

12.15 フォントの設定: Results window / Viewer

OSのバージョンによって、フォントの設定方法が異なります。

**Results window**

**Viewer**

DFKyokasho-W3-MP-RKSJ-Hのフォントを選択可能な場合は設定をおすすめします。

選択できない場合  
他のいずれのフォントを選択してもResults Windowのみ日本語が正しく表示されない場合があります。

12.16 フォントの設定: Data Editor / Command window

OSのバージョンによって、フォントの設定方法が異なります。

**Data Editor**

**Command window**

12.17 フォントの設定: Variables window / Review window

OSのバージョンによって、フォントの設定方法が異なります。

**Variables window**

**Review window**



## データの読み込み

Stataにて読み込み可能なデータ形式

**CSVファイル**からのデータの読み込みをおすすめします

▶ **Excelファイル** (.xls または .xlsx)  
Stata12以降ではExcelファイルからの読み込みが可能  
ただし、40メガバイトを超えるExcelファイルからの読み込みは不可

▶ **Stataファイル** (.dta) ← 開く  
ExcelまたはCSVファイルからStataに読み込んだデータをStata形式で保存したファイルを開く

## データの取り込み方法

- ▶Excelデータのコピーを貼り付け
- ▶スクリプト(doファイル)を用いたデータの取り込み

## データの取り込み方法

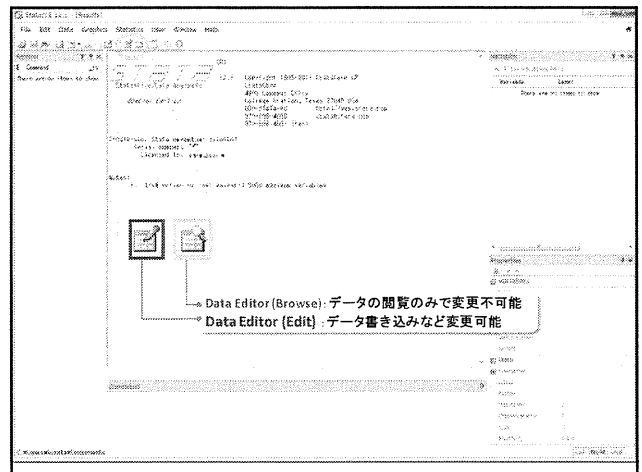
- ▶Excelデータのコピーを貼り付け
- ▶スクリプト(doファイル)を用いたデータの取り込み

## Excelデータの貼り付け

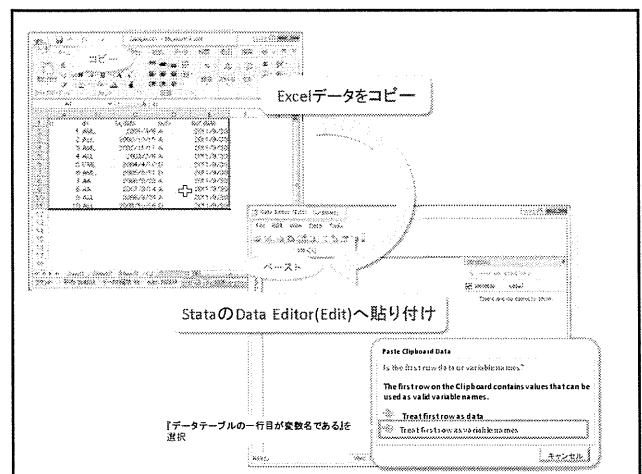
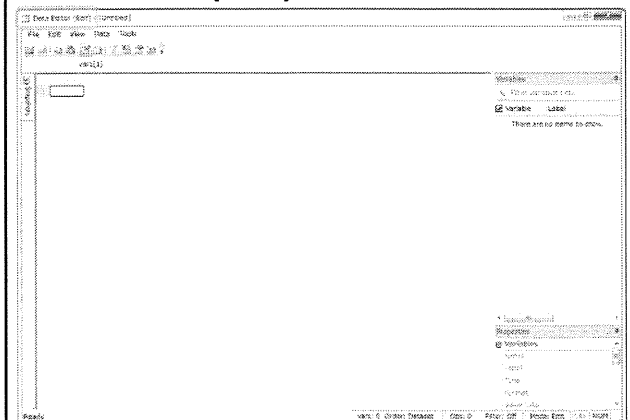
Excelからデータをコピーし、Data Editorに直接貼り付ける方法

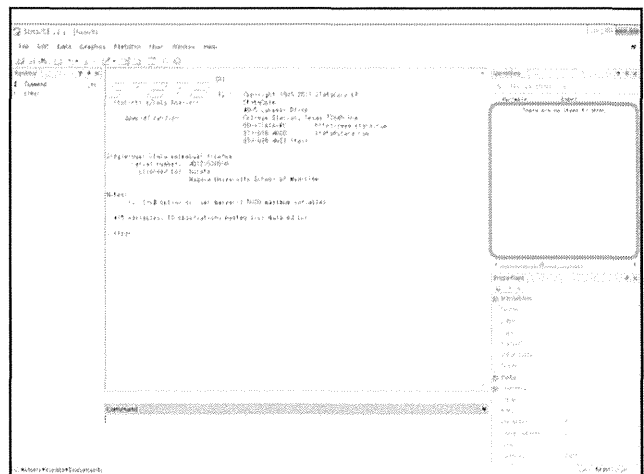
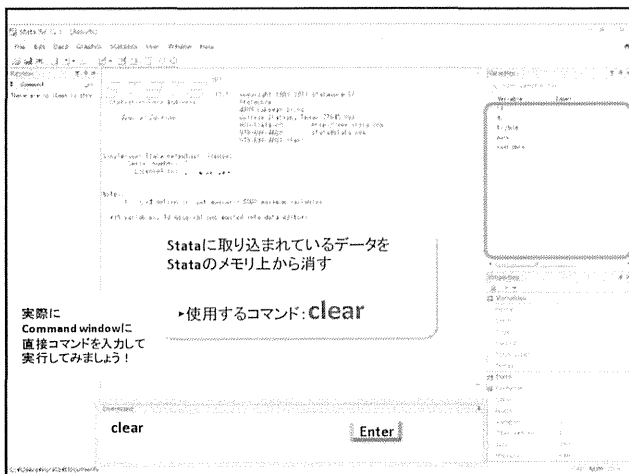
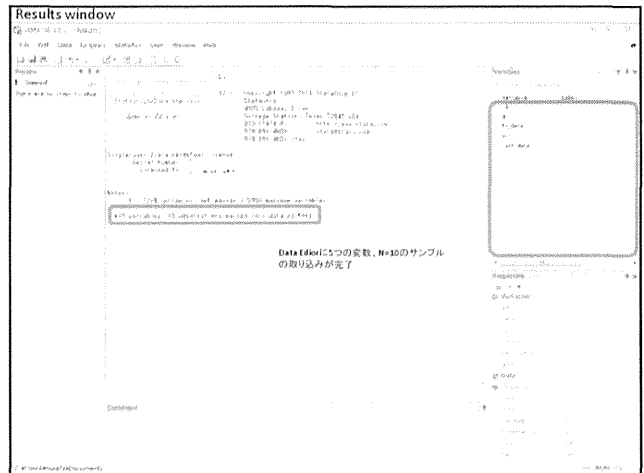
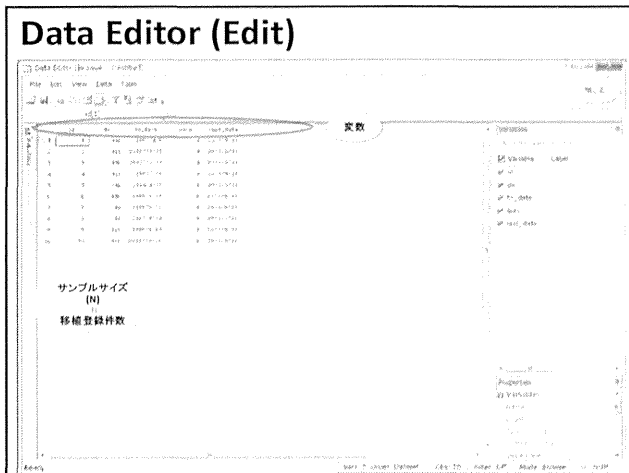
- ▶Stataに取り込むデータ (サンプルデータ)

