

JSHCT TRUMP データを用いた登録研究のための Stata 簡易マニュアル

version1.2 20120828

Stata のインストール (Stata12 for Windows)		
1. インストールする前に	… 1	
2. インストール	… 1	
3. ライセンス情報の入力	… 2	
4. Stata のアップデート	… 2	
Stata の起動		
1. Stata 画面の説明	… 3	
データの取り込み		
1. Excel ファイルからの取り込み	… 4	
2. CSV ファイルからの取り込み	… 5	
データの保存		
1. 取り込んだデータの保存	… 6	
フォルダの整理		
1. フォルダの整理術	… 7	
日本語表示(フォント)の設定		
1. 日本語表示の設定方法	… 8	
プログラムの実行		
1. プログラムエディタ(do-file Editor)の起動	… 10	
2. プログラム (do file)の読み込み	… 10	
3. プログラム (do file)の実行	… 11	
4. ログをとる	… 11	
コマンド		
1. 基本操作に用いるコマンド	… 12	
2. 変数の作成・変更・条件文に用いるコマンド	… 12	
3. データの要約に用いるコマンド	… 14	
4. 基本の検定に用いるコマンド	… 14	
5. 生存解析に用いる基本コマンド	… 14	
6. 多変量解析に用いる基本コマンド	… 14	
グラフの作成		
1. グラフ作成に用いるコマンド	… 15	
2. グラフの保存	… 15	
プログラムファイル(do.file)の内容		
1. 配布 do.file	… 16	
2. 配布 do.file の内容とプログラムの書き方	… 16	
ヘルプの使い方		… 26

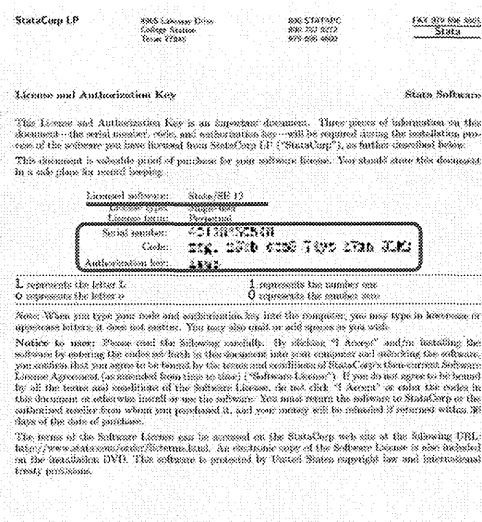
※本書では、Stata12 を Windows で使用する場合の操作方法を示しております。

Stata のインストール (Stata12 for Windows)

1. インストールする前に

インストールをするの前に以下をご確認ください。

- Stata のインストール CD-R をお手元にご用意ください。
- Serial number、Code、Authorization key が記載されている右の用紙をご用意ください。
(1.の CD-R のパッケージ内に付属しています。)
- インストールする Stata の種類を確認してください。
(右の用紙に記載されている **Licensed software** 種類を確認してください。)
- Stata をインストールする場所を決めてください。
以下のフォルダへのインストールをおすすめします。
32bit 版のインストール ⇒ **C:\Program Files\Stata12**
64bit 版のインストール ⇒ **C:\Program Files(x86)\Stata12**



- すでにコンピュータに旧バージョンの Stata をインストールされており、同じフォルダ内に Stata12 を追加でインストールする場合には、旧バージョンをアンインストールする必要があります。
※但し、アップグレードの場合には、アンインストールは不要です。
旧バージョンを残し、Stata12 を追加でインストールする場合は、異なるフォルダを作成する必要があります。複数のバージョンの Stata をインストールにより誤作動を生じる可能性があるため、おすすめはいたしません。

2. インストール

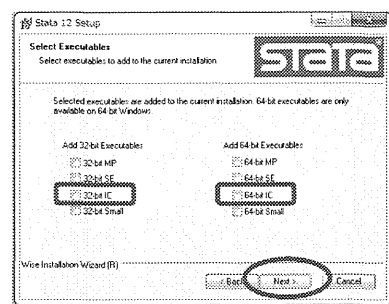
実際に Stata12 をインストールします。

- ① CD-ROM に Stata12 のソフトウェアを入れて、インストールを開始します。
 - 自動再生の設定がされている場合 ⇒ 自動的にインストールが始まります。
 - 自動的にインストールが始まらない場合 ⇒ CD-ROM を開き、setup.exe ファイルをクリックします。



- ② インストールする Stata の種類を選択します。
前述の「1.インストールする前に」をご参照うゑ **Licensed software** が記載されている用紙にてご確認ください。

