

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 鋤塚 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)
名古屋大学大森キャンパス

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。→ 講師も若くなりました。


注意点

データの収集・集積状況が一律でない

- 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、
読めるのが面倒。
- 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

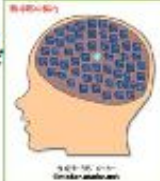


対策

- “大発見” “説明できない結果” に注意
(例) vooの有無で、予後に差はない。
「voo有」 ≠ 「voo無」
実は、「voo有」 ≠ 「voo無、他の合併症有」
 << 「合併症無(当然voo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...

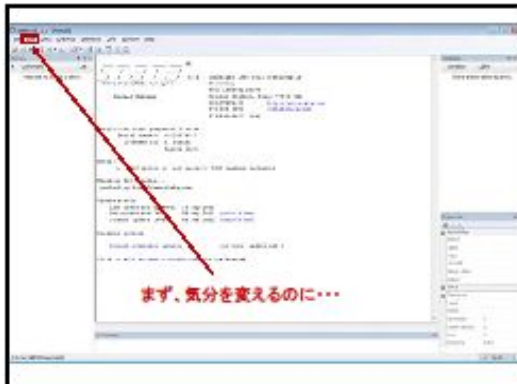
- 世の中には、統計ソフトというものがあるらしい。
- どうも、そいつを使って“統計解析”しないと、論文はアクセントされないようだ。
- 造血細胞移植学会WGで配られたデータも、やたら数が多いが、そういうのを使えば何とかなるらしい。
- WGで、タダでくれるというからもらった。
- Stata っていうらしいが、立ち上げたけどよく判らない。 → **そのまま放置**
- EZR というのもあるらしいが、もっとよく判らない。



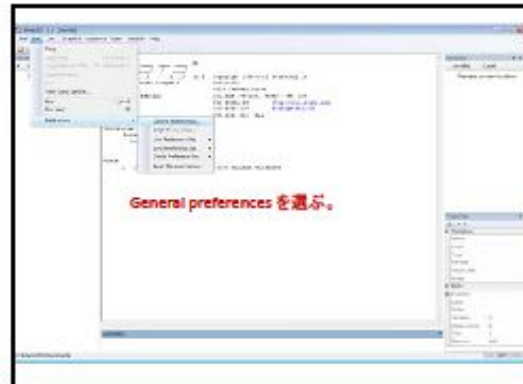
今日、Stata を立ち上げたら...



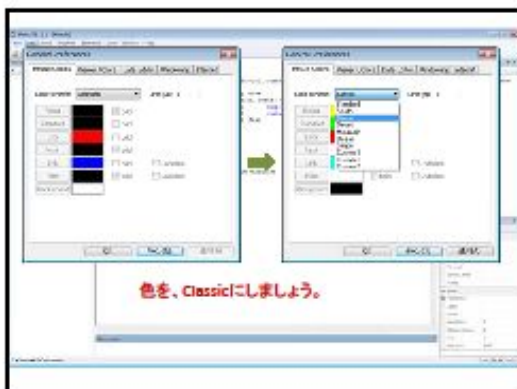
これが出た。



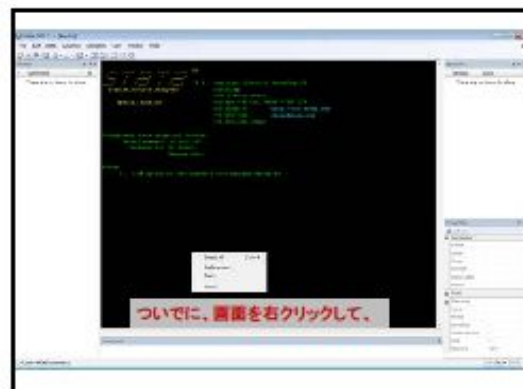
まず、気分を変えるのに...