 Sample.xlsx



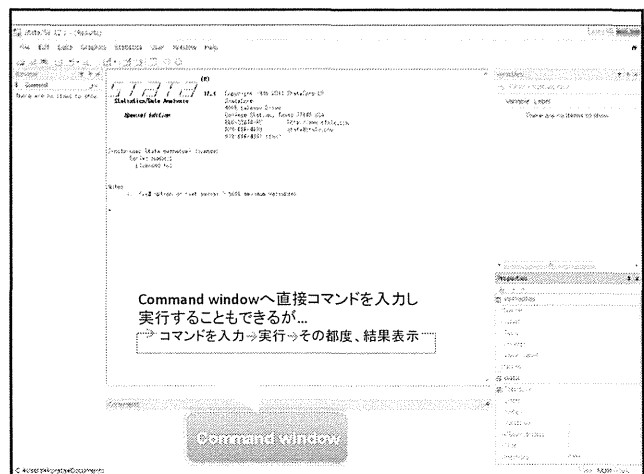
## Data Editor (Edit)





## データの取り込み方法

- ▶ Excelファイルデータのコピーと貼り付け
- ▶ スクリプト(doファイル)を用いたデータの取り込み



## スクリプト(doファイル)を用いたデータの取り込み

### doファイルとは…

コマンドを実行する順に書き並べたファイル

- ▶複数のコマンドを一度に実行することができる  
(実行したいコマンドのみを選択し実行することも可能)
- ▶作業メモなども書き留めることができる
- ▶doファイルとして保存できるため、その都度コマンドを打ち込む必要がなく、同じ作業を繰り返し行うことができる

JSHCT TRUMPデータを用いた登録研究の解析を行う場合、WG研究のための統計解析スクリプトをJSHCTホームページよりダウンロード可能

## スクリプト(doファイル)を用いたデータの取り込み

配布したスクリプトファイル(do file)を用いてデータを取り込む方法

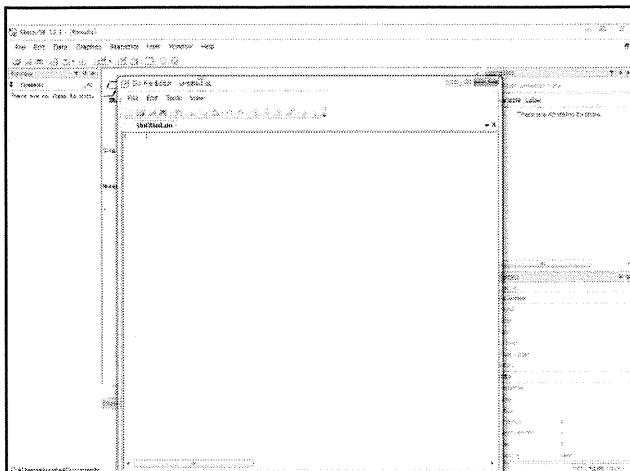
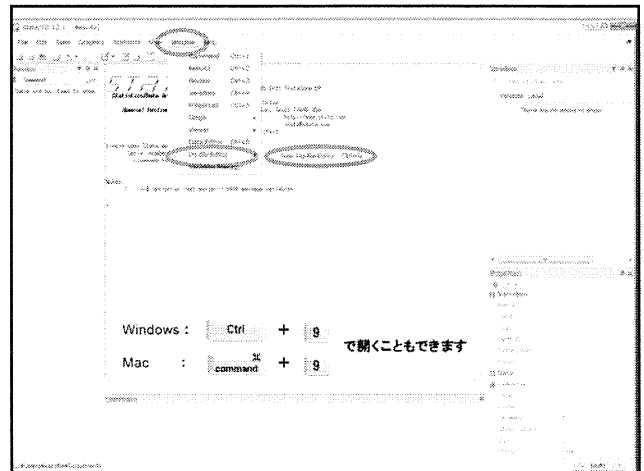
### ▶取り込むデータ

 TRUMPSeminar2014.csv

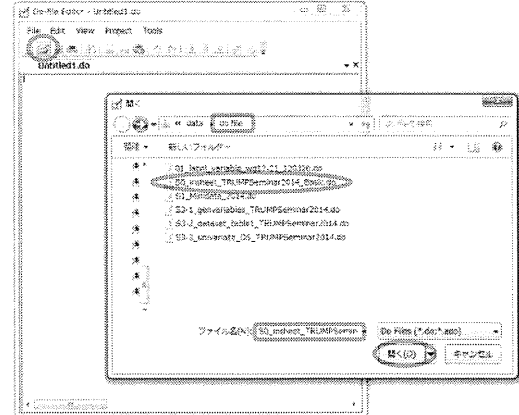
### ▶実行ファイル(スクリプト)

 S0\_insheet\_TRUMPSeminar2014\_Basic.do

## スクリプト実行のための Do-file Editorの起動



## Do-file Editor

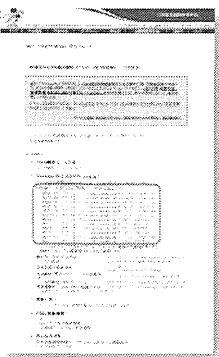




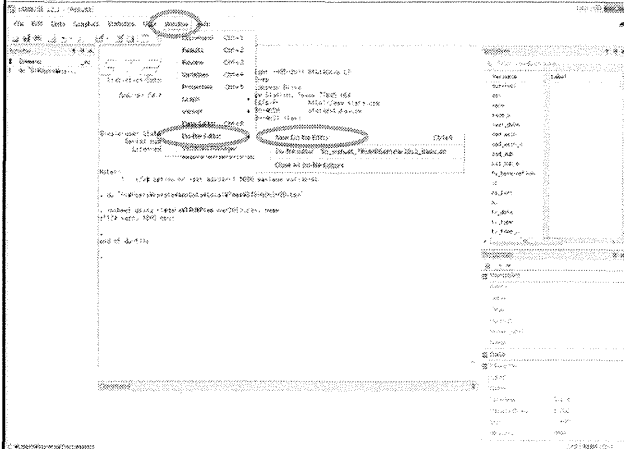
## 変数に日本語ラベルを付ける

変数の内容がわかるように変数

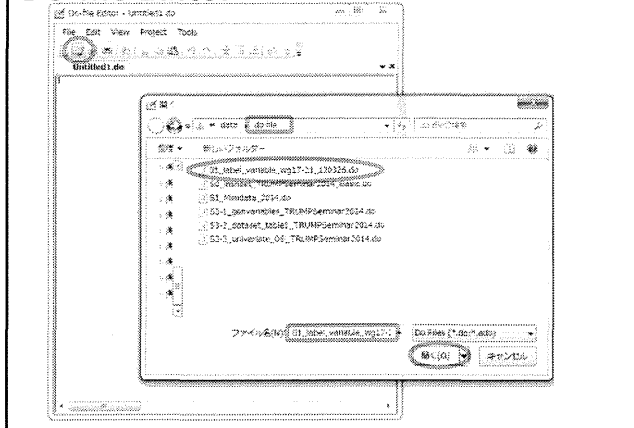
JSHCT TRUMPデータを用いた登録  
場合、各WGのデータセットに対応  
スクリプトをJSHCTホームページよ



