

# SPSSシンタックスによる データ加工と分析入門

2024/04

村瀬 洋一

1

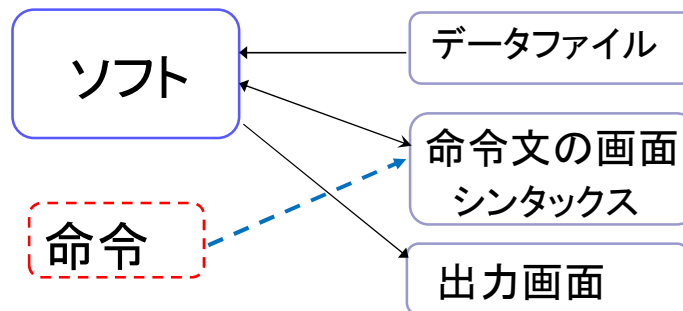
## 1. Syntaxとは何か

- 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
- シンタックスの利点
  - ・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
    - 後で同じ操作を簡単に再現できる

シンタックスを覚えると便利！

2

## SPSSの仕組み



- ・人間が命令を実行すると、出力画面に文字や表が出る
- ・起動すると初めは空のデータ画面のみがある
- ・データにはテキスト形式データ(txt)と、  
SPSS形式データ(sav)がある

3

## パソコン使用の注意点

- たくさんウィンドウを開かない。メモリを消費して動きが遅くなる。**開くのは最小限に！**
- ファイル名やフォルダ名、ユーザー名は、全角文字を使わない。分析ソフトを使った時にフリーズしやすい。
- ファイル名は半角文字で短めに。22fuku.txt など
- 手元のパソコンにファイルを保存する。ネット上に置いたまま開かない。
- ZIPファイルから、CDドライブなど手元のハードディスクにまずデータファイルを出す。その後に、SPSSなどでデータを開く。
- エクセルの「セルの書式設定」を使いこなす。
- エディターソフトを使う。作業効率がよくなる。

4

## 2. シンタックスの使い方

- 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

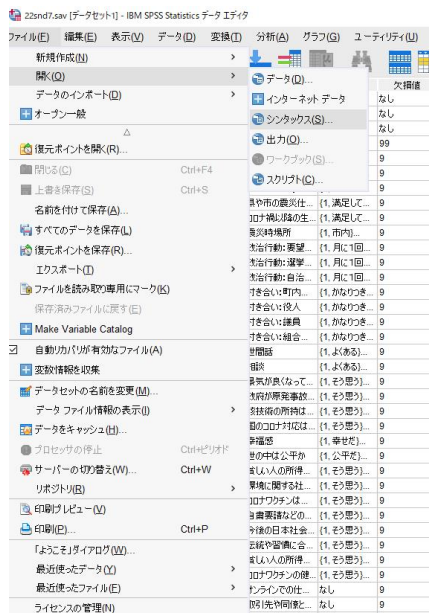
- 実行する(ctrl+R または実行ボタン▲を押す)

```

SPSS Statistics シンタックス エディタ
データ(D) 変換(O) 分析(A) グラフ(G) ユーティリティ(U) 実行(R) ツール(S) 拡張機
COMPUTE EDU=Q49.
38 RECODE EDU(1=6)(2=9)(3=12)(4=13)(5=14)(6=16)(7,9=99).
39 MISSING VALUES edu (99).
40
41 ***** 変数にラベルを付ける *****
42 VARIABL LABELS NENDAI '回答者年代' / EDU '教育年数'.
43
44
45 ***** GAKUREKI KATEGORI SAKUSEI *****
46 COMPUTE EDUCT=Q49.
47 RECODE EDUCT(1,2=1)(3,4=2)(5,6=3)(7,9=9).
48 VALUE LABELS educt 1 '中卒以下' 2 '高卒専門学校'
49 3 '短大高専大卒以上' 9 'わからない'.
50
    
```