Windows が 64bit で、64bit 版の Stata をインストールする場合には、
 64bit IC のチェックボックスを選択してください。



- ③ インストールする場所の設定を行います。
デフォルトでは、以下にフォルダに設定されています。デフォルト設定のフォルダを選択することをおすすめします。
 - 32bit 版のインストール ⇒ **C:\Program Files\Stata12**
 - 64bit 版のインストール ⇒ **C:\Program Files(x86)\Stata12**
 解析作業をすすめるにあたっては、「フォルダの整理」の章をご参照ください。

- ⑤ インストールが終了したら[Finish]をクリックしてください。

3. ライセンス情報の入力

Stata をインストール後、最初に起動する際に、ライセンスキーなどの入力画面が出てきます。Name、Organization、Serial Number*、Code*、Authorization*を入力します。
(*:前述の「1.インストールする前に」をご参照うえ、入力してください。)

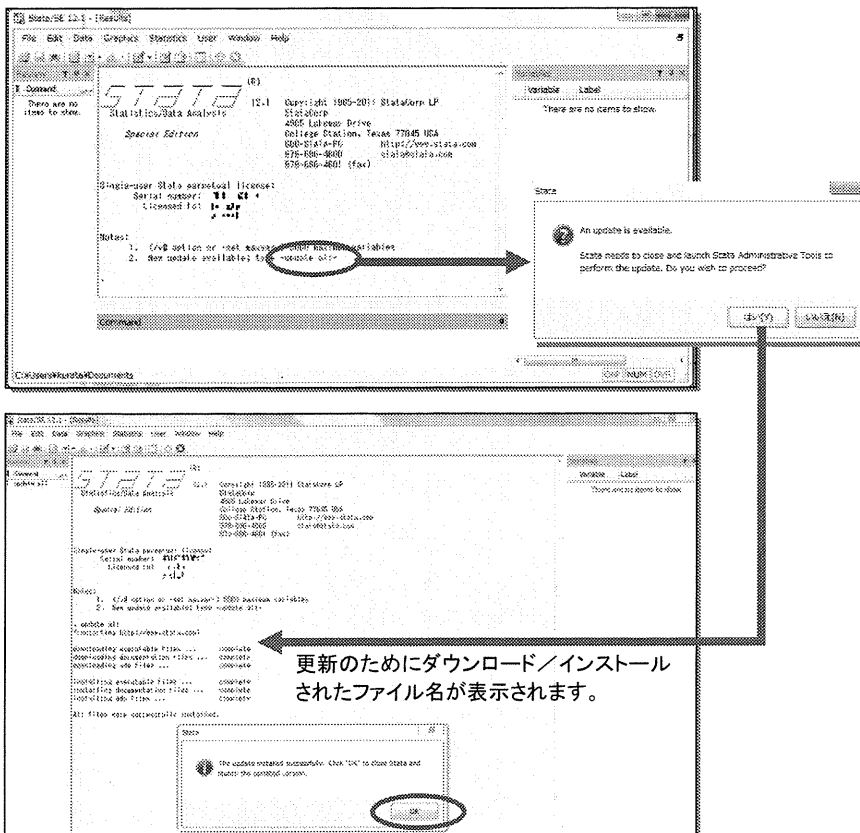
4. Stata のアップデート

しばしば Stata のアップデートがリリースされます。頻繁にアップデートを行うことをおすすめします。

4-1. アップデートの方法

Stata 起動時に Result window に表示される - update all - をクリックするか、または以下のコマンドを実行します。

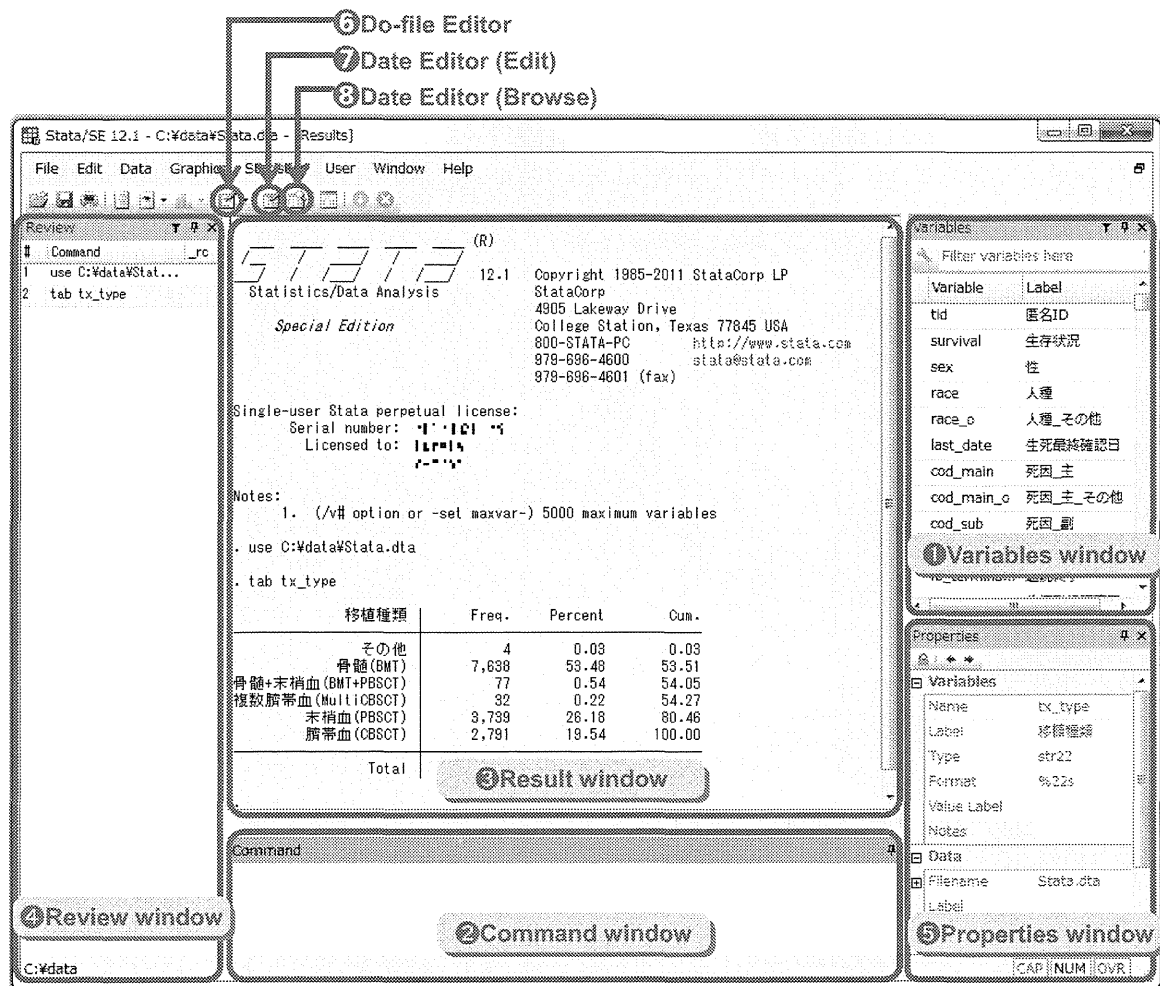