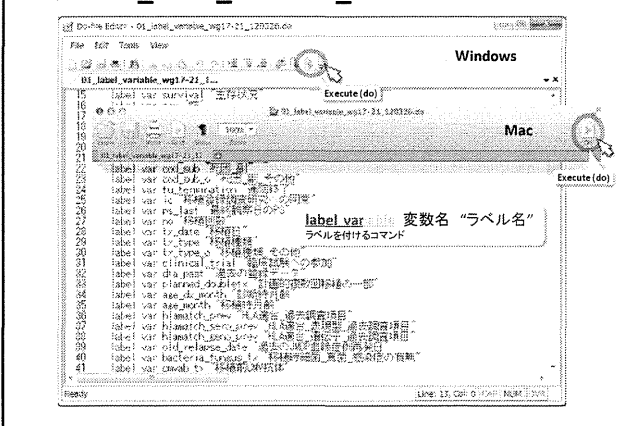
## 日本語ラベルを付けるコマンドを書く: Do-file Editor



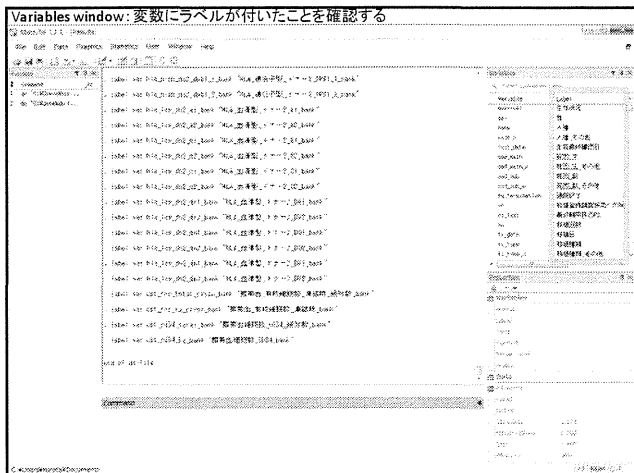
## Do-file Editor



## スクリプト: 01\_label\_variable\_ .do

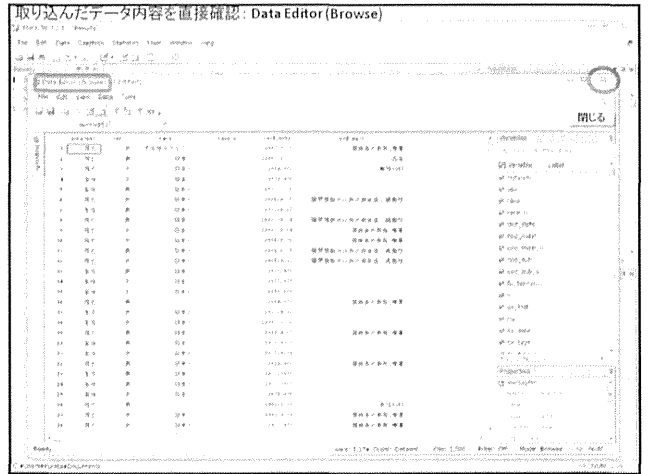
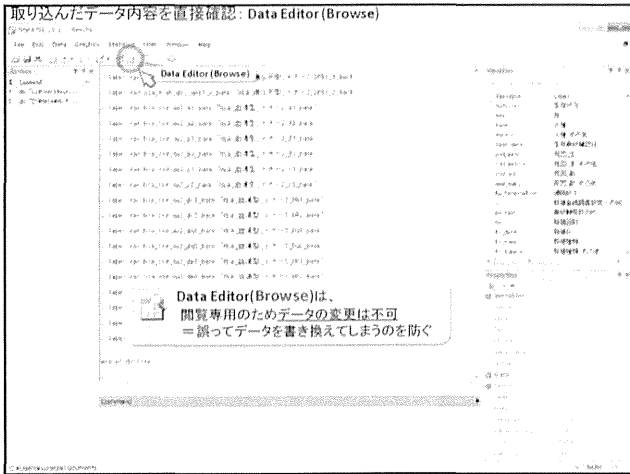


## Variables window: 変数にラベルが付いたことを確認する



## 取り込んだデータの確認

- ▶取り込んだデータ内容を直接確認
- ▶標本数と変数の属性を確認  
(各変数のタイプ、ラベル名、観測値、変数の数など)
- ▶連続変数の要約値を確認  
(指定した連続変数の観測値、平均値、標準偏差、最小値/最大値)
- ▶カテゴリー変数の要約値を確認  
(指定したカテゴリー別の頻度、割合)



## 取り込んだデータの確認

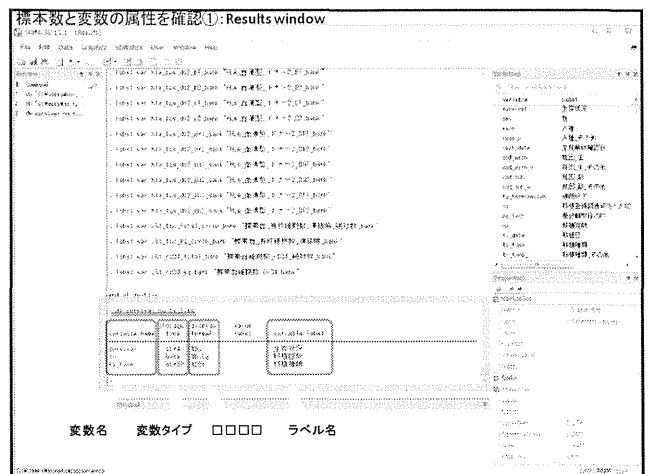
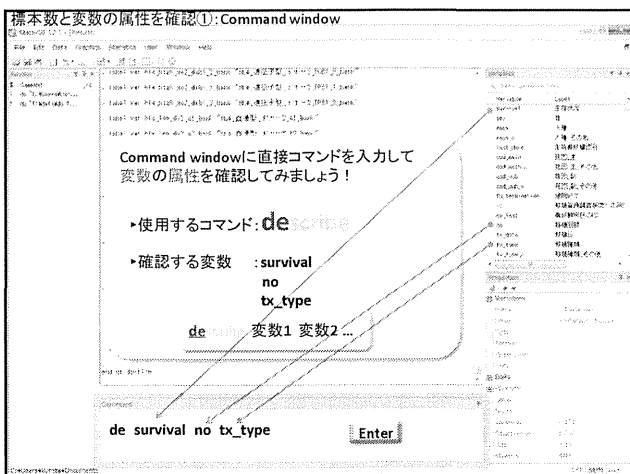
- ▶ 標本数と変数の属性を確認  
(各変数のタイプ、ラベル名、観測値、変数の数など)
  - ▶ 連続変数の要約値を確認  
(指定した連続変数の観測値、平均値、標準偏差、最小値・最大値)
  - ▶ カテゴリー変数の要約値を確認  
(指定した)
- ①実際にコマンドを入力し実行  
②配布スクリプトを実行  
結果が一致することを確認しましょう!