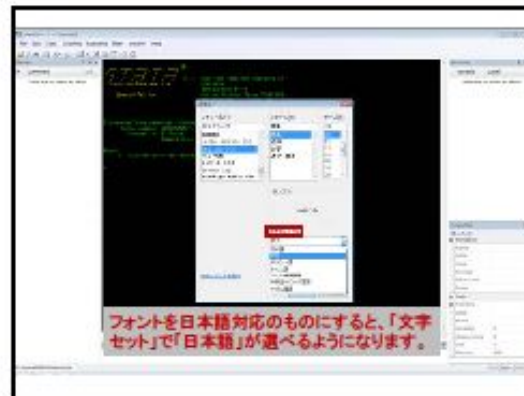
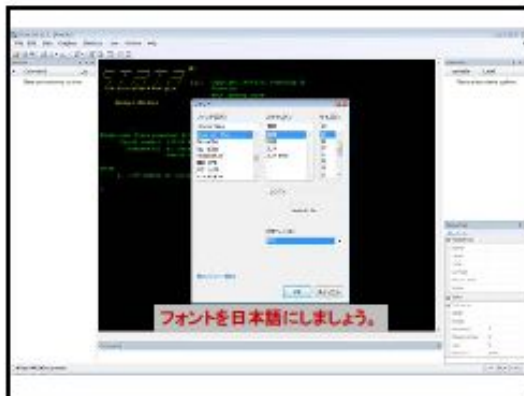
General preferences を選ぶ。



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かったりします。

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らないから、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、`state` 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
      58.3% (95% CI: 27.7-84.8%)
    