update query



Stata の起動

1. Stata 画面の説明

Stata を立ち上げると以下のような画面が表示されます。



説明

ウィンドウ	① Variables window	データ項目(変数名・ラベル名)の一覧が表示されます。 変数を入力する際は、変数名を手入力する必要はなく、このリスト上の変数をクリックすることにより入力できます。
	② Command window	コマンドを入力し、エンターキーにて実行をします。 (配布のプログラムを使用する際には⑥の Editor にて実行してください。)
	③ Result window	コマンドを実行した集計/解析結果やメッセージが順次に表示されます。
	④ Review window	過去に実行されたコマンドがすべて表示されていきます。表示されているコマンドをダブルクリックすると、そのコマンドが再度実行されます。
メニューボタン	⑤ Properties window	変数名、ラベル名、変数タイプ(str: 文字列、int: 整数型)、データの数、観測数 などを確認することができます。
	⑥ Do-file Editor	プログラムを書き、実行をします。(後述の「プログラムの実行」の章をご参照ください。)
	⑦ Date Editor(Edit)	取り込んだデータへの書き込みなど変更が可能なモードです。 (データの書き換えなどの誤操作を防ぐために、データ解析作業中は⑦ Browse モードでの操作をおすすめします。)
	⑧ Date Editor(Browse)	取り込んだデータの内容を見ることはできるが書き込みなど変更は不可能なモードです。

