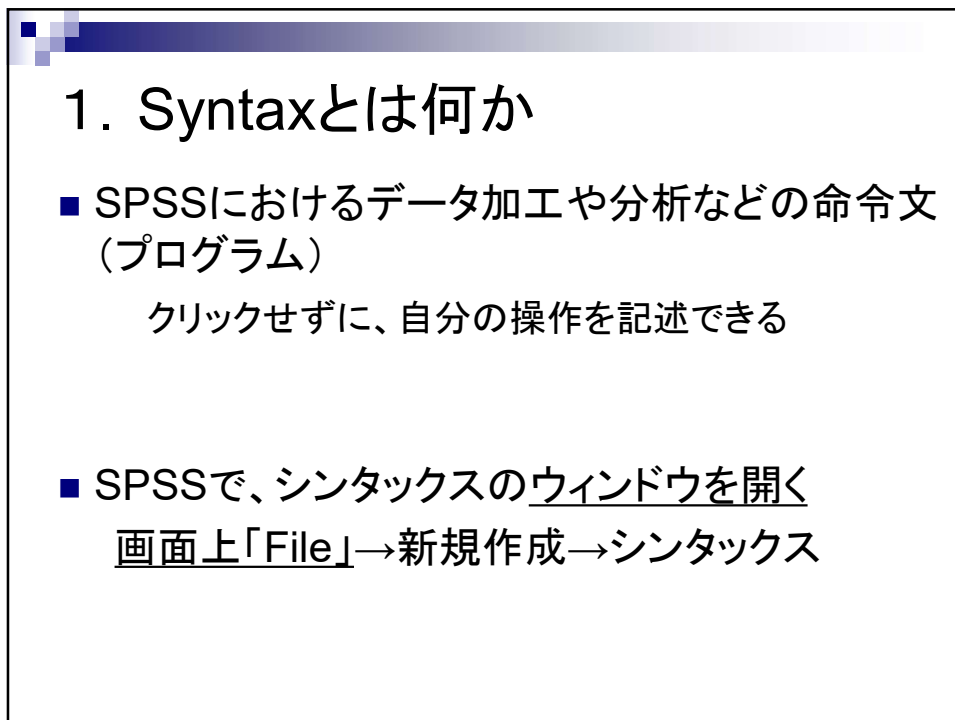
The slide features a decorative background of overlapping squares in various shades of blue and purple. The main title is centered in a dark blue rectangular area.

# SPSSシンタックスによる データ加工と分析入門

立教大学社会学部  
村瀬 洋一

1

The slide has a decorative header with a blue and purple gradient and a small square graphic. The content is presented in a clean, white background with blue bullet points.

## 1. Syntaxとは何か

- SPSSにおけるデータ加工や分析などの命令文（プログラム）  
クリックせずに、自分の操作を記述できる
- SPSSで、シンタックスのウィンドウを開く  
画面上「File」→新規作成→シンタックス

