

研究 番号	課題名	PI	学会発表	論文	業績 番号
18-8	非血縁者間臍帯血移植と GVH 方向 1 抗原以内不適合血縁者間移植の移植成績の比較	諫田淳也	JSH2012	Leukemia 2013;27:286	WGP 18-1
18-9	ABO 血液型不適合が同種移植成績に与える影響—移植細胞ソースによる違い	木村文彦	JSH2012		
18-10	重症再生不良性貧血に対する血縁者間造血細胞移植成績の国際比較	木村文彦	EHA2013		
18-12	HLA 一致血縁ドナーからの成人造血悪性腫瘍に対する骨髄破壊的前処置による同種造血幹細胞移植 移植ソース 骨髄と末梢血幹細胞の比較	長藤宏司	JSH2012		
19-1	シクロスポリンおよびタクロリムスによる GVHD 予防法の比較検討	酒井リカ	EBMT2012,他		
19-2	血液悪性腫瘍に対する同種造血細胞移植における抗リンパ球グロブリンの臨床的検討	加藤剛二	JSH2012,他		
19-5	既存データを用いた年齢別の急性 GVHD 発症後の予後の検討	中根孝彦	JSH2012,他		
19-6	急性 GVHD に対するステロイド一次治療の成績	村田誠	JSHCT2012,他	BBMT 2013;19:1183	WGP 19-1
19-7	GVHD と TMA の関連性の検討	吾郷浩厚	JSH2012		
19-8	既存データを用いた臓器別慢性 GVHD の発症様式、発症頻度、予後の解析	諫田淳也	JSH2012	BMT (in press)	WGP 19-3
19-9	一元化管理事業データに基づく造血幹細胞移植後の閉塞性細気管支炎の解析	仲宗根秀樹	JSHCT2012	Transpl Int 2013;26:631	WGP 19-2
19-10	GVHD と GVL 効果に対するドナーとレシピエントの性別の影響	大島久美	Tandem2012, 他		
20-1	造血幹細胞移植後サイトメガロウイルス感染症の発症頻度、危険因子、予防法に関する研究	西田徹也	JSHCT2013		
20-2	非血縁者間移植患者におけるウイルス感染症の検討	森有紀	JSH2013		
20-3	同種造血幹細胞移植後の出血性膀胱炎 (HC) に対する標準的予防法・早期治療法の確立に向けた抗ウイルス薬の HC 発症抑制効果に関する検討	中沢洋三	JSHCT2013		
20-4	造血幹細胞移植後合併症と長期予後に与える HCV 既感染の影響	仲宗根秀樹	JSHCT2012,他	Am J Hematol 2013;88:477	WGP 20-1
20-5	同種造血幹細胞後の深在性真菌症に関する検討	大島久美	JSH2012,他		
20-6	一元化管理事業データに基づく同種造血幹細胞移植後の器質化肺炎 (COP/BOOP) の解析	仲宗根秀樹	JSH2012	BMT 2013;48:1317	WGP 20-2

TRUMP 統計セミナー資料集 (基礎編)

* 模擬プロトコール・変数表・Stata 簡易マニュアルは省略する

厚生労働科学研究費補助金難治性疾患等克服研究事業(免疫アレルギー疾患等予防・治療研究事業
移植医療研究分野)「本邦における造血細胞移植一元化登録研究システムの確立」

平成 25 年度 TRUMP 統計セミナー 基礎編

日時： 平成 25 年 10 月 20 日 (日) 10:00~16:00

場所： 名古屋大学大学院医学系研究科・医学部保健学科 東館 4F 大講堂
〒461-8673 名古屋市東区大幸南一丁目 1 番 20 号

《プログラム》

9:30~ 受付、資料配布

10:00-10:30

TRUMP データを用いた移植登録研究を実施する際の注意点

(名古屋大学大学院医学系研究科 造血細胞移植情報管理・生物統計学 鈴木 律朗)

10:30-11:30

Stata の基本動作

(日本造血細胞移植データセンター 倉田 美穂)

11:30-13:00

変数の作成の基本的事項

(名古屋大学医学部附属病院 先端医療・臨床研究支援センター 鍬塚 八千代)

13:00-14:00

昼食

14:00-16:00

模擬プロトコールに基づいたデータマニピュレーション・統計解析(単変量解析)の実際

(名古屋大学大学院医学系研究科 造血細胞移植情報管理・生物統計学 熱田由子)

TRUMP データを用いた移植登録研究を実施する際の注意点

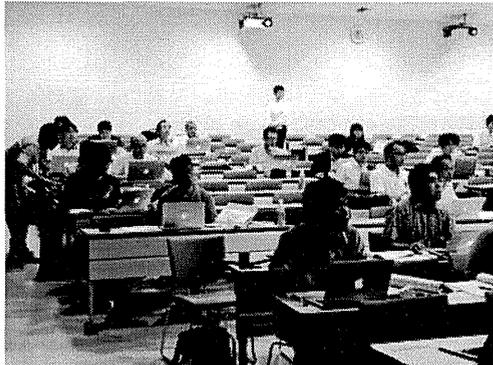
2013年10月20日(日)
名古屋大学大幸キャンパス

TRUMPデータを用いた移植登録研究を実施する際の注意点

名古屋大学大学院医学系研究科
造血細胞移植情報管理・生物統計学

鈴木律朗

今回は第2回です (基礎編)