## Stataコマンド

- ▶ Stataのコマンドは、命令が識別できる文字まで省略することも可能

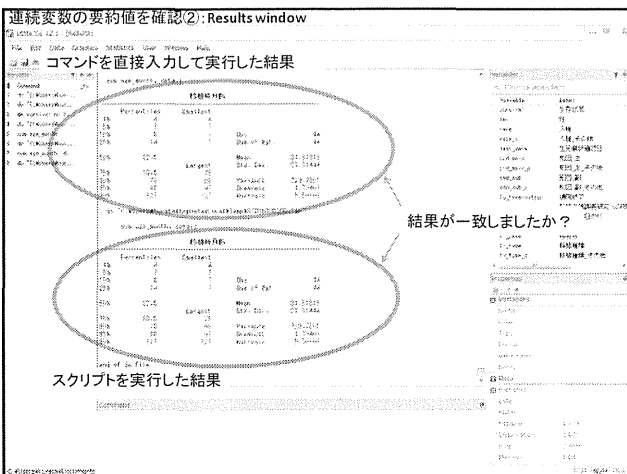
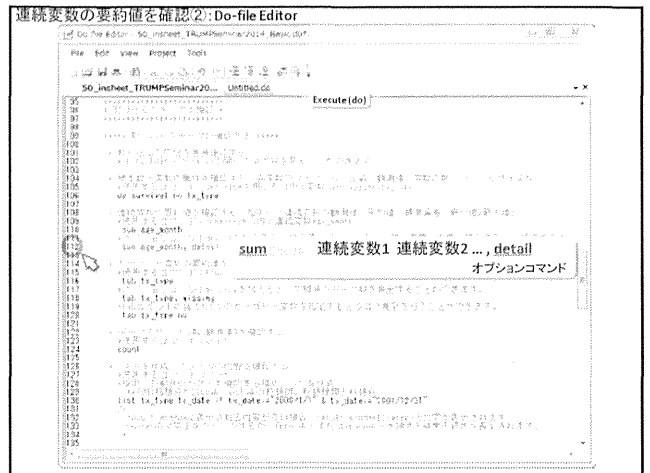
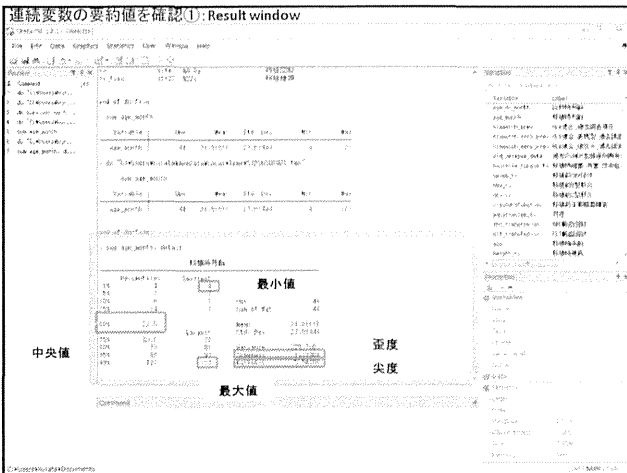
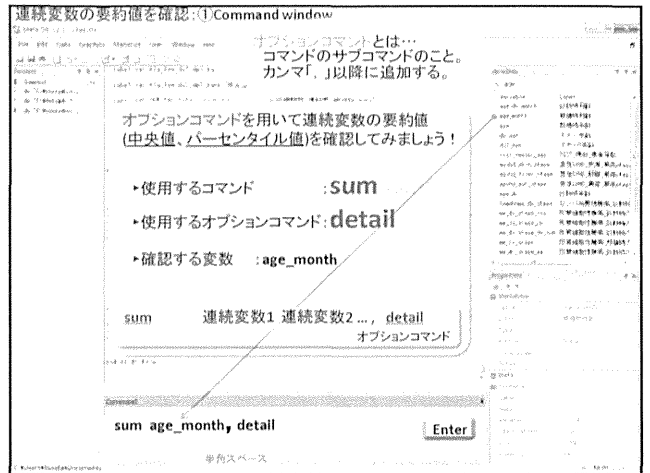
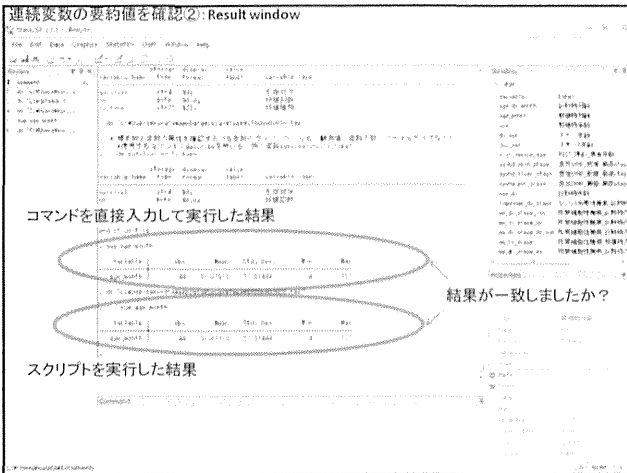
本日使用するコマンドの一例

`describe` → `de`  
`summarize` → `sum`  
`tabulate` → `tab`  
`list` → `l`  
`save` → `sa`









## 取り込んだデータの確認

- ▶ 標本数と変数の属性を確認  
(各変数のタイプ、ラベル名、観測値、変数の数など)
- ▶ 連続変数の要約値を確認  
(指定した連続変数の観測値、平均値、標準偏差、最小値/最大値)
- ▶ カテゴリー変数の要約値を確認  
(指定した) ①実際にコマンドを入力し実行  
②配布スクリプトを実行  
結果が一致することを確認しましょう！

### カテゴリ変数の要約値を確認①: Command window

Command windowに直接コマンドを入力してカテゴリ変数の要約値を確認してみましょう!

- 使用するコマンド: `tabulate`
- 確認する変数: `tx_type`

```
tab tx_type
```

Enter

### カテゴリ変数の要約値を確認①: Result window

tx_type	Freq.	Percent	Prb.
1	260	82.26	40.06
2	19	6.10	2.85
3	6	1.93	1.09
4	6	1.93	1.09
5	2	.64	.32
6	2	.64	.32
7	2	.64	.32
8	1	.32	.16
9	1	.32	.16
10	1	.32	.16
11	1	.32	.16
12	1	.32	.16
13	1	.32	.16
14	1	.32	.16
15	1	.32	.16
16	1	.32	.16
17	1	.32	.16
18	1	.32	.16
19	1	.32	.16
20	1	.32	.16
21	1	.32	.16
22	1	.32	.16
23	1	.32	.16
24	1	.32	.16
25	1	.32	.16
26	1	.32	.16
27	1	.32	.16
28	1	.32	.16
29	1	.32	.16
30	1	.32	.16

Summary Statistics

Category: tx\_type

Measure: Relative Frequency

Frequency: 260 (82.26%)

Percentage: 6.10%

Probability: 0.32

### カテゴリ変数の要約値を確認②: Do-file Editor

```

execute `do' "D:\stata\bin\stata17\stata17_64.exe"
clear all
use "D:\stata\bin\stata17\stata17_64.exe"
tab tx_type

```

Execute (do)

Category: tx\_type

Measure: Relative Frequency

Frequency: 260 (82.26%)

Percentage: 6.10%

Probability: 0.32

### 連続変数の要約値を確認②: Results window

コマンドを直接入力して実行した結果

tx_type	Freq.	Percent	Prb.
1	260	82.26	40.06
2	19	6.10	2.85
3	6	1.93	1.09
4	6	1.93	1.09
5	2	.64	.32
6	2	.64	.32
7	2	.64	.32
8	1	.32	.16
9	1	.32	.16
10	1	.32	.16
11	1	.32	.16
12	1	.32	.16
13	1	.32	.16
14	1	.32	.16
15	1	.32	.16
16	1	.32	.16
17	1	.32	.16
18	1	.32	.16
19	1	.32	.16
20	1	.32	.16
21	1	.32	.16
22	1	.32	.16
23	1	.32	.16
24	1	.32	.16
25	1	.32	.16
26	1	.32	.16
27	1	.32	.16
28	1	.32	.16
29	1	.32	.16
30	1	.32	.16

結果が一致しましたか?

Summary Statistics

Category: tx\_type

Measure: Relative Frequency

Frequency: 260 (82.26%)

Percentage: 6.10%

Probability: 0.32

スクリプトを実行した結果

### カテゴリ変数の要約値を確認①: Command window

オプションコマンドを用いてカテゴリ変数の欠損値を含む要約値を確認してみましょう!

- 使用するコマンド: `tabulate`
- 使用するオプションコマンド: `missing`
- 確認する変数: `tx_type`

```
tab tx_type, missing
```