2

## シンタックスの長所

- 大量の操作(データ加工、分析)ができる
- 自分が実行した操作を保存できる
- 後で同じ操作を再現できる

3

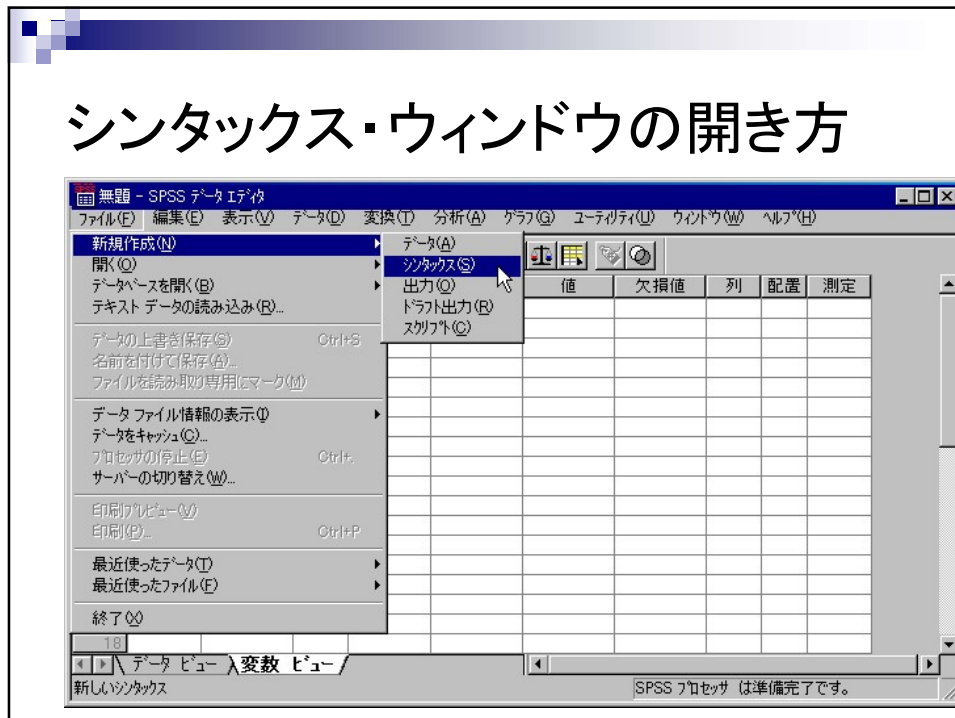
## 2. 使い方

- シンタックス・ウィンドウを開く  
SPSSには、シンタックス・ウィンドウの他、データウィンドウ、出力ウィンドウがある
- 自分でシンタックスを書く
- 実行する(実行ボタンを押す)



4

## シンタックス・ウィンドウの開き方



5

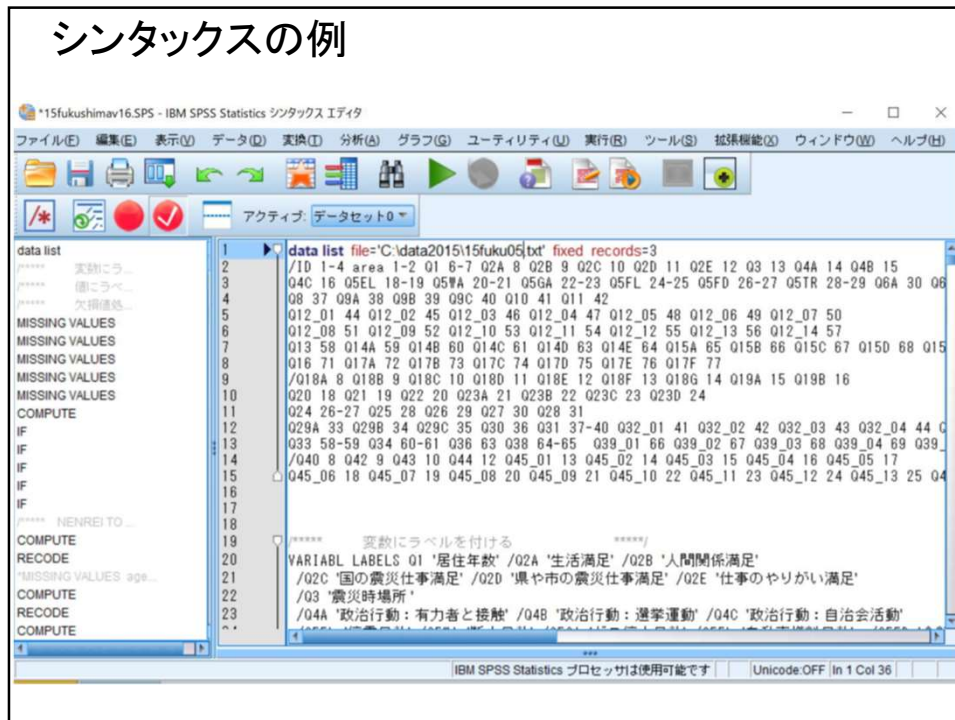
- この後、自分でシンタックスを書いた後に、実行する

実行ボタンを押すか、あるいは、

[ctrl+A\(全選択\)](#), [ctrl+R\(実行\)](#)

- シンタックスのファイルは、好きな名前を付けて保存しておくこと

6



7

- 書くときの注意点
  - シNTAXは大文字と小文字を区別しない。
  - 半角文字のみ。全角文字は使わない。
  - **全角空白**があるとエラーが出て止まってしまうので注意。
  - 命令文の最後に1つだけピリオドをつける。
- 変数名(調査ならば質問項目の名前)は、  
実在する名前を正確に書くこと。変数名は  
SPSSのデータ・ウィンドウで確認すればよい。