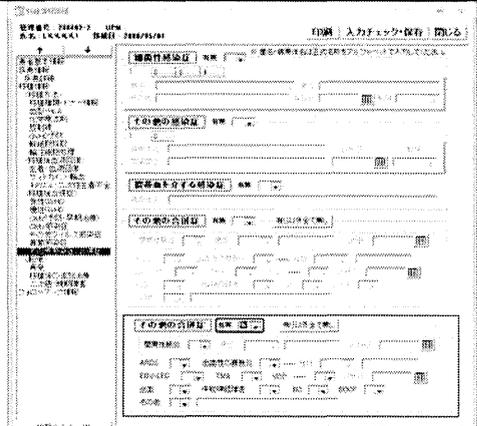


前回： 2012. 9. 3. TRUMP統計セミナー @ 名古屋大学
今回は、参加者の年齢層が若くてよかったです。 → 講師も若くしました。

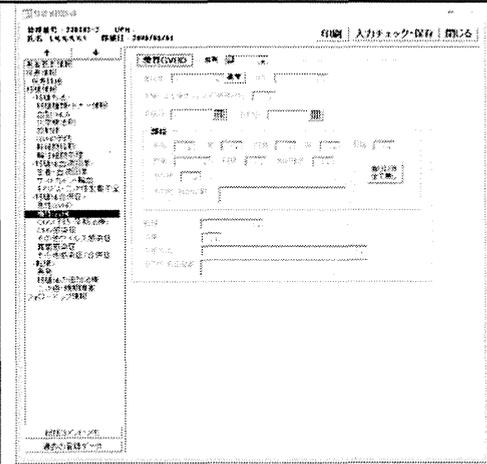
注意点

データの収集・集積状況が一様でない

- ◆ 登録時期(年)によって、調査項目が異なる
過去の報告書に調査項目が載っているが、調べるのが面倒。
- ◆ 条件を満たした場合にのみ、入力される項目がある
入力の際には便利だが、解析の時は要注意



「その他の合併症」は、「有」の場合のみ入力できる

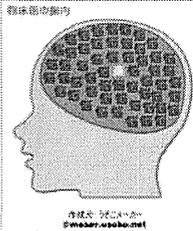


対策

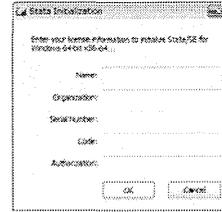
- ◆ “大発見” “説明できない結果” に注意
(例) VODの有無で、予後に差はない。
「VOD有」 ⇔ 「VOD無」
実は、「VOD有」 ⇔ 「VOD無、他の合併症有」
 << 「合併症無(当然VOD無)」
- ◆ 集計コマンドによる予備解析で、関連を把握
sum (summarize) コマンド
tab (tabulate) コマンド

一般的な臨床医の頭の中・・・

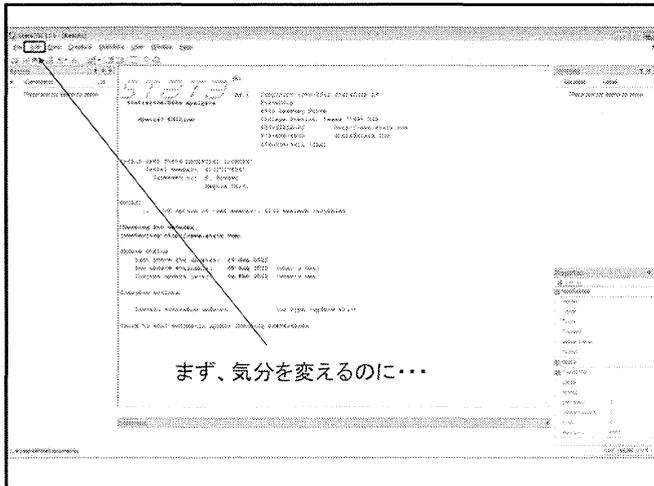
- 世の中には、統計ソフトというものがあるらしい。
- どうも、そいつを使って“統計解析”しないと、論文はアクセプトされないそうだ。
- 造血細胞移植学会WGで配られたデータも、やたら数が多いが、そういうのを使えば何とかなるらしい。
- WGで、タダでくれるというからもらった。
- Stata っていうらしいが、立ち上げたけどよく判らない。→ そのまま放置
- EZR というのもあるらしいが、もっとよく判らない。