データの取り込み

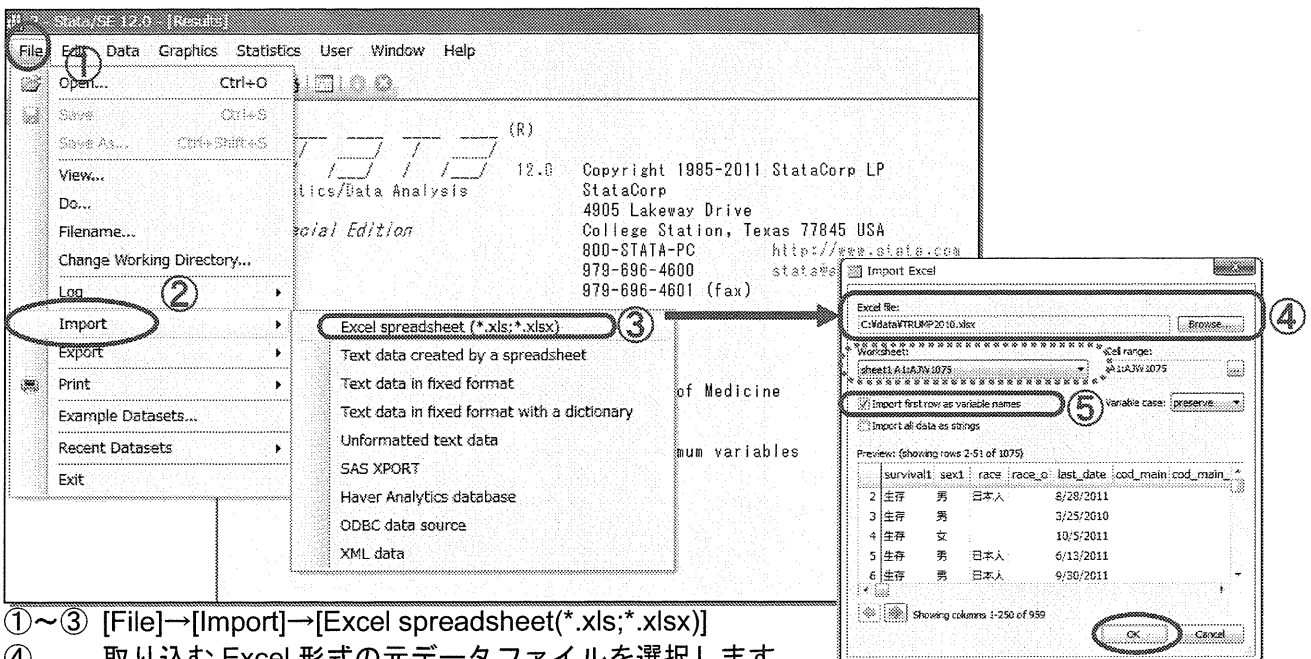
1.Excel ファイルからの取り込み

1-1.プルダウンメニューボタンを用いる取り込み方法

Stata12 では、Excel ファイルを直接インポートすることが可能です。

Excel 1997-2003(.xls)ファイル、Excel 2007-2010(.xlsx)ファイルの双方に対応しています。

※横断的 WG でフルデータを取り込む場合など、Excel ファイルの大きさが 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ができません。その場合は、CSV ファイルから取り込んでください。



①～③ [File]→[Import]→[Excel spreadsheet(*.xls;*.xlsx)]

④ 取り込む Excel 形式の元データファイルを選択します。

⑤ データの第一行目を変数名として取り込むために[import first row as variable names]を選択します。

※Excel ファイル内のワークシートが複数に分かれている場合、選択した 1 つのワークシートのデータしか取り込むことができません。

各 WG へはワークシートが分割されていない Excel データを配布しております。そちらのご使用をおすすめします。

ワークシートが分割されているデータを使用される場合には、①～⑤の手順にてワークシートごとにデータを取り込み、各 Stata データセット作成した後、本書 15 ページの merge コマンド実行にて各データセットを結合する必要があります。

1-2.コマンド実行での取り込み方法

以下の取り込みコマンドを実行します。

※Excel ファイル名に日本語が含まれる場合にはデータの取り込みができません。

予め、半角英数字のファイル名にしておく必要があります。(例：TRUMP2010.xlsx)

```
import excel using C:¥data¥TRUMP2010.xlsx, firstrow
```