5

## シンタックス・ウィンドウの開き方



6



## ■ シンタックスの注意点

- 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。  
ピリオドが多いとエラーが出るので注意。

## ■ 変数名(調査ならばQ3Aなど質問項目の名前)は、実在する名前を正確に書くこと。

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

9



10

## 社会調査結果の例

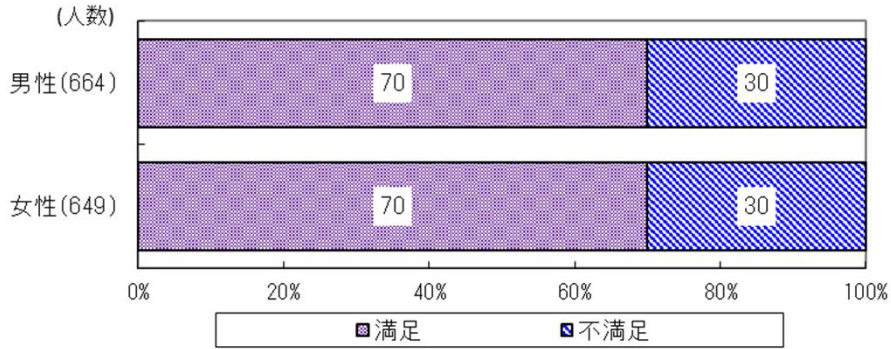


図1. 無関連状態（架空の例） 「あなたは現在の生活全般に満足していますか、あるいは不満ですか」  
2022仙台市調査調査

11

## 社会調査結果の例

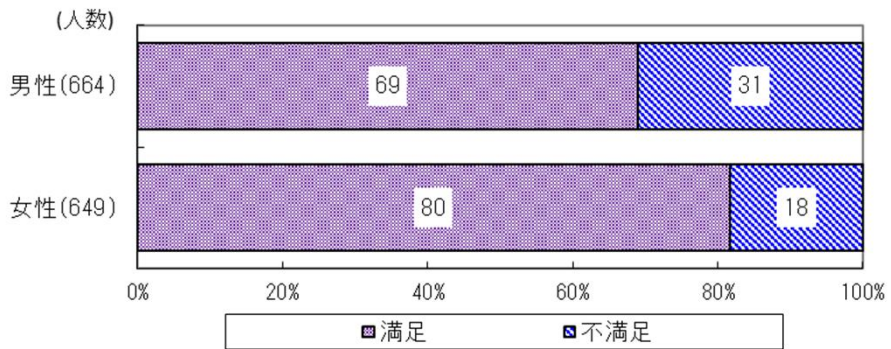


図2. 男女で差がある状態（架空の例） 「あなたは現在の生活全般に満足していますか、あるいは不満ですか」  
2022仙台市調査調査

12

### 3. データの読み込み

- テキスト形式データの場合は、まず、データ読み込み命令文を書く必要がある。
- SPSS形式データ(**sav**)が既にある場合は、それをSPSSで開くだけでよい。

13

### テキスト形式データの例

- 行が個人、列が変数となる数字の行列
- 具体例 3人分のデータ行列の例

```
00101 21412508 2421111111  
00102 21611402 1221213132  
00103 12714806 1222212121
```

- 5カラム目までがサンプル番号。
- 半角数字のみを入力。

データ分析とは、このようなデータ行列の分析

14

### 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

以下はドキュメント内の`data2015` 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。

```
DATA LIST FILE=' C:\%Users%\username
¥Documents¥data2015¥da09.txt' FIXED RECORDS=2
/1 ID 1-4 Q1 5 Q2 6 Q3 7 Q401 8 Q402 9 ...
```

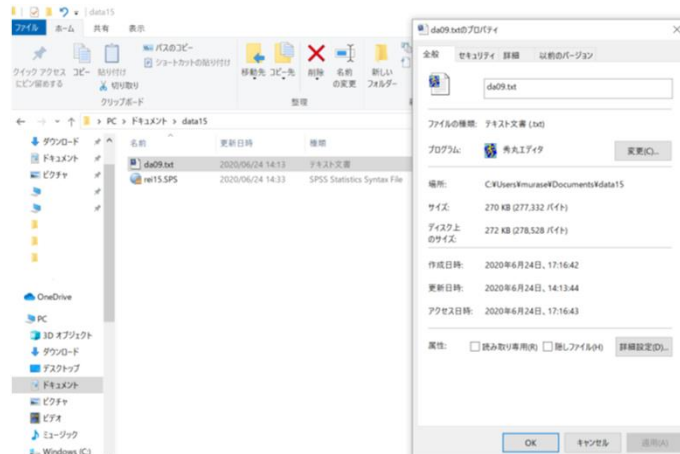
DATA LIST FILE シンタックスによりデータファイルを開く

15

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



16



## マックOSにおいて、ファイルやフォルダの場所(フルパス)を取得するには

- ファイル情報を見る

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。

その中の **一般情報 > 場所** がフォルダのパス。  
/Users/以降すべてが場所。

- 詳しくは「MAC フルパスの書き方」でグーグル検索するとよい。

17

## 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

18

## 3.2. データ読み込み命令文

### SPSS形式データファイルsavの場合

savファイルをダブルクリックすればデータが開く

以下のシンタックスを書いて開いてもよい。savの場合は桁指定を書く必要はない

GET FILE シンタックスによりデータファイルを開く

- ドキュメント内の **data20** というフォルダ内にある **rei7.sav** ファイルを読み込め、という命令文

```
GET FILE='C:\Users\¥username¥Documents¥data20¥rei7.sav'.
```