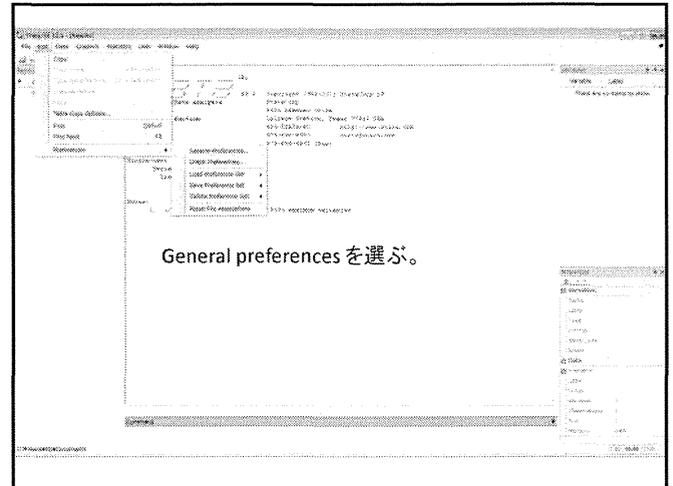
今日、Stata を立ち上げたら・・・



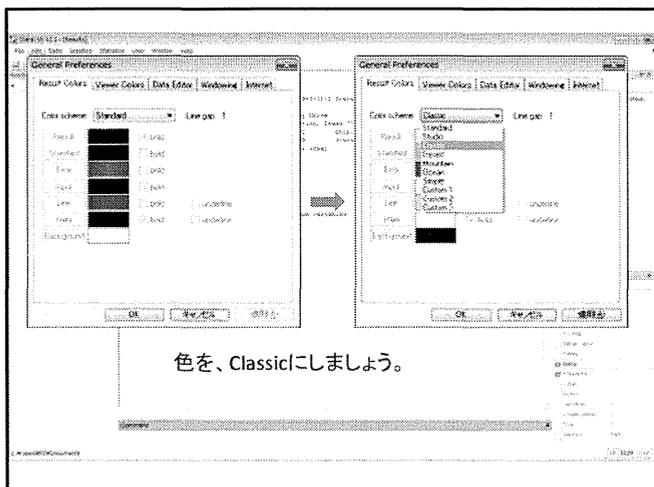
これが出た。



まず、気分を変えるのに・・・



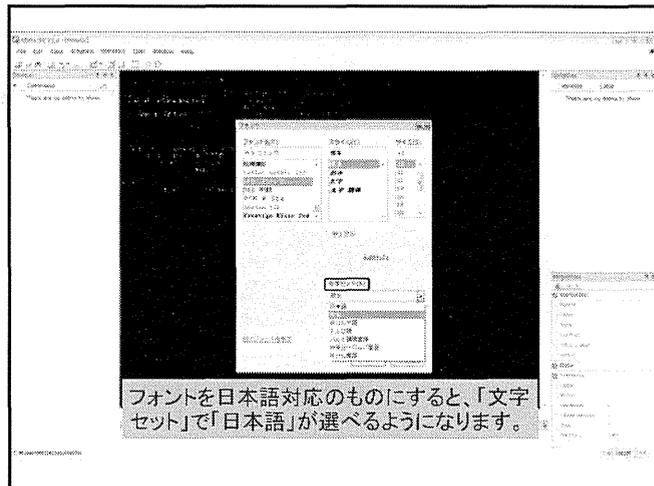
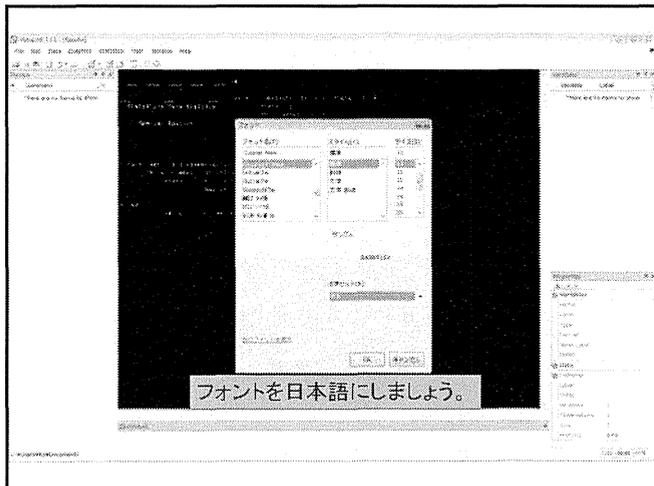
General preferences を選ぶ。



色を、Classic にしましょう。



ついでに、画面を右クリックして、



まずは、アップデートしましょう

ネットにつないで、以下のコマンドをタイプ。

update all

ado ファイルと呼ばれる、新しいコマンドファイルをインストールします。これをしておかないと、コマンドが使えなかったりプログラムが動かなくなったりします。

コマンドは、基本的にタイプして使った方が分かりやすいです。

本日、ここではネットに繋がせませんから、家に帰って行ってください。

Immediate command を使ってみよう

Immediate command

= コマンドの最後に i のつくコマンドで、データファイルを使わずに直接データを入力するコマンドです。

- データが何もなくて、Stataを立ち上げたらすぐに使えるので便利。
- 電卓的に使える。
- cii, tabi, csi, cci, iri など。
- 臨床試験の症例数も簡単に算定できる (sampsi コマンド)

各 command の意味

- cii confidence interval immediately
- tabi tabulate immediately
- csi chi square immediately
- cci case control immediately
- iri incidence rate immediately
- sampsi sample size

信頼区間

急性GVHDの治療を12例にしたたら、7例で有効であった。発表に使うので、有効率の95%信頼区間を求めたい。

以下のコマンドをタイプ。

cii 12 7

```

.cii 12 7
+-----+-----+
Variable | Obs   Mean   Std. Err.      +-----+-----+
+-----+-----+-----+-----+-----+-----+
          | 12   .5833333 .1429188      +-----+-----+
          |                                     .2766897   .8483478
+-----+-----+-----+-----+-----+-----+

```

58.3% (95% CI: 27.7-84.8%)

90%信頼区間

以下のコマンドをタイプ。

cii 12 7, level(90)

```
. cii 12 7, level(90)
```

Variable	Obs	Mean	Std. Err.	Binomial Exact [90% Conf. Interval]	
	12	.5833333	.1423188	.3152376	.8189752

58.3% (95% CI: 27.7-84.8%)
58.3% (90% CI: 31.5-81.9%)