Enter

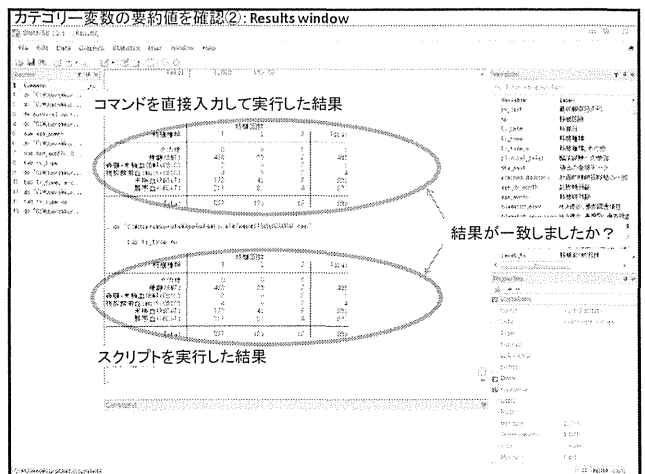
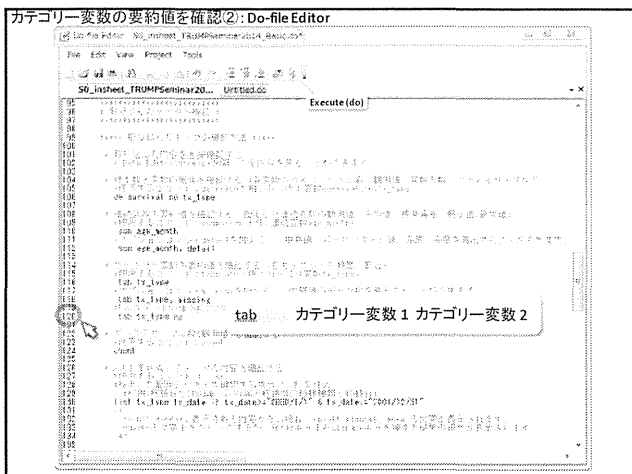
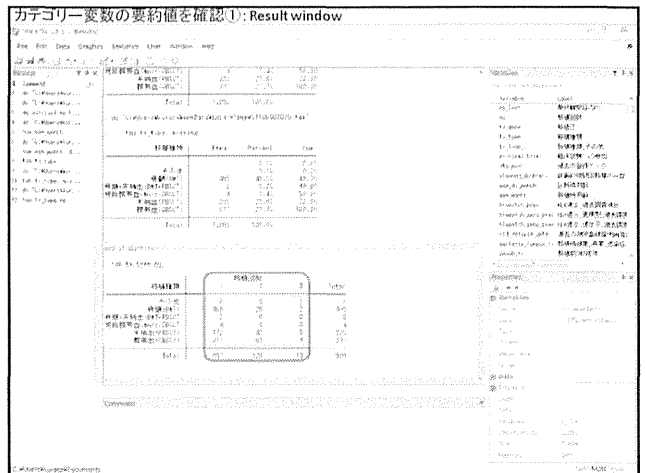
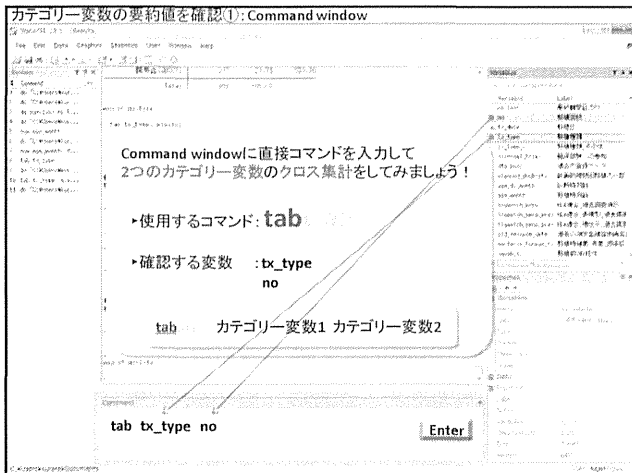
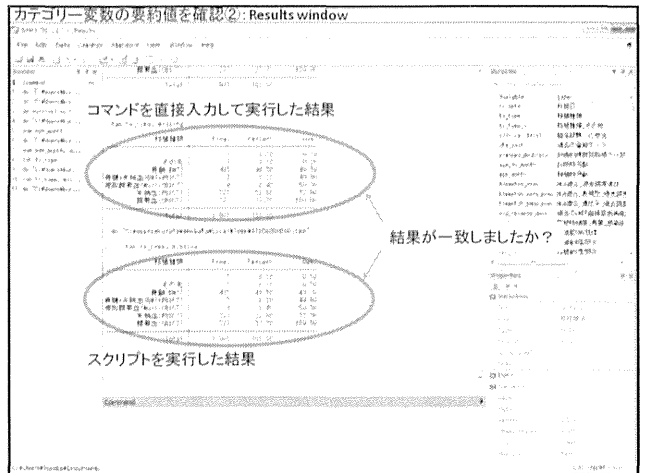
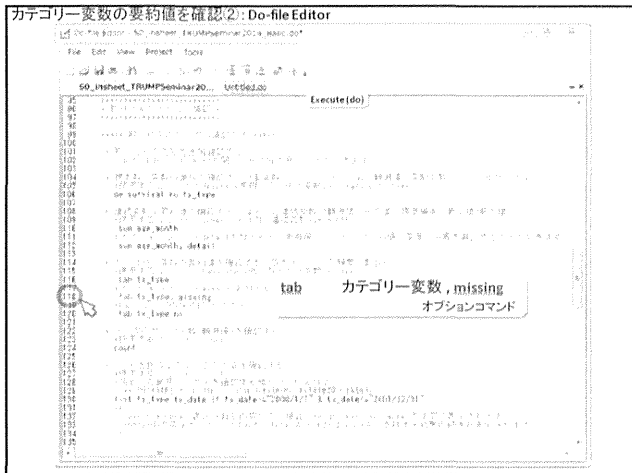
### カテゴリ変数の要約値を確認①: Results window

tx_type	Freq.	Percent	Prb.
1	260	82.26	40.06
2	19	6.10	2.85
3	6	1.93	1.09
4	6	1.93	1.09
5	2	.64	.32
6	2	.64	.32
7	2	.64	.32
8	1	.32	.16
9	1	.32	.16
10	1	.32	.16
11	1	.32	.16
12	1	.32	.16
13	1	.32	.16
14	1	.32	.16
15	1	.32	.16
16	1	.32	.16
17	1	.32	.16
18	1	.32	.16
19	1	.32	.16
20	1	.32	.16
21	1	.32	.16
22	1	.32	.16
23	1	.32	.16
24	1	.32	.16
25	1	.32	.16
26	1	.32	.16
27	1	.32	.16
28	1	.32	.16
29	1	.32	.16
30	1	.32	.16
31	1	.32	.16
32	1	.32	.16
33	1	.32	.16
34	1	.32	.16
35	1	.32	.16
36	1	.32	.16
37	1	.32	.16
38	1	.32	.16
39	1	.32	.16
40	1	.32	.16
41	1	.32	.16
42	1	.32	.16
43	1	.32	.16
44	1	.32	.16
45	1	.32	.16
46	1	.32	.16
47	1	.32	.16
48	1	.32	.16
49	1	.32	.16
50	1	.32	.16

欠損の数を含まない

欠損値データの数、割合

欠損の数を含み



データのサンプル数を確認: Command window

Command windowに直接コマンドを入力してデータのサンプル数を確認してみましょう!

▶使用するコマンド: **count**

count

Enter

データのサンプル数を確認: Results window

Results window showing the output of the 'count' command, displaying a table with columns for '移動種別', '移動種別', '移動種別', and 'Total'. The table contains data for various categories like '転居', '転居', '転居', etc.

データ一覧を表示: Command window

Command windowに直接コマンドを入力してデータの一覧を表示してみましょう!

2000年~2001年の移住例の、移住種別と移住日のリスト

▶使用するコマンド: **list**

▶確認する変数: **tx\_type**  
**tx\_date**

指定した範囲のリストを作成する

▶条件: **2000年~2001年の移住**

list 変数1 変数2 変数3 ... if 条件式

list 変数1 変数2 変数3 ... if 条件式

if以下構文に用いる演算子(一例)

>	より大きい	+	加算
<	より小さい	-	減算
>=	以上	*	乗算
<=	以下	/	除算
==	等しい	^	べき乗
!=	等しくない(not equal)		

& かつ (and)  
| または (or)  
! 否定

その他の演算子は、**help operator** のコマンドを入力し実行すると調べることができる。

help operator

help operator

Title

[M] 13.2 operators

Syntax

Arithmetic	Logical	Relational
+ addition	& and	> greater than
- subtraction	! not	< less than
* multiplication	! not	>= greater than or equal
/ division	! not	<= less than or equal
^ power		== equal
^ concatenation		!= not equal
		!= not equal

A double equal sign (==) is used for equality testing.

The order of evaluation (from first to last) of all operators is: (ex: +), ^, - (negation), /, \* (multiplication), &, ==, >, <, >=, <=, ==, &, and |.

Examples

```

sysuse auto
generate weight2 = weight^2
count if rep78 > 4
count if rep78 > 4 & weight < 3000
list make if rep78 == 5 & wgt > 25

```

list 変数1 変数2 変数3 ... if 条件式

例: 2000年~2001年の移住

>	より大きい	&	かつ (and)
<	より小さい		または (or)
>=	以上	!	否定
<=	以下		
==	等しい		
!=	等しくない(not equal)		

list 変数1 変数2 変数3 ... if tx\_date>="2000/1/1" & tx\_date<="2001/12/31"

データ一覧を表示: Command window

Command windowに直接コマンドを入力してデータの一覧を表示してみましょう!