詳しくは村瀬他『SPSSによる多変量解析』オーム社 p.47  
 参照

8



9

### 3.1. データ読み込み命令文 テキスト形式データファイルの場合

普通、ハードディスクはCドライブ

例 ドキュメント内のdata2009 というフォルダ内にある

da09.txt ファイルを読み込め、という命令文

username はWindowsにログインした時のIDを入れる

```
DATA LIST FILE=' C:¥Users¥username
¥Documents¥data2009¥da09.txt' FIXED RECORDS=2
/1 ID 1-4 Q1 5 Q2 6 Q3 7 Q401 8 Q402 9 ...
```

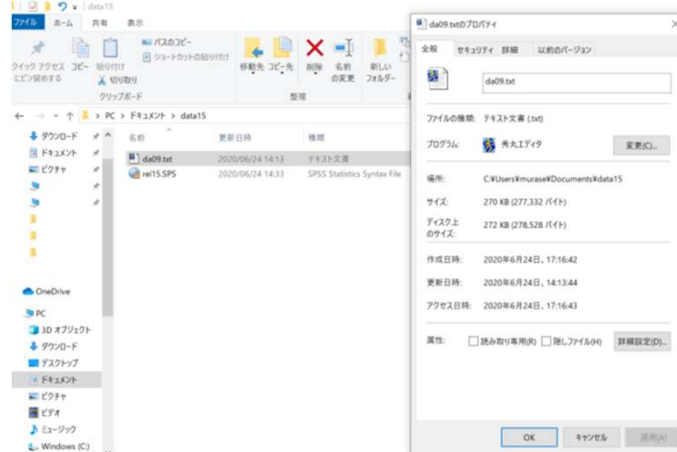
DATA LIST FILE シンタックスによりデータファイルを開く

10

ドキュメント内のdata15というフォルダの中にある  
da09.txtというファイル ... の場所(フルパス)は?

ファイルを右クリックして、プロパティを見る  
→ 場所 の部分

C:\Users\username\Documents\data15



11

マックOSにおいて、ファイルやフォルダの場所(フルパス)を見るには

txtファイルなどを選択し、コンテキストメニュー(右クリック、あるいはctrl+クリック)、またはメニューの情報を見る(Cmd+i)で、情報ウィンドウを表示させる。

一般情報 > 場所 がフォルダのパス。/Users/以降すべてが場所。

12

## データの場所の書き方 Mac OSの場合

ハードディスク内のdata15 というフォルダ内にあるda09.txt  
ファイルを読み込め、という命令文

```
DATA LIST FILE= '/Users/data15/da09.txt' FIXED
RECORDS=2
/1 ID 1-4 Q1 5 Q2 6 Q3 7 Q401 8 Q402 9 ...
```

DATA LIST FILE シンタックスによりデータファイルを開く

13

## 3.2.データ読み込み命令文 SPSS形式データファイルの場合

普通、ハードディスクはCドライブ

- ドキュメント内のdata2009 というフォルダ内にある  
drei.savファイルを読み込め、という命令文  
GET FILE= 'C:¥Users¥username  
¥Documents¥data2009¥drei.sav
- Hドライブ内のmydata というフォルダ内にあるdrei.sav  
ファイルを読み込め、という命令文  
GET FILE= 'H:¥mydata¥drei.sav

GET FILE シンタックスによりデータファイルを開く

14

## データ・ウィンドウの例

	ID	Q1	Q2	Q3	Q401	Q402	Q403	Q404	Q501	Q502	Q503	Q504	Q505	Q506	Q507	Q508
1	101	1	3	4	1	1	1	3	4	3	3	2	2	3	2	
2	102	2	2	3	3	2	1	2	2	3	2	2	2	3	1	
3	103	2	2	2	2	2	2	2	3	2	1	1	1	2	1	
4	104	2	4	4	2	1	1	1	4	4	4	1	1	1	1	
5	105	2	3	2	1	2	1	1	3	3	2	2	3	1	1	
6	106	3	4	3	2	1	2	2	3	2	2	1	2	3	2	
7	107	2	2	3	4	2	1	2	3	4	1	2	2	3	2	
8	108	1	4	4	3	1	1	2	4	4	2	3	1	1	1	
9	109	2	2	3	2	1	2	2	2	2	3	3	2	2	1	
10	110	3	3	4	2	2	2	3	1	3	2	1	1	1	1	
11	111	2	3	3	2	2	1	2	3	2	2	3	2	2	2	
12	112	1	4	3	4	2	2	3	4	4	4	4	2	2	3	
13	113	2	2	2	4	1	2	1	3	2	2	1	1	1	1	
14	114	2	2	3	2	2	1	2	4	3	3	1	2	1	1	
15	116	2	2	4	4	3	2	2	2	3	3	3	1	9	3	
16	117	2	2	2	2	2	2	2	3	3	3	2	2	2	2	