群間比較

Aという化学療法を25例に実施したら、20例が寛解に入った。
以前のB治療では、53例中寛解は29例であった。比較したい。

以下のコマンドをタイプ。

csi 20 5 29 24

```
. csi 20 5 29 24
```

	Exposed	Unexposed	Total
Cases	20	5	25
Non-cases	29	24	53
Total	49	29	78
Risk	.4081633	.1724138	.3205128
Point estimate [95% Conf. Interval]			
Risk difference	.2357495	.0412269	.430272
Risk ratio	2.380382	1.300000	1.500000
Attr. frac. exp.	.6726882	.4081633	.5402603
Attr. frac. pop.	.462069	.2927294	.3773967

chi2(1) = 4.65 Pr>chi2 = 0.0311

群間比較

または、

tabi 20 5 ¥ 29 24, chi2

```
. tabi 20 5 ¥ 29 24, chi2
```

row	col	1	2	Total
1		20	5	25
2		29	24	53
Total		49	29	78

Pearson chi2(1) = 4.6492 Pr = 0.031

χ²乗検定をしています。

群間比較

一か所、5という数があるので、Fisherの直接確立検定をしたいという場合は..

以下のコマンドをタイプ。

csi 20 5 29 24, exact または
tabi 20 5 ¥ 29 24

```
. csi 20 5 29 24, exact
```

	Exposed	Unexposed	Total
Cases	20	5	25
Non-cases	29	24	53
Total	49	29	78
Risk	.4081633	.1724138	.3205128
Point estimate [95% Conf. Interval]			
Risk difference	.2357495	.0412269	.430272
Risk ratio	2.380382	1.300000	1.500000
Attr. frac. exp.	.6726882	.4081633	.5402603
Attr. frac. pop.	.462069	.2927294	.3773967

1-sided Fisher's exact P = 0.0283
2-sided Fisher's exact P = 0.0442

```
. tabi 20 5 ¥ 29 24
```

row	col	1	2	Total
1		20	5	25
2		29	24	53
Total		49	29	78

1-sided Fisher's exact P = 0.0283
2-sided Fisher's exact P = 0.0442

結果を残しておきたい場合

(正式)

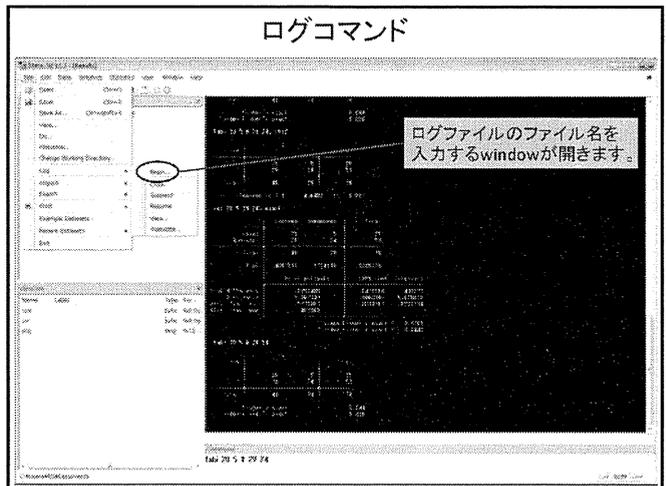
Stataを立ち上げた時、ログコマンドを用いる。
立ち上げた時でなくてもよいけど、「後で」と思っていると忘れる。

(裏技)