```


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

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
できますが、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

変数	平均	標準偏差	最小	最大
female	1.28750	.4418011	0.0	2.000000
male	0.712500	.4418011	0.0	2.000000
wt	160.709	73.9993	1.45	310.0000
height	1.687500	.0618011	1.57	1.800000
weight	160.7090	73.9993	1.45	310.0000
totalwt	160.7090	73.9993	1.45	310.0000

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認しておくか、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、分かりやすい。

Stata の基本動作

配布ファイルの保存場所

Windows

配布フォルダ

Macintosh

配布フォルダ

★ Macをご利用の場合は、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

Review window 実行したコマンドの結果が表示される

Result window 実行コマンドの結果が表示される

Command window 実行コマンドを入力する

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、データの値などが表示される

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

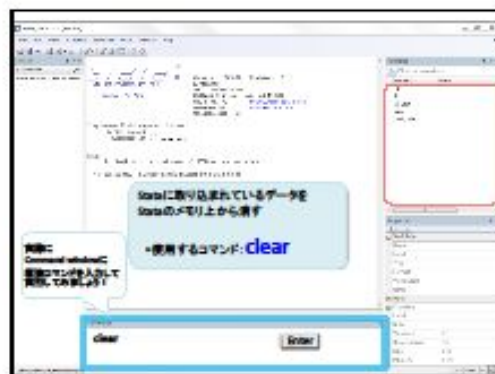
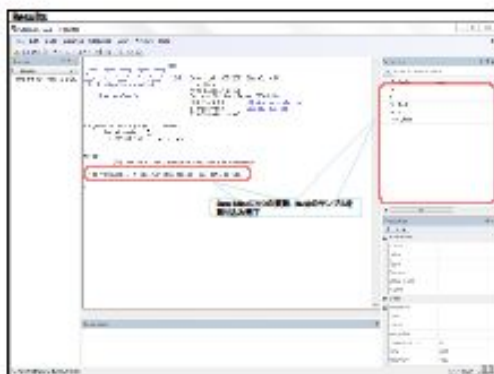
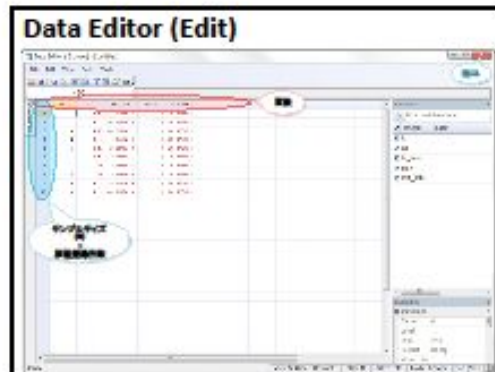
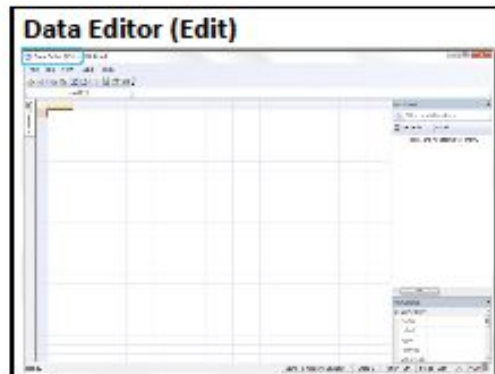
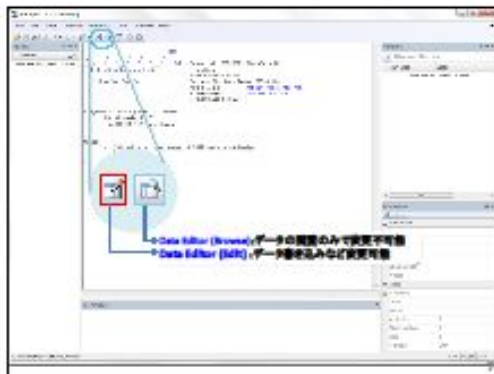
Stataにて読み込み可能なデータ形式

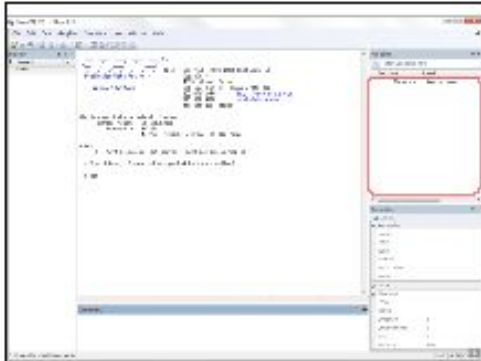
- ▶ **Excelファイル (.xls または .xlsx)**
Stataは標準ではExcelファイルからの取り込みが可能
ただし、40メガバイトを超えるExcelファイルからの取り込みは不可
- ▶ **CSVファイル (.csv)**
ファイルサイズが40メガバイトを超える場合は、Excelファイルからの取り込みは不可 → CSVファイルから取り込む