15

## 4. データの中身について

### ■ 横1行が1人の回答

1000人の調査回答なら1000行ある  
Q1の質問項目(変数)で1と答えたなど

### ■ 4段階回答だが、9と答えている人がいる

非該当8、無回答(DK/NA)9 とすることが多いため  
DK/NA: Don't Know, No Answer

### ■ 複数回答

いくつでも○をつけてください、という問の場合  
○をつけた人は1、なければ0、という1,0のデータになっている

16



## 5. データ加工の例

### リコード文によるカテゴリ合併

- 例 変数Q2の4段階回答を2段階にする。

以下は、Q2という変数をもとに、N2という新変数を作成。  
新変数名がN2ということを理解する。新変数は、何でも好きな名前でもよい。

リコード文によりN2の回答を2段階にする例。

```
COMPUTE  N2 = Q2 .
RECODE  N2 (1, 2=1) (3, 4=2) .
```

```
CRO                                     ← クロス集計命令
```

```
  /TAB = N2 by Q2
```

```
  /CEL = ROW .
```

- 上記のシンタックスを実行すると、N2を用いたクロス集計結果が出る。変数名は、自分の使いたい変数の名前を書けばよい。
- リコード文のみを実行しても何もおきないように見えるが、その後に新変数を使って分析を行えばよい。

17

## 例 収入カテゴリの作成

- 収入の問がQ15 だとする

回答は1～14までの選択肢

```
COMPUTE ZAISANCT=Q15 .
RECODE  ZAISANCT (1 thru 4 =1) (5, 6, 7=2)
(8, 9=3) (10, 11=4) (12, 13, 14=5) .
```

この場合、Q15が元の変数、ZAISANCTが新変数名。

18

## 6. 例 年齢を10歳刻みに分けるには

- リコード文によるカテゴリー合併の例

AGEという細かい変数を、NENDAIという5段階の変数（新しい変数）に変換。

新変数名がNENDAIということを理解する。

```
COMPUTE      NENDAI =AGE.
RECODE      NENDAI (1 THRU 29 =2) (30 THRU 39 =3)
              (40 THRU 49 =4) (50 THRU 59 =5) (60 THRU 97 =6) .
```

```
CRO
/TAB = NENDAI by Q3
/CEL = ROW .
```

- AGEの部分は、年齢の変数名に書き変える
- Q3 の部分は、自分が分析したい変数名に書き変える  
上記のシンタックスを実行すると、年代ごとの集計が出る。

19

## 7. 例 学歴を教育年数に変換するには

- リコード文による値の書き換えの例

Q16という本人学歴の変数を、EDUという教育年数の変数（新変数）に変換。

以下のシンタックスを書いて実行すればよい。

```
COMPUTE      EDU=Q16.
RECODE      EDU (1=6) (2=9) (3=12) (4=13)
              (5=14) (6=16) (7, 9=99) .
MISSING VALUES EDU (99) .
```

1と答えた人（小学校卒）なら6年、などに変換している。

7のその他や9の無回答は欠損値としている。

- Q16の部分は、学歴の変数名に書き変える。
- これだけでは何もおきないが、この後に新変数EDUを使って、クロス集計など分析を行えばよい。

20

- 無回答は普通、9や99を入力している。
- そのままだと、分析上問題が起きるので、上記のようにMISSING VALUES コマンドを使い、欠損値処理をする。
- きちんと新変数ができたかどうか、単純集計やクロス集計表を出して、新変数の内容を確認してから分析すること。

21

## 8. ある値の人にデータを限定

- SELECT文を使うと、ある条件の人のみのデータとなる。
- データ人数が減るので注意。
- 例1 70歳未満の人のみを使って分析したい場合

```
SELECT IF AGE < 70.
```