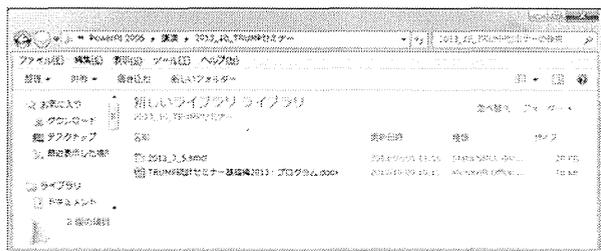
画面を全部選んで、Wordにコピーして保存する。

一長一短があります。

ログコマンド

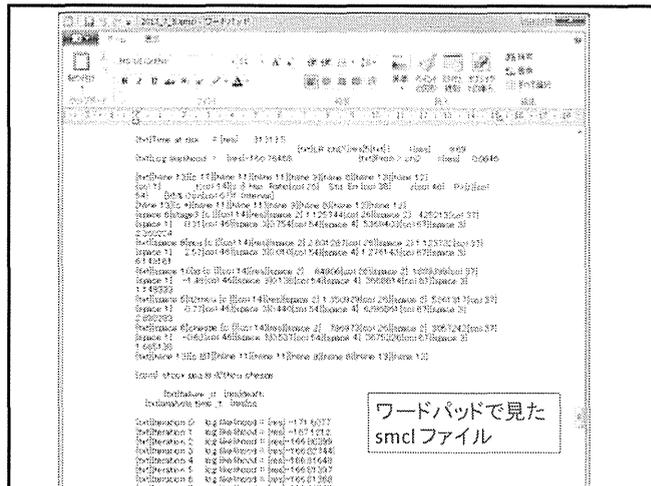


Stata ログファイルは .smcl ファイルになります

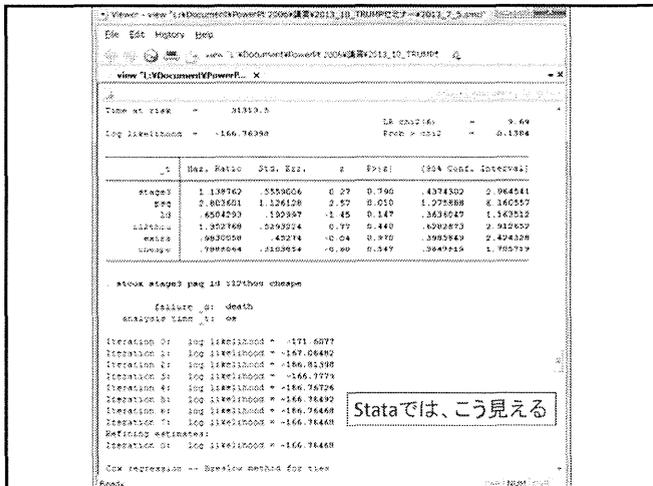


ダブルクリックすると、Stataが立ち上がって内容を見ることができですが、中身は text file です。

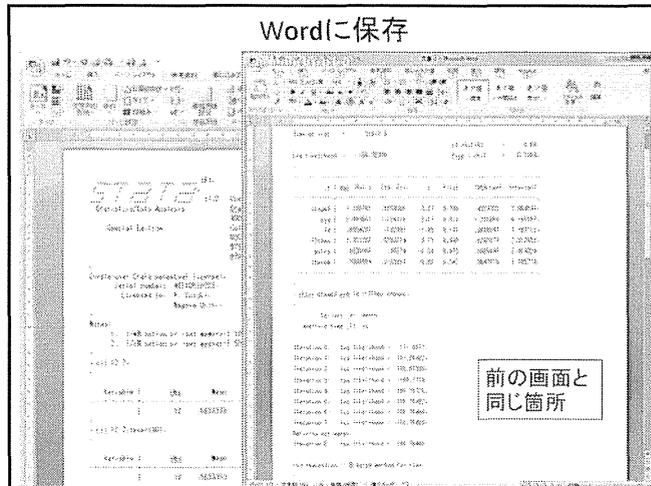
でも、通常の text reader では、Stata の 描画コマンド (line など) が邪魔をして、かなり慣れないと理解不能です。



ワードパッドで見た smcl ファイル



Stataでは、こう見える



Wordに保存

前の画面と同じ箇所

Word でのログファイルの保存

- 後からできるので、ものぐさ(=悪い子)に適している。
- 他のテキスト形式でも可。
- Stataとは別画面で見られる。
- コマンド行をStataにコピーして、動作を再現することが可能。
- ただし、何千行も解析する時は、最初の結果が失われることもあるので、十分なバッファ量を確保しておくか、正式なログコマンドを使いましょう。
- Stataがインストールされていない別のパソコンで見るとは、分かりやすい。

Stata の基本動作

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を作成のうえ、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。

Stataへデータを取り込む前の準備

データセットの前準備

- ▶ファイル名は必ず**英語名**であること
- ▶項目名が**英語名**のデータセットを使用する
- ▶WGデータセットを使用する場合、
開封パスワードの設定を解除のうえ、保存したデータセットを使用する

各種window (デフォルト)

Review window
実行したコマンドの履歴が表示される

Result window
集計/解析結果など実行された操作の結果が表示される

Command window
実行コマンドを入力する
1つのコマンドを入力
→Enter keyにて実行
→その都度、結果表示

Variables window
取り込んだデータの項目名の一覧が表示される

Properties window
変数のタイプ、データの数などが表示される

表示フォントの設定

Stataは標準設定の状態では、
変数名、コメントに日本語を使用する場合
文字化けが生じる

- ▶日本語を表示する設定に変更
(windowごとに設定の変更が必要)