2000年~2001年の移植種類と移植日のリスト

- 使用するコマンド: list
- 確認する変数 : tx\_type  
tx\_date
- 指定した範囲のリストを作成する
- 条件 : 2000年~2001年の移植

list 変数1 変数2 変数3 ... if 条件式

list tx\_type tx\_date if tx\_date >= "2000/1/1" & tx\_date <= "2001/12/31"

Enter

リストを作成しデータの内容を確認: Results window (1)

移植日 (tx\_date) と移植種類 (tx\_type) のリストが表示されています。

このテキストの文字上を **クリック** または **Enter** または **Space** または **Tab** を押すと、結果の続きが表示される

リストを作成しデータの内容を確認: Results window (2)

移植日 (tx\_date) と移植種類 (tx\_type) のリストが引き続き表示されています。

データ一覧を表示: Command window

if 構文を用いて、条件を満たすデータのサンプル数を確認してみましょう!

移植時年齢60歳以下の男性かつ、同種移植の件数

- 使用するコマンド: count
- 確認する変数 : age  
sex  
donor
- 指定した条件のサンプル数を算出
- 条件 : age (60以下)  
sex (男)  
donor (同種\_allogeneic)

count if 条件式

count if age <= 60 & sex == "男" & donor == "同種\_allogeneic"

Enter

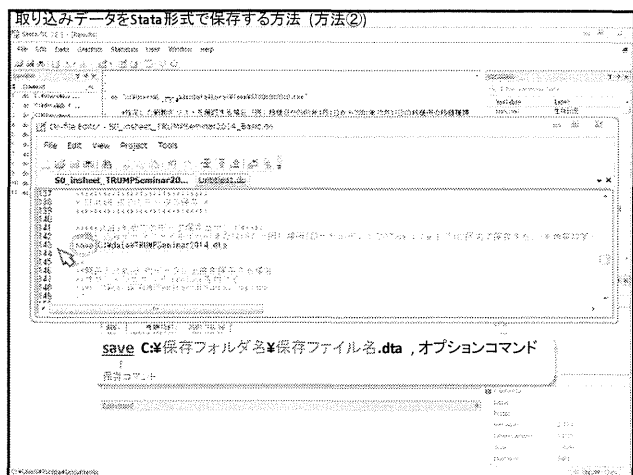
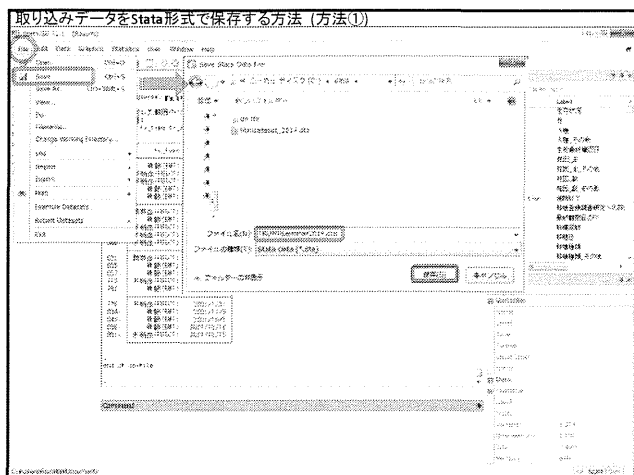
条件を満たすデータのサンプル数を確認: Results window

移植時年齢60歳以下の男性かつ、同種移植の件数は 10 件と表示されています。

## 取り込んだデータの保存

ExcelまたはCSVファイルからStataに取り込んだデータをStata形式で保存する。

▶Stata形式ファイル(.dta)での保存

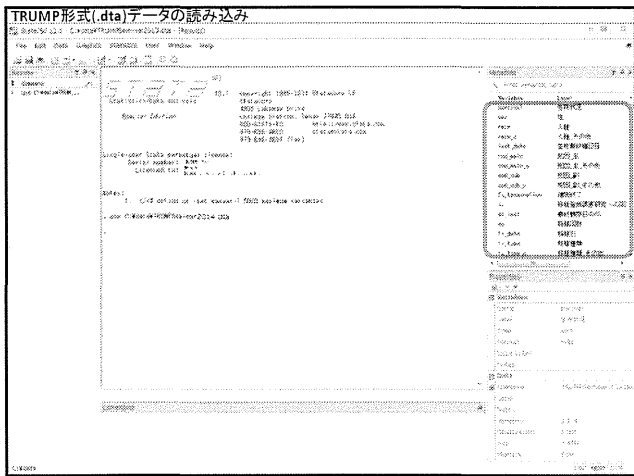


### TRUMP形式(.dta)データの読み込み

Stata形式(.dta)のデータを開く

- ▶フォルダ内に保存されたファイル(.dta)を直接クリックして開く
- ▶ファイル(.dta)の読み込みコマンドにて開く

### TRUMP形式(.dta)データの読み込み



### ログファイルの作成

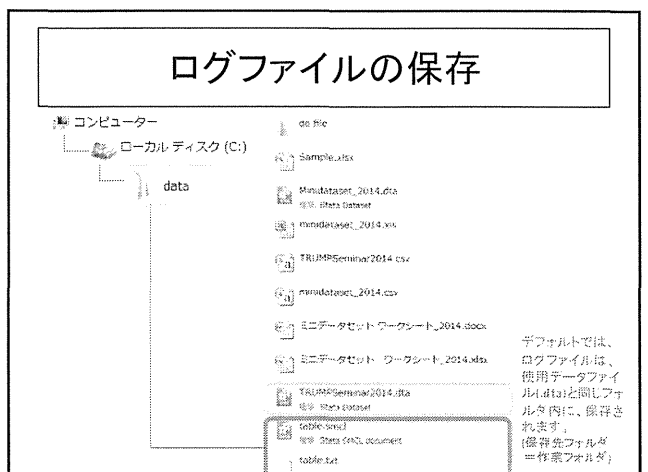
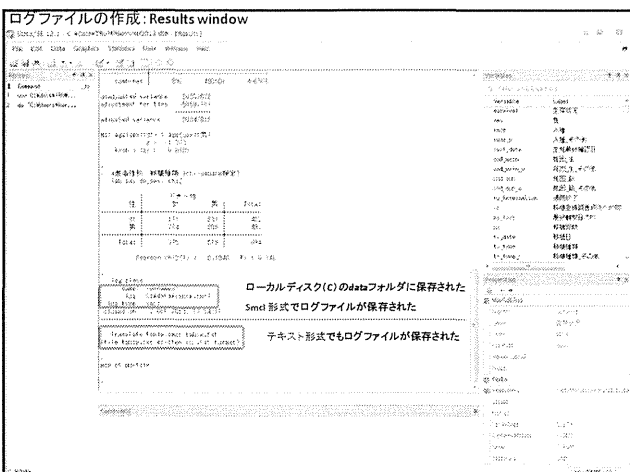
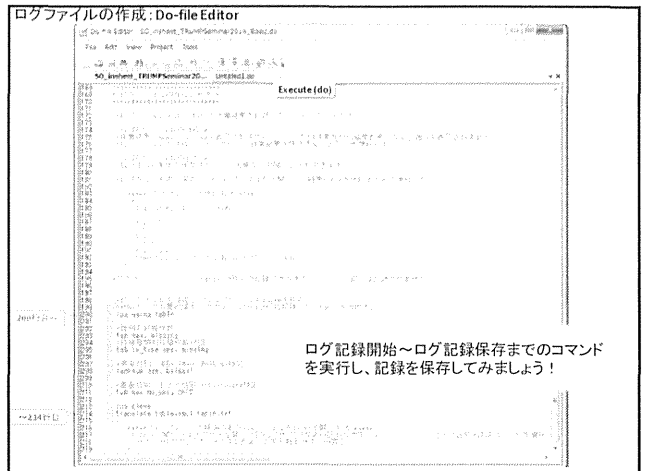
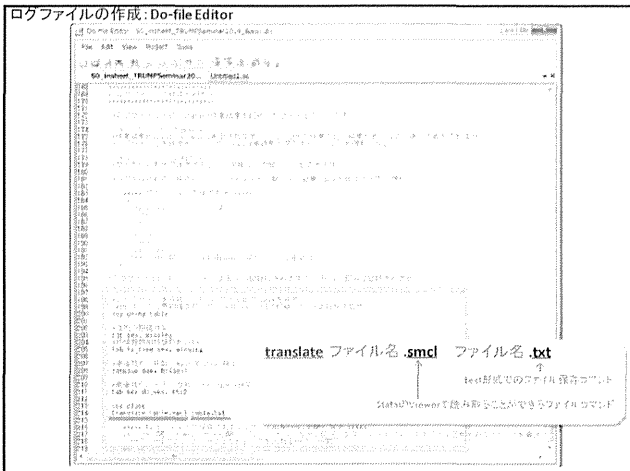
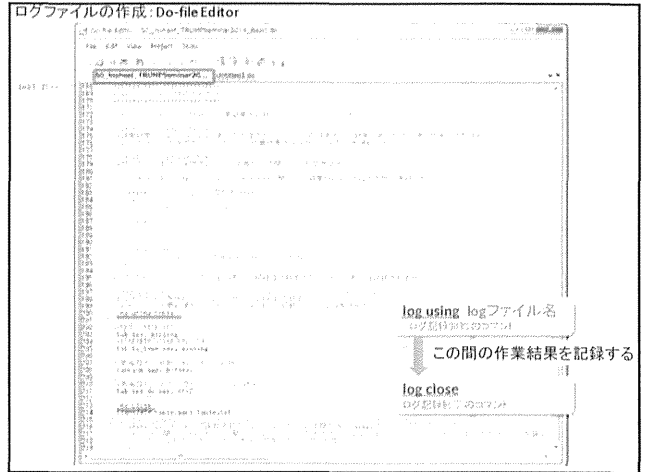
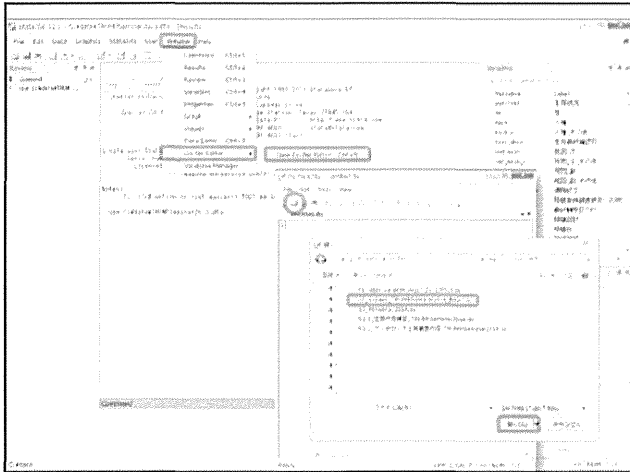
**ログファイルとは…**

