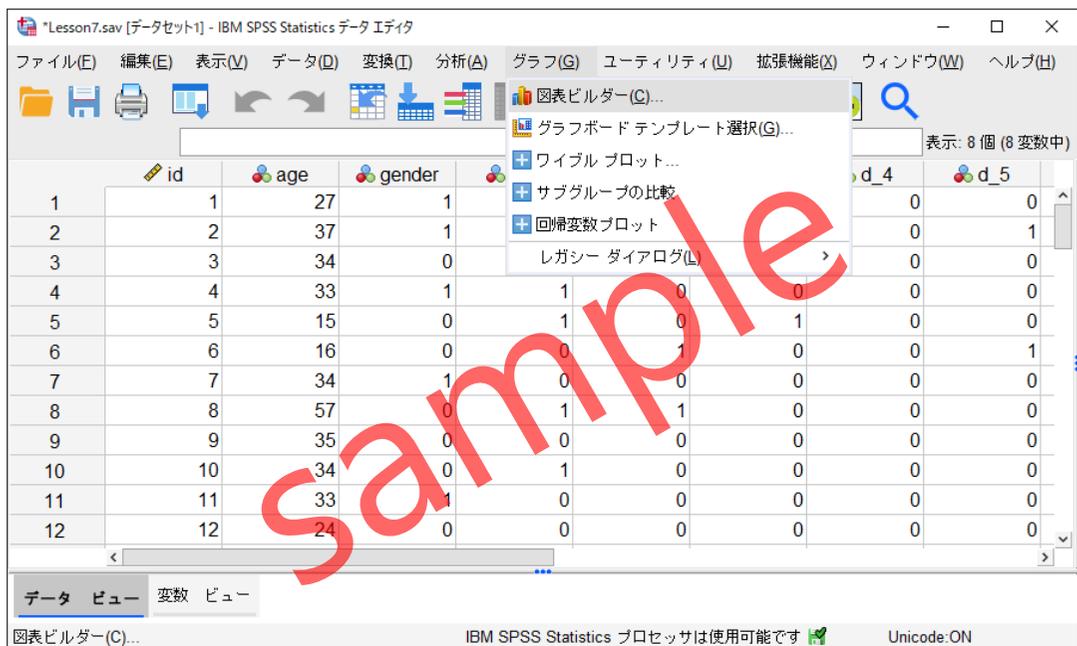


Figure7.3.5 図表ビルダーメニュー

操作手順

2. 確認メッセージの**OK**ボタンをクリックします。

図表ビルダーが表示されます。変数のリストを見るとデータメニューで作成した多重回答グループ \$d_all が含まれていることが分かります。ここでは、多重回答グループを利用して単純棒グラフを作成します。

操作手順

3. 図表ビルダーの**ギャラリー**タブの以下から選択から**棒グラフ**を選択します。
4. **単純棒グラフ**のアイコン(上段左端)をプレビュー領域にドラッグします。



Figure7.3.6 図表ビルダーメニュー

図表ビルダーのプレビュー表示が単純棒グラフに切り替わります。また、同時にグラフの詳細設定を行うための要素のプロパティウィンドウが開きます。ここでは、X軸に多重回答グループである**\$d_all**を指定します。

操作手順

- 変数から**\$d_all**を選択して**X軸?**のボックスに移動します。



Figure7.3.7 多重回答グループ\$d_allをX軸に配置した図表ビルダー

Y軸は自動的に**度数**の表示に切り替わります。この例ではこのまま実行しますが、Y軸の表示内容を変更したい場合は、要素のプロパティを利用します。

操作手順

6. **OK**ボタンをクリックします。

多重回答グループの棒グラフが出力されます。集計テーブルと異なり、棒グラフは視覚的に表示されるため、カテゴリ間の比較が行いやすくなります。例えば、治療中の病気としては、高血圧がもっとも多いことが分かります。

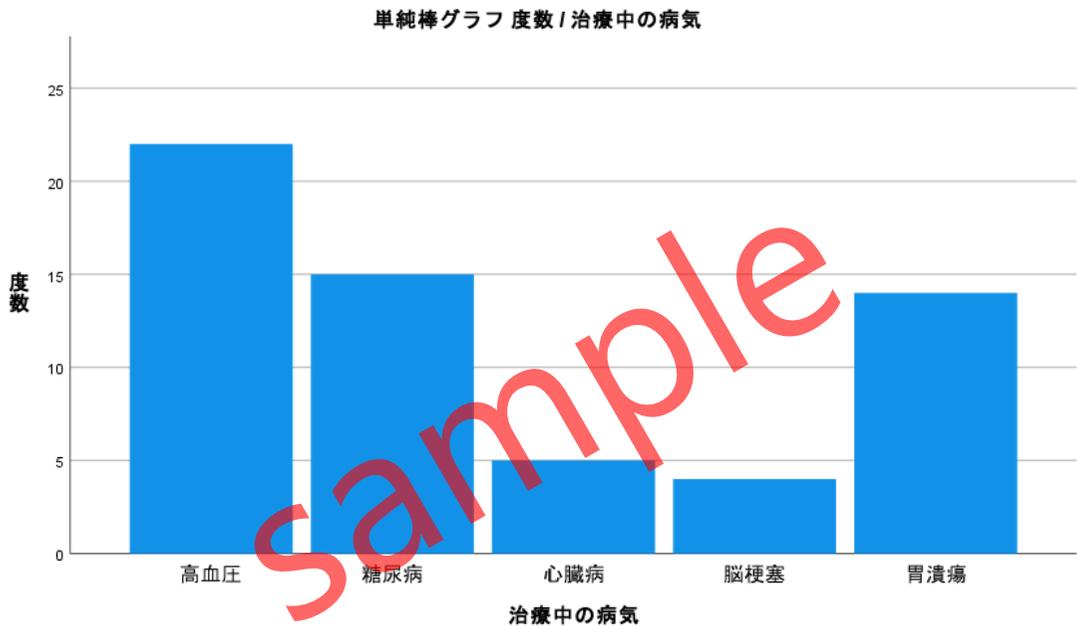


Figure7.3.8 多重回答グループ\$d_allの棒グラフ

Lesson5で確認したとおり、グラフをダブルクリックすることで**図表エディタ**を表示させ、グラフのデザインやレイアウト、カテゴリの並び順の変更や、データ値ラベルの表示などを行うことができます。