データの読み込み

Stataにて読み込み可能なデータ形式

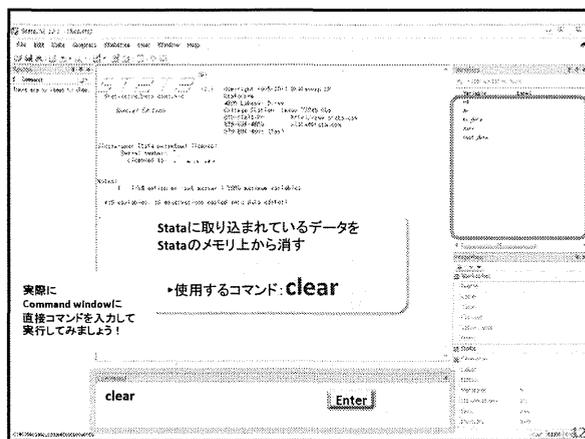
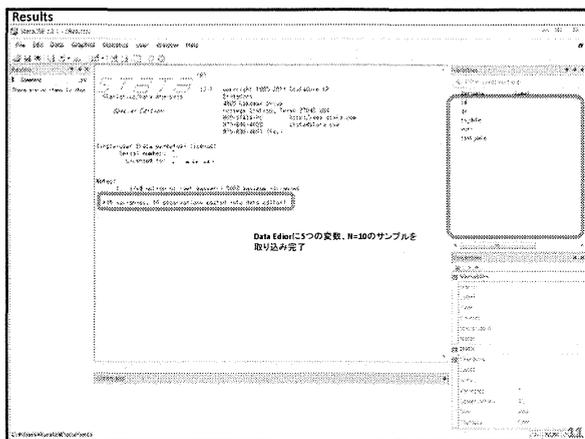
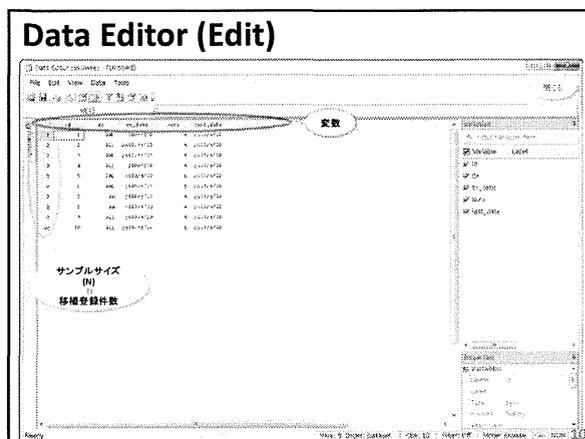
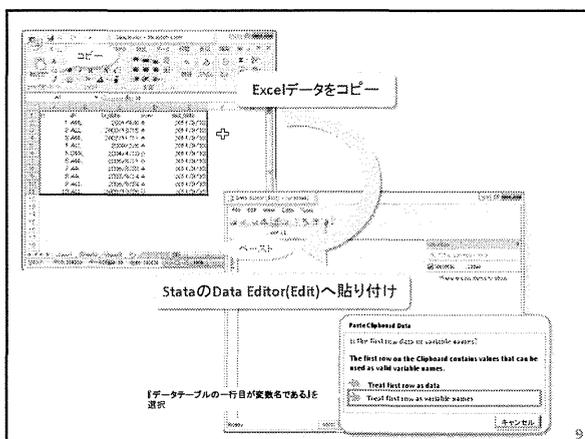
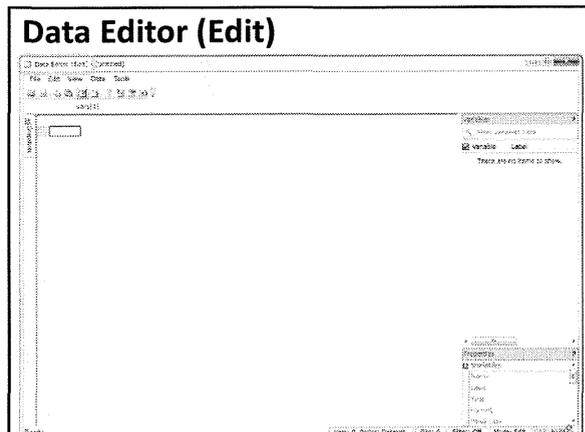
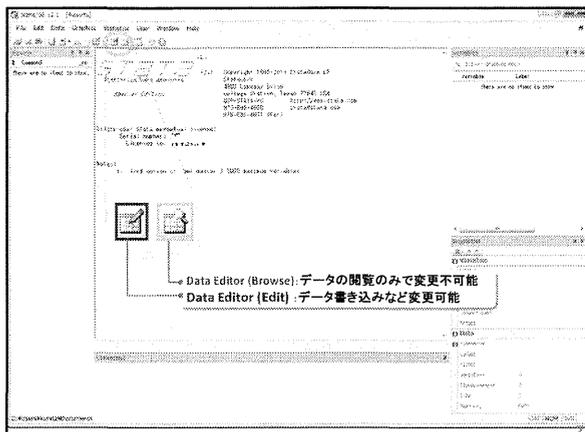
- ▶ **Excelファイル (.xls または .xlsx)**
Stata12以降ではExcelファイルからの取り込みが可能
ただし、40メガバイトを超えるExcelファイルからの取り込みは不可
- ▶ **CSVファイル (.csv)**
ファイルサイズが40メガバイトを超える場合は、Excelファイルからの取り込みは不可 ***CSVファイルから取り込む
(横断的WGでのフルデータなど)
- ▶ **Stataファイル (.dta)**
ExcelまたはCSVファイルからStataに取り込んだデータをStata形式で保存したファイルを読み込む

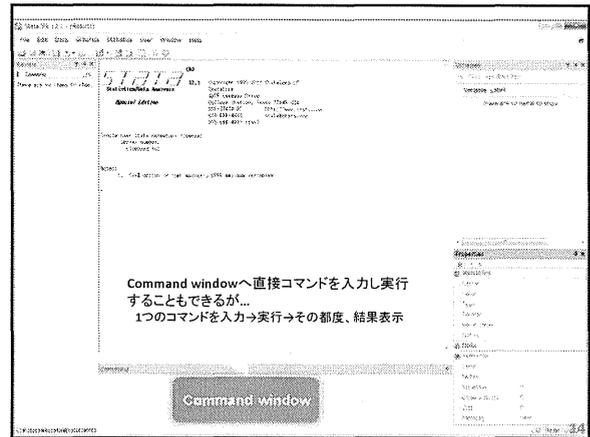
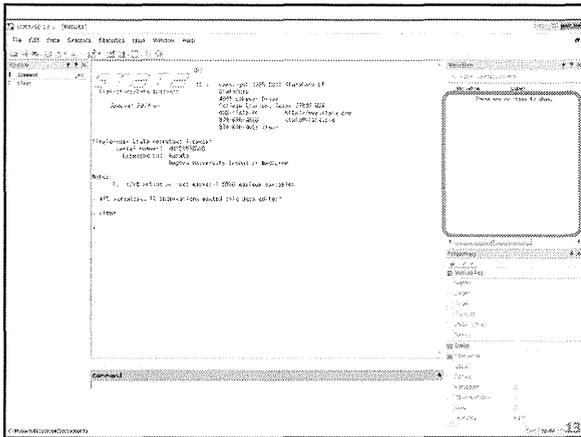
Excelデータの貼り付け