②

③

①

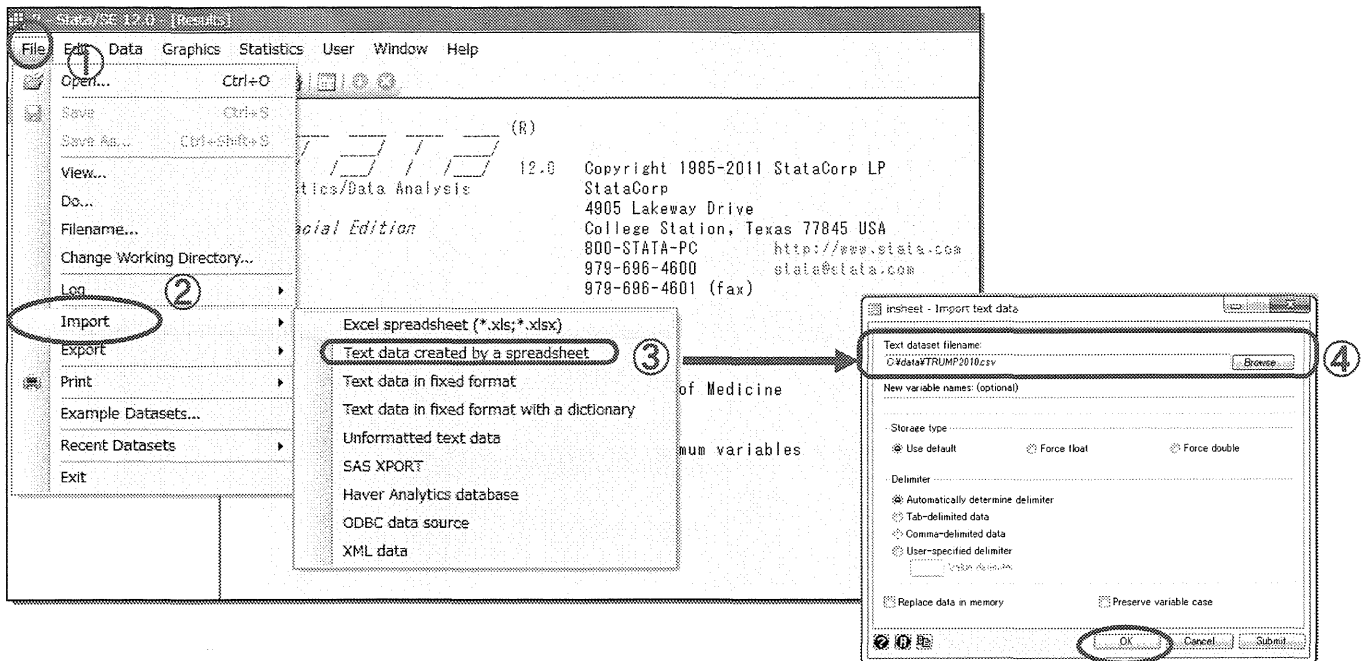
① using 以降はファイルの場所(directory)です。

② C ディスクの中に“data”というフォルダを作成し、予め Excel 形式の元データを入れておいてください。

③ オプションコマンド*の firstrow はデータファイル第一行目のデータが項目名であることを指示します。

2.CSV ファイルからの取り込み

2-1.プルダウンメニューボタンを用いる取り込み方法



- ①～③ [File]→[Import]→[Text data created by a spreadsheet]
 ④ 取り込む CSV 形式の元データファイルを選択します。

2-2.コマンド実行での取り込み方法

以下の取り込みコマンドを実行します。

※CSV ファイル名に日本語が含まれる場合にはデータの取り込みができません。
 予め、半角英数字のファイル名にしておく必要があります。(例：TRUMP2010.csv)

insheet using C:\data\TRUMP2010.csv,name

② ③
 ①

- ① using 以降はファイルの場所(directory)です。
 ② C ディスクの中に“data”というフォルダを作成し、予め CSV 形式の元データを入れておいてください。
 ③ オプションコマンド*の name は、データファイル第一行目のデータが項目名であることを指示します。