- Hドライブ内の **mydata** というフォルダ内にある **rei7.sav** ファイルを読み込め、という命令文

```
GET FILE='H:\¥mydata¥rei7.sav'.
```

19

## データ・ウィンドウの例

画面下にデータビューと変数ビューのタブがある

名前	型	幅	小数桁数	ラベル	値	欠損値	列	配置	尺度	役割
1 ID	数値	4	0	なし	なし		6	右	スケール	入力
2 area	数値	2	0	なし	なし		6	右	スケール	入力
3 pid	数値	2	0	なし	なし		5	右	スケール	入力
4 Q1	数値	2	0	居住形態	なし	999	4	右	スケール	入力
5 Q2A	数値	1	0	生活満足	{1.満足して...}		5	右	名義	入力
6 Q2B	数値	1	0	周囲の人々関係	{1.満足して...}		5	右	名義	入力
7 Q2C	数値	1	0	国の農林仕事	{1.満足して...}		5	右	名義	入力
8 Q2D	数値	1	0	県や市の農林仕事	{1.満足して...}		5	右	名義	入力
9 Q2E	数値	1	0	コトが満足度の生	{1.満足して...}		5	右	名義	入力
10 Q3	数値	1	0	職業の種類	{1.市町}		4	右	名義	入力
11 Q4A	数値	1	0	政治行動:選挙	{1.月に1回...}		5	右	名義	入力
12 Q4B	数値	1	0	政治行動:選挙	{1.月に1回...}		5	右	名義	入力
13 Q4C	数値	1	0	政治行動:自治	{1.月に1回...}		5	右	名義	入力
14 Q5A	数値	1	0	付き合い:町内	{1.かたがた...}		5	右	名義	入力
15 Q5B	数値	1	0	付き合い:役人	{1.かたがた...}		5	右	名義	入力
16 Q5C	数値	1	0	付き合い:議員	{1.かたがた...}		5	右	名義	入力
17 Q5D	数値	1	0	付き合い:組合	{1.かたがた...}		5	右	名義	入力
18 Q6A	数値	1	0	世間話	{1.よくある...}		5	右	名義	入力
19 Q6B	数値	1	0	相談	{1.よくある...}		5	右	名義	入力
20 Q7A	数値	1	0	暴気があって	{1.そう思う...}		5	右	名義	入力
21 Q7B	数値	1	0	政治的関心	{1.そう思う...}		5	右	名義	入力
22 Q7C	数値	1	0	政治的関心	{1.そう思う...}		5	右	名義	入力
23 Q7D	数値	1	0	国のコトが好きな	{1.そう思う...}		5	右	名義	入力
24 Q8	数値	1	0	幸福感	{1.幸せだ...}		4	右	名義	入力
25 Q9	数値	1	0	世の中は公平か	{1.公平だ...}		5	右	名義	入力
26 Q10A	数値	1	0	富しい人の所得	{1.そう思う...}		5	右	名義	入力
27 Q10B	数値	1	0	富しい人の所得	{1.そう思う...}		5	右	名義	入力
28 Q10C	数値	1	0	コトが好きな	{1.そう思う...}		5	右	名義	入力
29 Q10D	数値	1	0	自営業者などの	{1.そう思う...}		5	右	名義	入力
30 Q11A	数値	1	0	今後の日本社会	{1.そう思う...}		5	右	名義	入力
31 Q11B	数値	1	0	伝統や習慣に	{1.そう思う...}		5	右	名義	入力
32 Q11C	数値	1	0	富しい人の所得	{1.そう思う...}		5	右	名義	入力
33 Q11D	数値	1	0	コトが好きな	{1.そう思う...}		5	右	名義	入力
34 Q12_01	数値	1	0	オンラインでの住	なし		5	右	名義	入力
35 Q12_02	数値	1	0	取引先や同僚と	なし		5	右	名義	入力
36 Q12_03	数値	1	0	友人や知人と	なし		5	右	名義	入力

20

## 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  
データビュー画面で確認

- 1つを選択 (Single Answer)  
ただ、4段階回答だが、9という値があることも。  
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 (Multiple Answer)  
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型のデータ。