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





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ホームページよりダウンロード可能

15

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

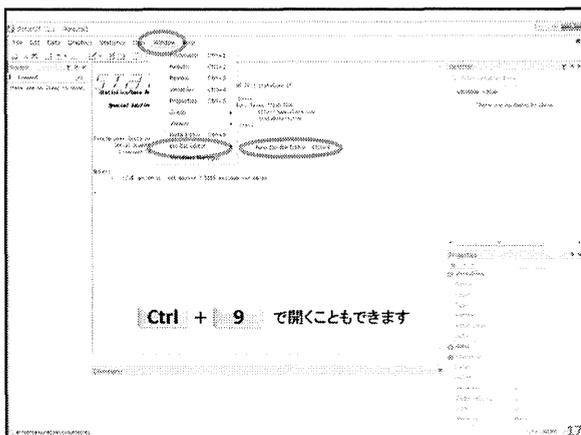
▶取り込むデータ

TRUMPSeminar2013.csv

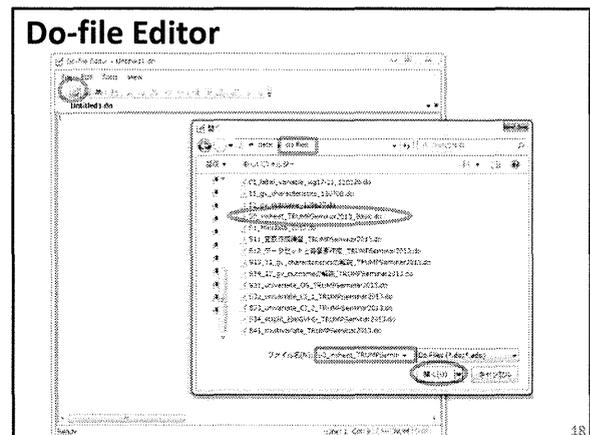
▶実行ファイル(スクリプト)

S0_insheets_TRUMPSeminar2013_basic.do

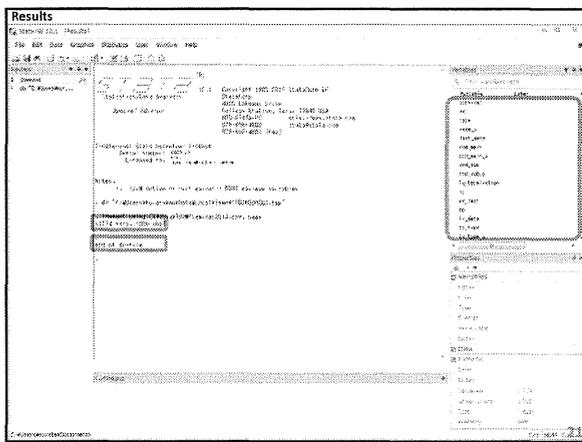
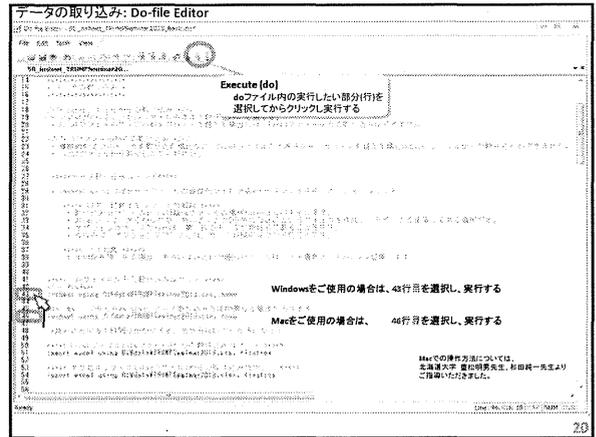
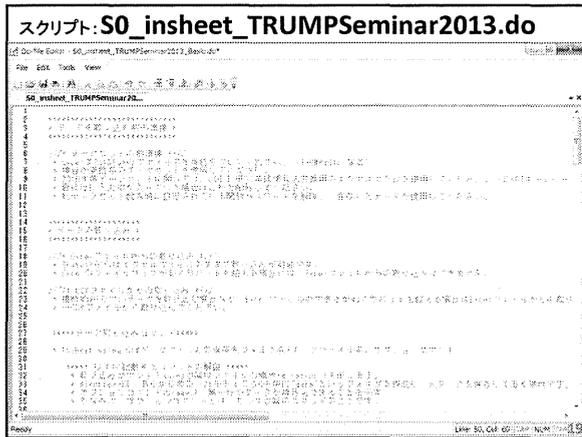
16