* オプションコマンドとは…
 カンマ以降のコマンドを示します。

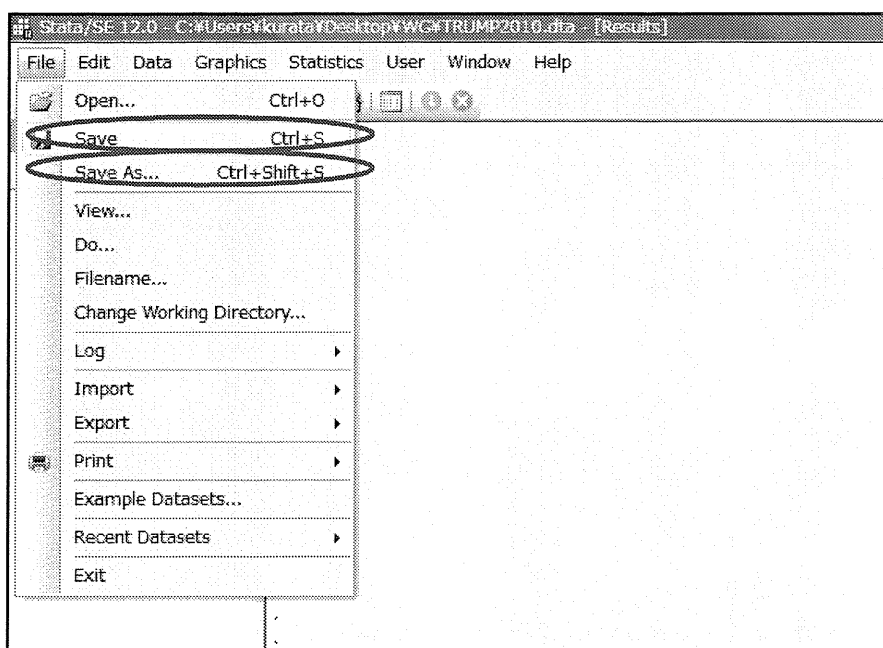
データの保存

1. 取り込んだデータの保存

取り込んだデータは Stata 形式(.dta)で保存しましょう。(半角英数字のファイル名で保存してください。) 次回ファイルを開ける際にダブルクリックにて Stata で開くことができるので便利です。

[File]→[Save]または[Save As]で保存してください。
また、以下のコマンドを実行し保存することもできます。

```
save C : %Data%TRUMP2010.csv.dta
```



* Stata 形式ファイルの拡張子について説明します。

拡張子	ファイル	内容
.dta	data file	取り込みデータ
.do	do file	一連のコマンドをプログラムとして書いたもの (次頁を参照)
.smcl	log file	解析結果のログを取ると結果を遡って見るのに便利です

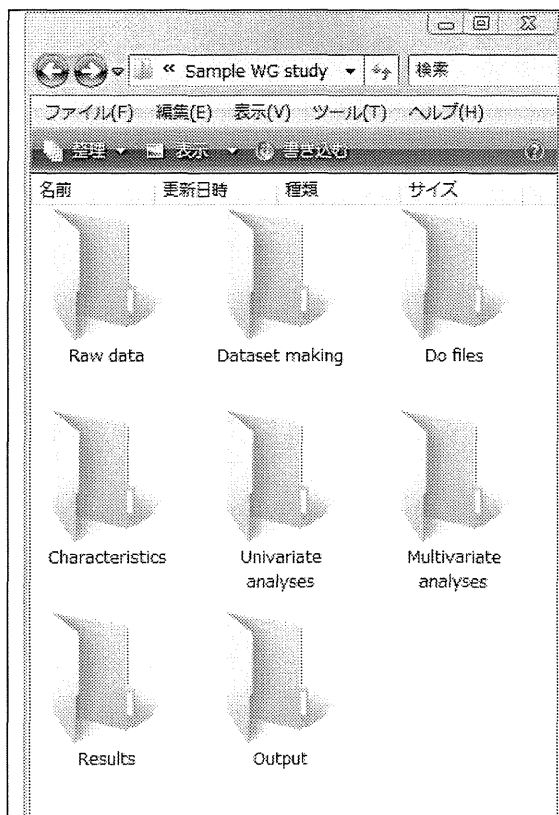
フォルダの整理

1. おすすめするフォルダの整理術

ご参考までに、熱田がおすすめするフォルダの整理術をご紹介します。

Stata での解析結果やログファイルは、作業を行うデータセットが保管されている同一のフォルダ内に保存されます。

下記のようにフォルダを作成して作業を進めていくと便利です。よろしければご活用ください。



▶Raw data

Excel ファイルや CSV ファイルなど、変更前の生データを保管します。

▶Dataset making[※]

解析データセットを作成するフォルダです。
この中で、解析に必要な基本変数の作成も行います。

▶Do files

該当研究の Do file はすべてこの中に保存します。

▶Characteristics[※]

患者背景表の集計・解析を行い、その解析結果、ログファイルを保存します。

▶Univariate analyses[※]

単変量解析を行い、作成した図や解析結果(検定結果含む)、ログファイルを保存します。
実際には、この中に Overall survival, Relapse-free survival, Cumulative incidence curves など、解析ごとの小フォルダを作成することが多いです。

▶Multivariate analyses[※]