21

## 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) .
```

```
GRO                                ← クロス集計命令
/TAB = N2 by Q2 .
```

- 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起きてい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何らかの結果が出る。

22

## 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 = Q3 BY NENDAI
/CEL = COL .
```

- **AGE**の部分は自分が使うデータの年齢の変数名を書く。
- **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。

23

## 6. 例 収入カテゴリーの作成

- 収入の問が**Q15**だとする  
 回答は1～14までの選択肢。  
 以下は5段階の新変数を作る例。

```
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。  
 元の変数名が間違っているとエラーになる。

24

## 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、学歴の変数名を書くこと。
- 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。

25

- 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。

26

## 8. 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、  
新変数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。

27

## 9. 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

28

- 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。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。

29

## 10. 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。

30

## 11. 変数や値にラベルをつける

- 変数にラベルをつける

VARIABLE LABELS

```
Q1      '年齢'
/ Q2    'Life Satisfaction' .
```

- 値にラベルをつける

VALUE LABELS Q6A TO Q6E

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

その他、詳しくは参考書など参照

31

## 12. 分析命令文

- 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 .
```

/TAB の行に自分が使い田変数名を書く。  
 /CEL の行に COL を書くと縦%、COUを書くと実数が出る。  
 /STA 行を付けると各種の関連係数などが出る。