(標準的WGでのフルデータなど)
- ▶ **Stataファイル (.dta)**
ExcelまたはCSVファイルからStataに取り込んだデータをStata形式で保存したファイルを読み込む

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)





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

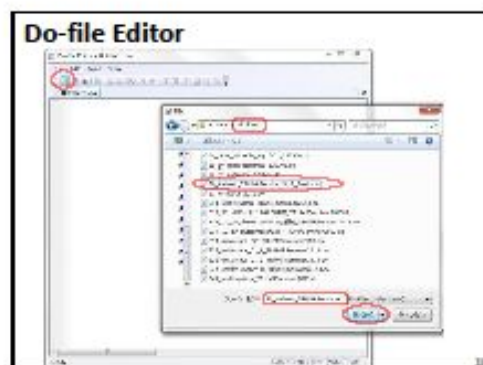
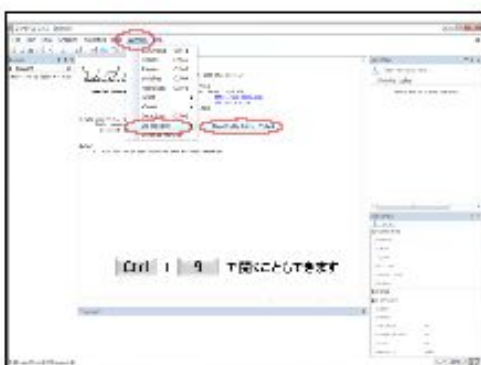
- 複数のコマンドを一度に実行することができる
 (実行したいコマンドのみを選択し実行することも可能)
- 作業メモなども書き留めることができる
- doファイルとして保存できるため、その都度コマンドを打ち込む必要がなく、同じ作業を繰り返し行うことができる

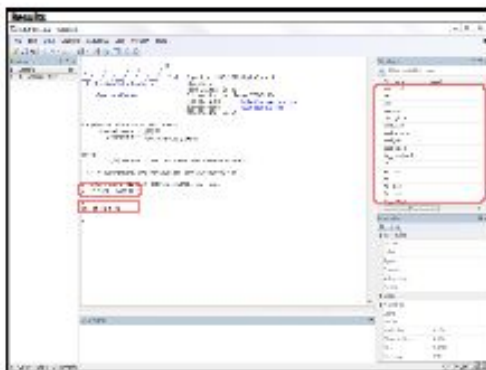
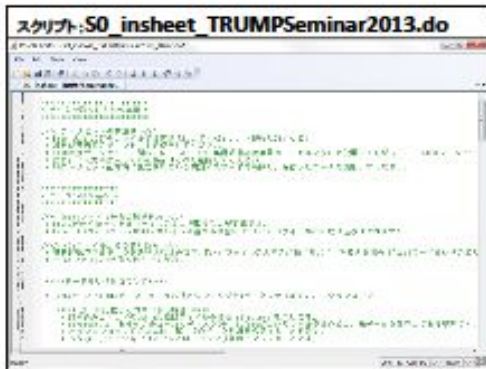
JSHQOT TRUMPデータを用いた登録研究の解析を行う場合、WG研究のための統計解析スクリプトをホームページよりダウンロード可能

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
[JSHQOT TRUMP Seminar 2013.csv](#)
- ▶ 実行ファイル(スクリプト)
[SO inahwet TRUMP Seminar 2013 basic.do](#)

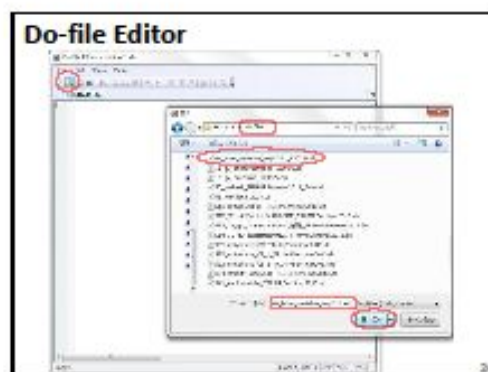
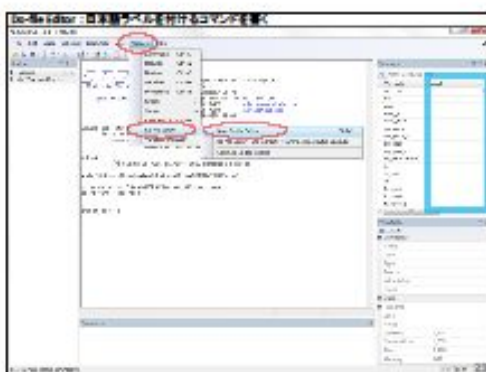


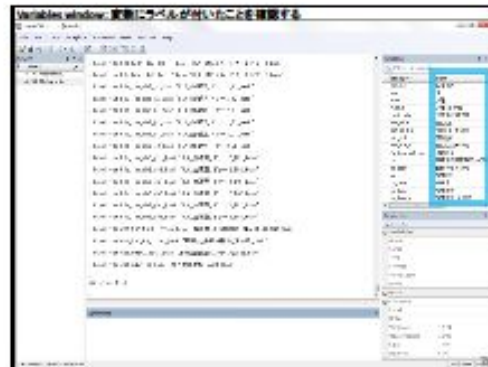
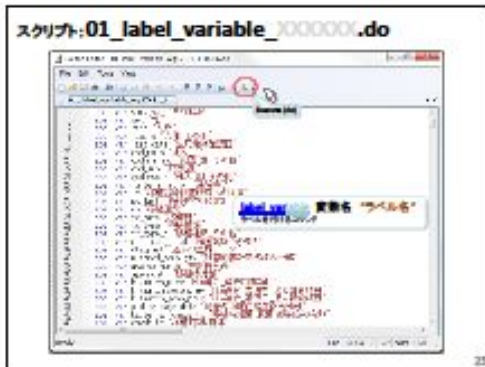


変数に日本語ラベルを付ける

変数の内容がわかるように変数ラベルを付ける

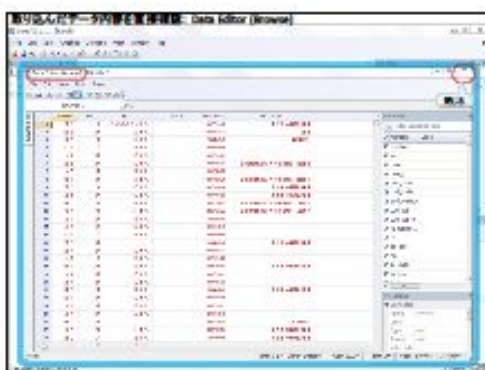
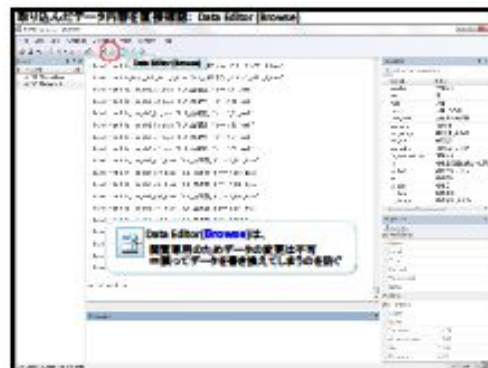
JSHCT TRUMPデータを用いた登録研究の解析を行う場合、各WGのデータセットに対応した日本語ラベルスクリプトをホームページよりダウンロード可能





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、実行の観音スクリーンを実行結果が一覧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	⇒	<u>d</u> e
<u>s</u> ummarize	⇒	<u>s</u> um
<u>t</u> abulate	⇒	<u>t</u> ab
<u>l</u> ist	⇒	<u>l</u>
<u>s</u> ave	⇒	<u>s</u> a