AGE 部分は存在する変数名を正確に書く

22

- 例2 ISSPデータで、日本のみに絞る。0歳未満の人のみを使って分析したい場合

```
SELECT IF country = 392.
```

country部分は、国を表す変数名を正確に書く。年によって異なる。値392が日本ならば、このように書くこと。値はデータウィンドウで確認する。

SELECT文を使うと、データファイル自体の人数が減るので注意する。データは新たに名前を付けて保存すること。

23

## 9. 男女別にデータを分割

```
SORT CASES BY Q21 .  
SPLIT FILE BY Q21 .
```

- Q21の部分は、性別の変数名に書き変える
- データ・ウィンドウを見ると、全員のデータが性別順に並べ替えられている。
- これを実行した後に分析を行うと、すべての結果が男女別に出る。

24

## 10. IF文の使い方

- IF文を使うと、ある条件の場合のみ、何らかの命令を実行可
- 例1 年齢が30歳以上の場合に、変数Q7Aの値に1を足す  
( )内に条件を書き、その後に、条件に合う場合に  
実行したい命令を書けばよい  
IF (age >29) Q7A=Q7A+1.
- 例2 専門職ダミー変数作成。変数Q25の値が2か3の場合に、  
変数SENMONの中身を1にする。  
COMPUTE SENMON =0.  
IF (Q25=2) SENMON =1.  
IF (Q25=3) SENMON =1.

Q25が2か3の場合のみ、SENMONの中身は1、あとは0になる。  
SENMONは新変数なので、何でも好きな名前でもよい。

25

## 11. 分析命令文の見本

- 度数分布(単純集計)シンタックス見本  
FRE VAR = Q2 Q3 .
  - クロス集計シンタックス見本  
CRO  
/TAB = V3 by V29  
/CEL = ROW  
/STA .
- /CEL の行に COL を書くと縦%、COUを書くと実数が出る。
- RECODE文やIF文だけでは何も出ない。度数分布やクロス集計を実行すること。

26

### ■ 重回帰分析の例

被説明変数がQ6Bの場合

```
REG /DEP Q6B  
/METHOD=ENT Q2 Q3 Q4A .
```

### ■ 分散分析の例

被説明変数がEDUの場合

```
UNI EDU by sex nendai  
/PLOT = PRO( nendai * sex ) .
```

- 変数名は自分が分析したい変数に書き換える。
- 最後に1つだけピリオドがつくのを忘れずに。

27

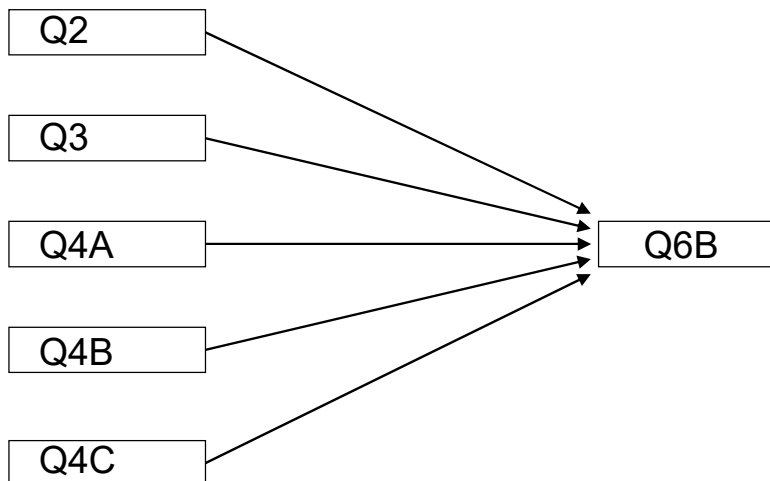


図1 重回帰分析のモデル

28

## 重回帰分析の注意点

- **多重共線性**に注意する
  - X同士の相関を確認する
- 分析前に
  - 変数の方向をそろえる
  - 欠損値処理をする
    - まず、用いる変数の分布を確認すること
- なお、男女別、国別にデータを分割後に分析した方が、説明力が上がることが多い。
- YもXがすべて連続量という点が、分散分析と異なる。できるだけ細かい変数を使うこと。

29

## 12. 変数や値に名前をつける

- 変数に名前をつける  
VARIABLE LABELS
 

```

Q1      '年齢'
/ Q2    'Life Satisfaction'
      
```
- 値に名前をつける  
VALUE LABELS Q6A TO Q6E
 

```

1 'agree'
2 'somewhat agree'
3 'somewhat disagree'
4 'disagree'
9 '無回答'
      
```

その他、詳しくは「応用調査実習」ホームページ参照

30