多変量解析を行い、作成した図や解析結果(検定結果含む)、ログファイルを保存します。

▶Results

解析結果ログファイルを元に作成した表や、図をパワーポイントファイルで整理したものを保存します。

▶Output

学会発表のプレゼンテーションファイルや論文ファイルを保存します。

※データセットと同一のフォルダ内に解析結果、図、ログファイルあるいは別名で保存するデータセットが保存されるため、Data making の作業、Characteristics、Univariate analyses、Multivariate analyses の解析を行う場合には、使用するデータセット(.dta)を必ず各作業フォルダに保存のうえ作業を行ってください。

日本語表示(フォント)の設定

1.日本語表示の設定方法

Stata は標準設定の状態では、変数名またはコメントに日本語を用いる場合に表示 window で文字化けが生じます。そのため、日本語を表示する設定に変更してください。Window ごとに設定が必要となります。

1-1.Result window の日本語表示設定

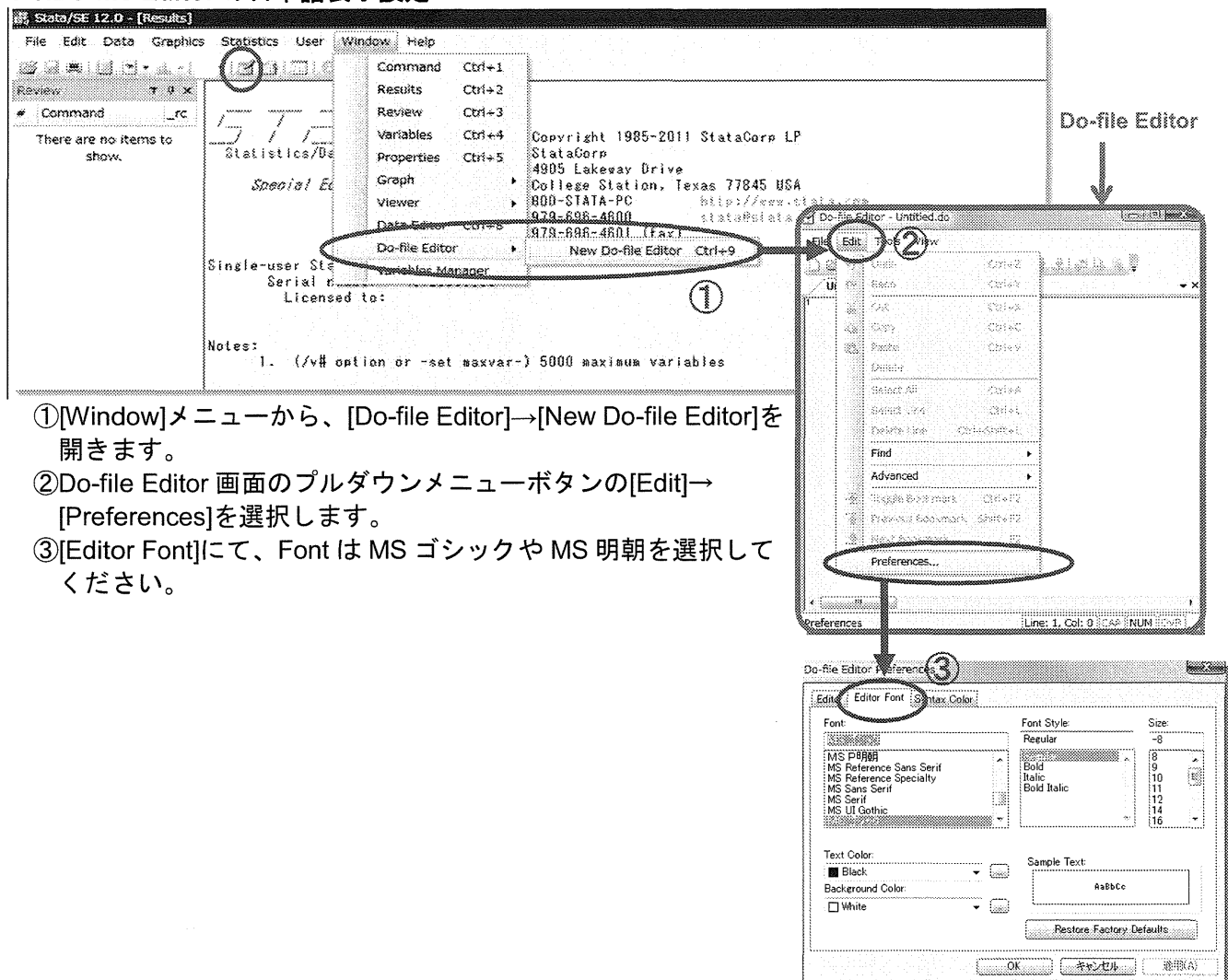
①-1, ②-1 Result window 内のどこかで右クリックをしてください。
 ①-2 Preference にて表示フォントの設定を行います。
 Color scheme の[Custom1]を選択し、[Result]の bold チェックを外してください。 bold → bold
 bold に設定されている場合には、Result window 内の日本語表示に文字化けを生じます。
 ②-2 [フォント名]は“MS ゴシック”や“MS 明朝”を選択し、[文字セット]は“日本語”を選択してください。フォント名の設定で@MS ゴシックや@MS 明朝を選択すると文字が左に 90°回転するので、@マークが付いていないフォントを選択してください。