Stataによる作業結果を記録したファイル

コマンド実行などの作業結果はResult windowに表示されるが、いくつかの作業を行うと結果が長くなり、すべて遡って表示されない。(最初の結果から消去されていく)

**ログファイルを作成すると、**

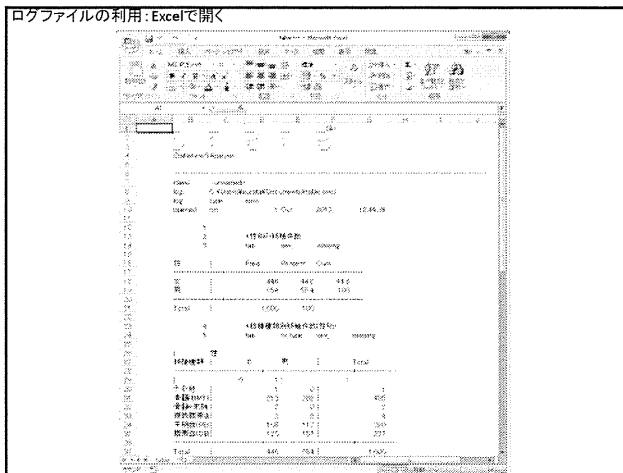
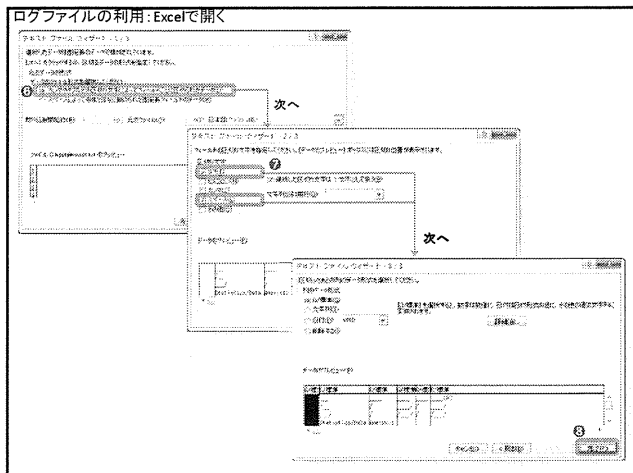
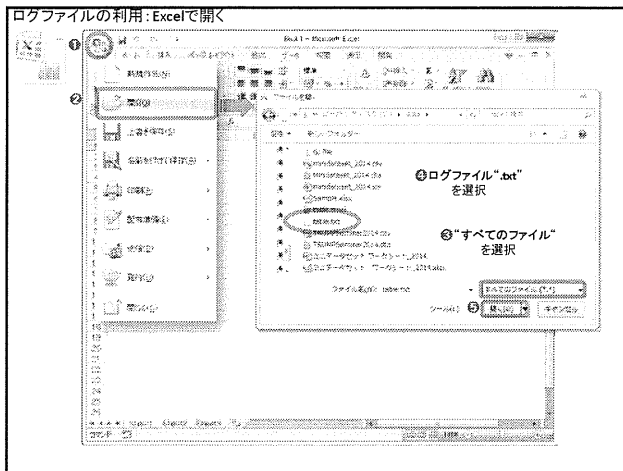
- ▶すべての結果を保存可能
- ▶text形式で保存し、Excelファイルから開くと、結果tableを作成するのに便利