17



18

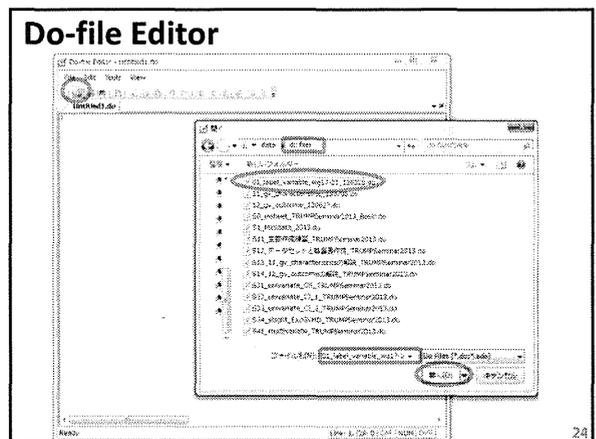
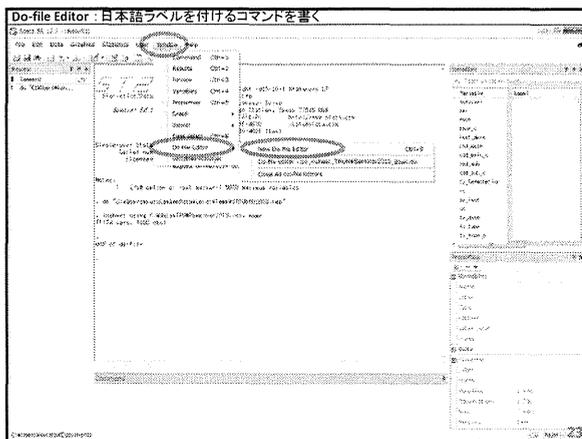


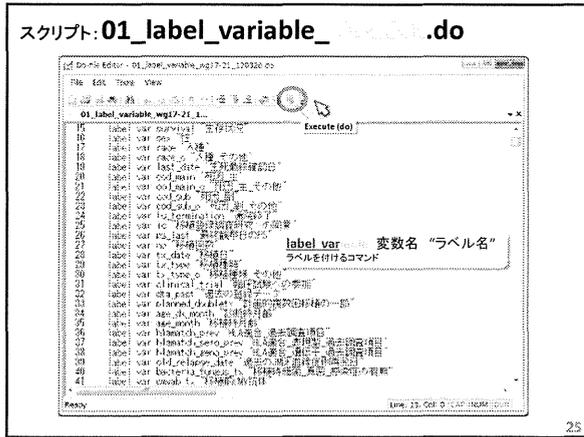
変数に日本語ラベルを付ける

変数の内容がわかるように変数ラベルを付ける

JSHCT TRUMPデータを用いた登録研究の解析を行う場合、各WGのデータセットに対応した日本語ラベルスクリプトをホームページよりダウンロード可能

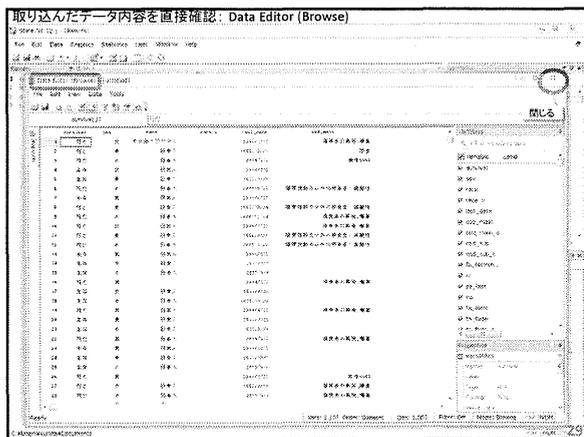
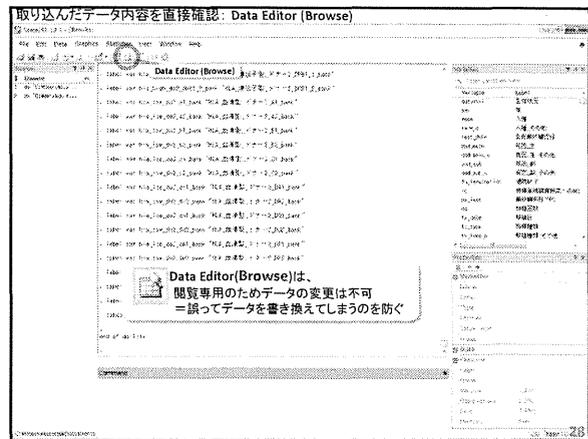
22





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
(指定したカテゴリー別の頻度、割合)

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てみましょう!

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

<u>d</u> escribe	→ de
<u>s</u> ummarize	→ sum
<u>t</u> abulate	→ tab
<u>l</u> ist	→ l
<u>s</u> ave	→ sa

標本数と変数の属性を確認①: Command window

Command windowに直接コマンドを入力して変数の属性を確認してみましょう!

- 使用するコマンド: **describe**
- 確認する変数 : survival no tx_type

describe 変数1 変数2 ...

de survival no tx_type Enter

標本数と変数の属性を確認①: Result window

変数名	変数タイプ	表示形式	ラベル名
survival	float	%9.0g	生存期間
no	float	%9.0g	性別
tx_type	float	%9.0g	治療法

標本数と変数の属性を確認②: Do-file Editor

```

1 * 生存期間と性別、治療法の属性を確認する
2 describe survival no tx_type
3
4 * 結果を確認する
5 list survival no tx_type
6
7 * 結果を保存する
8 save mydata.dta, replace
9
10 * 終了
11

```

describe 変数1 変数2 ...

連続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連続変数の要約値を確認してみましょう!

- 使用するコマンド: **summarize**
- 確認する変数 : age_month

sum 変数1 変数2 ...

sum age_month Enter

連続変数の要約値を確認①: Result window

観測値	平均値	標準偏差	最小値	最大値
age_month	50.000000	10.000000	30.000000	70.000000