32



### ■ 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。

33

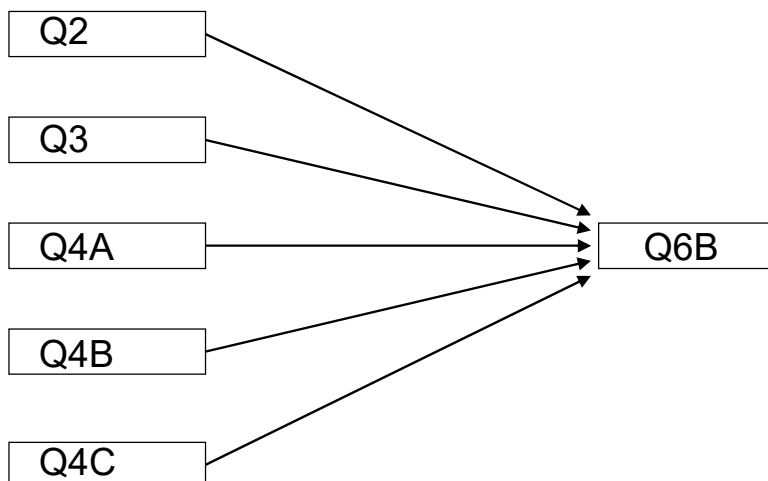


図5 重回帰分析のモデル

34

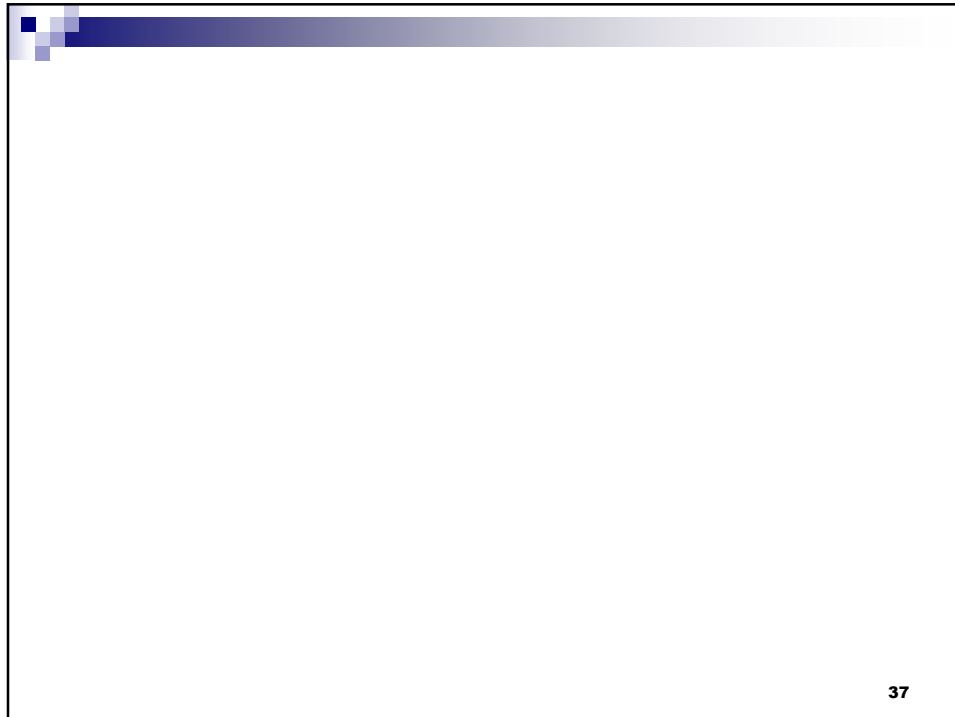
## 重回帰分析の注意点

- 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。

35

- 重回帰も分散分析も、複数のXを用いて、Yと関連が強いものを発見する、という点では、数学的にまったく同じモデルである。
- ただし重回帰分析は、XもYもすべて量的変数を使う。分散分析のXは質的変数(カテゴリ)であり、量的な意味はない。

36



## おまけ SPSSの文字化け対策

見本シンタックスを開いたら文字化けする時

SPSSの画面上「編集」→「オプション」 以下の図6

日本語の表記をする時は「ロケールの書き込み」を選択すること。

- ロケールエンコード：言語や国・地域ごとに異なる単位や記号などの表記規則を使う。
- Unicode エンコード：世界中の文字を共通の表記規則で扱う。

あるいは、シンタックスウィンドウに以下を書いて実行。

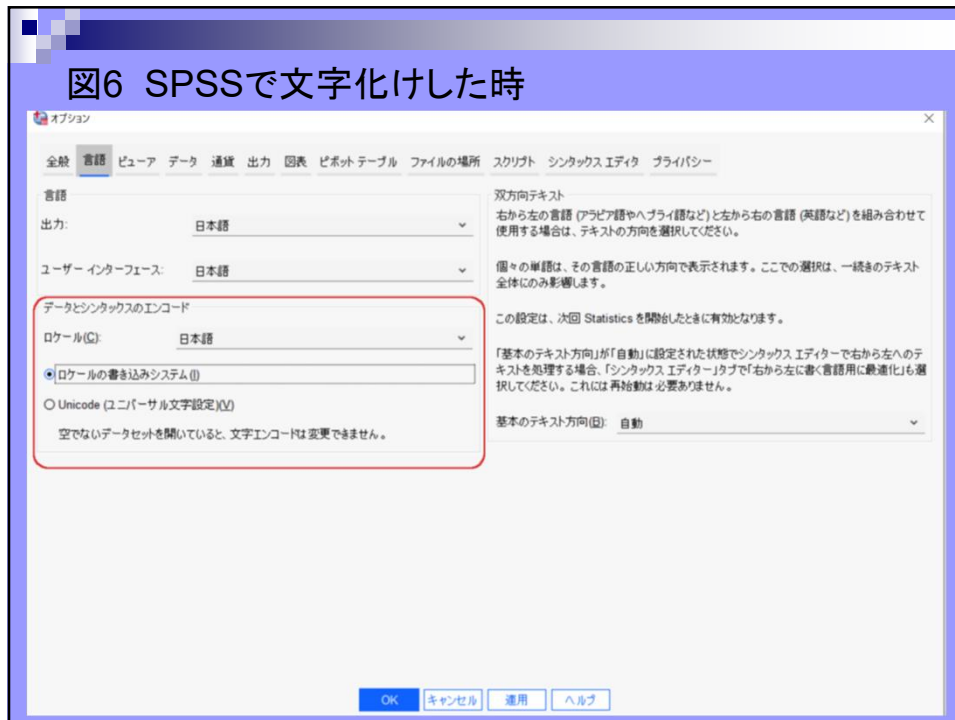
SET OLang=Japanese Unicode=No Locale=Japanese .

あるいは、SPSSの画面上「表示」→「フォント」

フォントを MSゴシックなど日本語フォントを選択し、[OK]ボタンを押す。

38

図6 SPSSで文字化けした時



## おまけ テキスト形式データの例

```

1001 29133332443 333444131222 343444414422323 0001100000353↓
1002 22322222443 344323332334 224144224424422 1111110100132↓
1003 03322212444 444444342334 224434222423212 1011010000333↓
1004 58122211343 343324231234 112144323413331 0110100100333↓
1005 43122211431 244422411231 223414423413312 0011100000231↓
1006 03223414443 344444344444 234444444444412 0010000010143↓
1007 50322222442 343423122233 223444414414212 0011100100323↓

```

- 普通は最初がID番号。
- 半角数字のみを入力。
- これをもとにSPSS形式データファイルを作る。