## ログファイルの利用

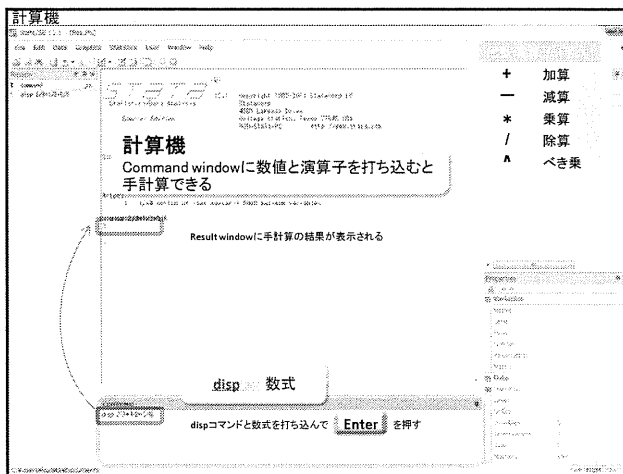
ログファイルを作成すると、

- ▶すべての結果を保存可能
- ▶text形式で保存し、Excelファイルから開くと、結果tableを作成するのに便利



## その他、Stataの便利機能

▶計算機





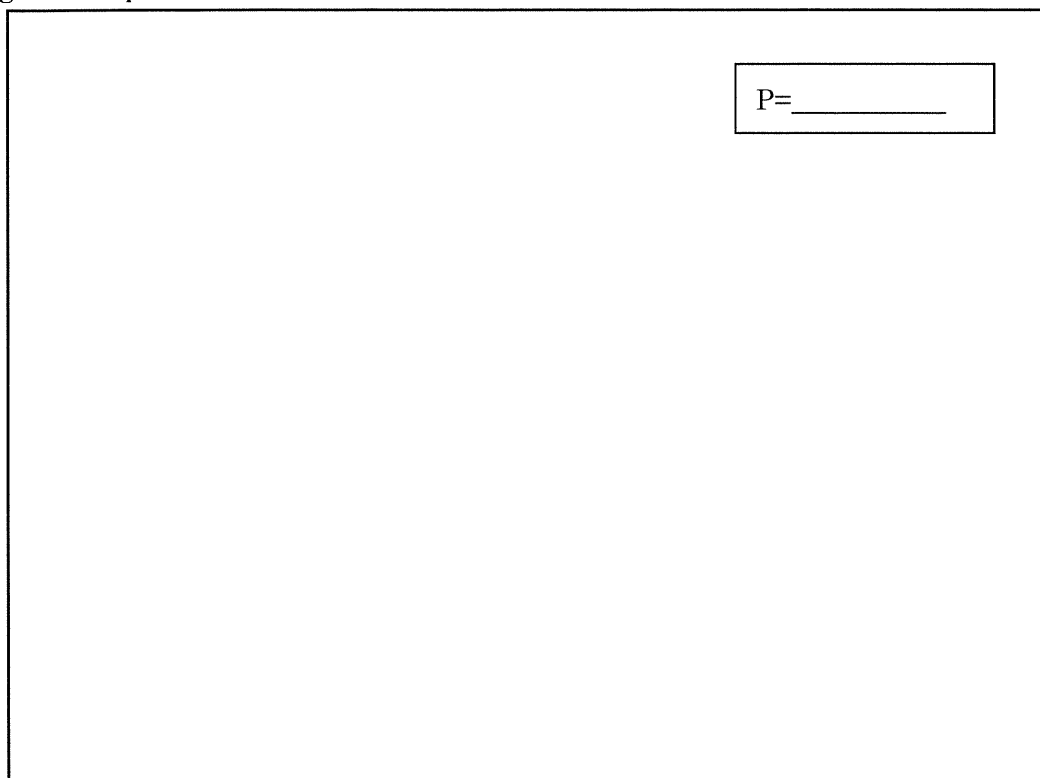
模擬データセットを用いた変数と背景表の作成、生存曲線の描出(単群)

ミニデータセット ワークシート

**Table 1. Characteristics of study patients**

Characteristic	HLA-matched	HLA-mismatched	P value
	N (%)	N (%)	
Total			
Age at transplant, years, median (range)	( - )	( - )	
Gender			
Male	( %)	( %)	
Female	( %)	( %)	
Disease			
AML	( %)	( %)	
ALL	( %)	( %)	
Year of transplant			
2000-2005	( %)	( %)	
2006-2010	( %)	( %)	

**Figure 1. Kaplan-Meier Survival Curve**



平成26年度 TRUMP統計セミナー 基礎編  
2014年9月13日

## 変数作成の基本的事項

名古屋大学附属病院  
先端医療・臨床研究支援センター  
鋸塚八千代

1

## このセクションの目標

- 50例、少数の項目のデータ
- Stataを使用した簡単なデータクリーニング
- 変数作成
- 多変量解析前のデータ準備

2

## 配布資料

- ① Minidataset\_2014 (Stata Dataset)
- ② minidataset\_2014 (Excel)
- ③ minidataset\_2014 (CSV)
- ④ S1\_minidata\_2014  
(Stata Do-file・プリント)
- ⑤ Minidataset\_WS\_2014  
(Excel・プリントワークシート)

3

## minidataset\_2014.xls

1	2	3	4	5	6	7	8	9	10	11	12
sex	bsdt_18yr	ts_18yr	age	gender	diagnosis	diagnosis_5	serumlaboz				
1	男	2001/8/18	2001/1/26	11	閉鎖_alkog	男性リンパ	2	1	0		
2	男	2011/7/11	2002/8/9	27	閉鎖_alkog	男性骨髄	1	4	2		
3	女	2007/7/21	2007/3/20	5	閉鎖_alkog	男性リンパ	2	4	1		
4	女	2011/8/2	2009/8/19	5	閉鎖_alkog	男性リンパ	2	1	0		
5	男	2011/8/1	2010/10/1	10	閉鎖_alkog	男性骨髄	1	2	0		
6	男	2007/5/7	2006/5/29	16	閉鎖_alkog	男性骨髄	1	4	0		
7	男	2005/12/24	2004/5/21	10	閉鎖_alkog	男性骨髄	1	1	0		
8	男	2004/11/1	2004/10/1	6	閉鎖_alkog	男性骨髄	1	1	0		
9	男	2013/8/17	2008/1/25	21	閉鎖_alkog	男性リンパ	2	1	0		
10	女	2007/11/2	2007/3/6	13	閉鎖_alkog	男性骨髄	1	1	0		
11	女	2001/2/13	2000/10/20	14	閉鎖_alkog	男性骨髄	1	1	0		
12	女	2009/3/11	2007/10/3	2	閉鎖_alkog	男性リンパ	2	4	0		
13	男	2001/3/13	2000/9/13	10	閉鎖_alkog	男性骨髄	4	4	0		
14	女	2007/1/26	2003/1/26	3	閉鎖_alkog	男性リンパ	2	2	1		
15	男		2003/9/10	7	閉鎖_alkog	男性リンパ	2	4	0		
16	女	2007/8/13	2001/5/16	20	閉鎖_alkog	男性リンパ	3	4	1		
17	女	2008/6/12	2006/5/16	18	閉鎖_alkog	男性リンパ	2	1	1		
18	男	2004/10/1	2004/3/31	1	閉鎖_alkog	男性リンパ	2	4	0		
19	男	2002/2/2	2002/1/19	3	閉鎖_alkog	男性リンパ	2	1	0		
20	女	2011/4/6	2010/12/2	14	閉鎖_alkog	男性リンパ	2	1	1		

## Minidataset\_2014.dta

## S1\_minidata\_2014.do Stata Do-file & 印刷資料

```

Do-file Editor

File Edit View Project Tools

S1_minidata_2014 Urnt5ac.do
1 *** ***** NOTICE *****
2 ***
3 *****
4 *****
5 *****
6 *****
7 *****
8 *****
9 *****
10 *****
11 *****
12 *****
13 *****
14 *****
15 *****
16 *****
17 *****
18 *****
19 *****
20 *****
21 *****

```

### 配布ファイルの保存場所

Windows

Macintosh

\* Macをご使用の場合は、フォルダCを作成の上、保存してください。

フォルダの保存先が上記と異なる場合、本日のセミナーで使用するプログラム、コマンド内容を変更していただく必要があります。

配布ファイルを、直接コピーペーストして C ディスク、dataフォルダ内に保存してください。

### このセクションのゴール

- Table 1  
患者背景表の作成
- Figure 1  
OS curveの作成

ゴール

### データセットの作成 S1¥Minidata\_2014 34行目~

```

**** 解析データセットの作成 ****
*Stataを起動して、ファイルの中にある、minidataset_2014.xls
を取り込んでください。

*Mac使用者は、Stataを起動して、ファイルの中にある、
minidataset_2014.csv を取り込んでください。

```

\*あるいは、配布DVD/USB内の"Minidataset\_2014"をダブルクリックして開く

資料37行目~  
\*配布DVD/USB内の"Minidataset\_2014"をダブルクリックして開く

- コマンドを使用して開く代わりに、クリックで開くことも可能です。
- 配布資料 Minidataset\_2014.dta をクリック

```

Notes:
. use C:\data\Minidataset_2014.dta

```

ファイルのクリックでデータを取り込んだ場合

```

Command

```

### 解析データセットの作成(コマンド使用する場合)

Win: S1¥Minidata\_2014: 40行目

```
import excel "C:\data\minidataset_2014.xls", firstrow
```

Mac: S1¥Minidata\_2014: 42行目

```
insheet using "C:\data\minidataset_2014.csv", name
```

```
import excel "C:\data\minidataset_2014.xls", firstrow
```

### エクセルデータ取り込み (コマンド使用)

S1¥Minidata\_2014: 40行目

```

Notes:
. log using "C:\data\min(20131020_sec)"
. use "C:\data\min(20131020_sec)"
. log type: SRG

```

```
import excel "C:\data\minidataset_2014.xls", firstrow
```

エラーは出ていませんか?