1-2.Command window の日本語表示設定

Command window のほか Review window、Variables window においても同様の設定を行ってください。

① Command window 内のどこかで右クリックをしてください。
 ② [フォント名]は MS ゴシックや MS 明朝を選択し、[文字セット]は日本語を選択してください。


1-3. Do-file Editor の日本語表示設定



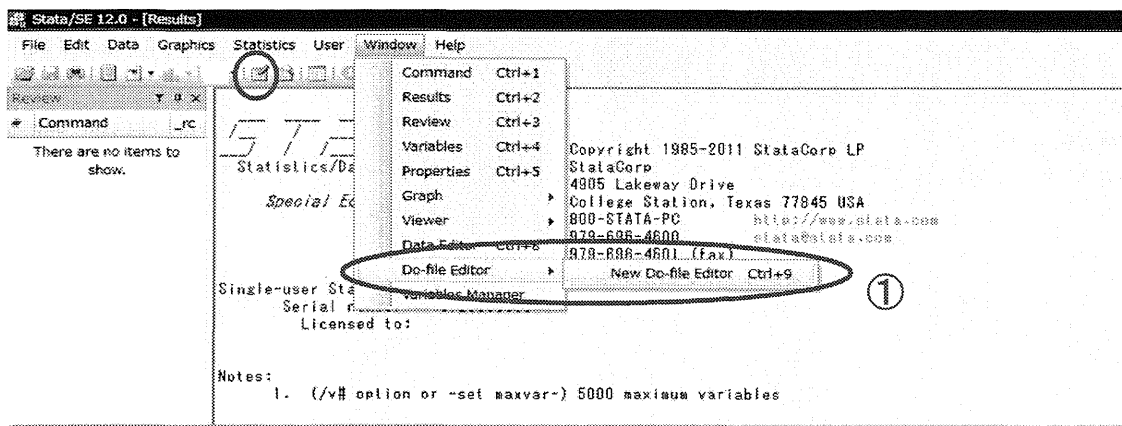
- ①[Window]メニューから、[Do-file Editor]→[New Do-file Editor]を開きます。
- ②Do-file Editor 画面のプルダウンメニューボタンの[Edit]→[Preferences]を選択します。
- ③[Editor Font]にて、Font は MS ゴシックや MS 明朝を選択してください。

プログラムの実行

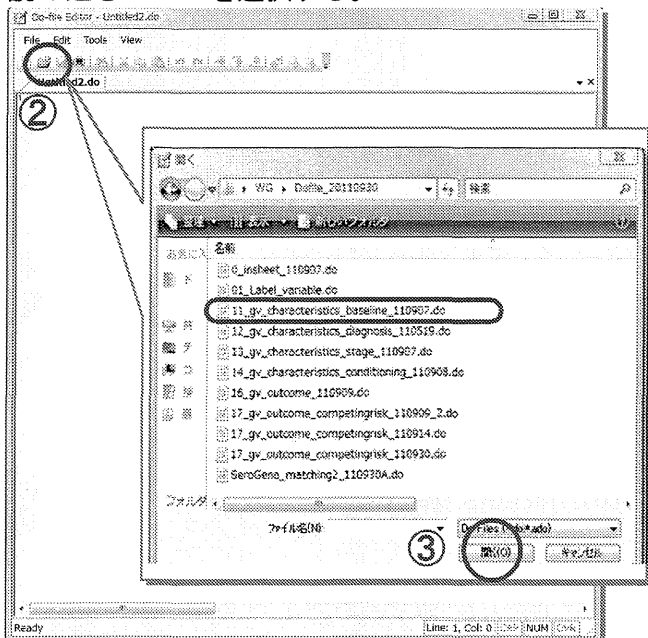
1. プログラムエディタ(do-file Editor)の起動

プログラムを書き実行するにあたり、do-file editor を起動します。
 [Window]メニューから、[Do-file Editor]→[New Do-file Editor]を選択します。
のアイコンをクリック、または、[Ctrl]+[9]の key で起動することもできます。

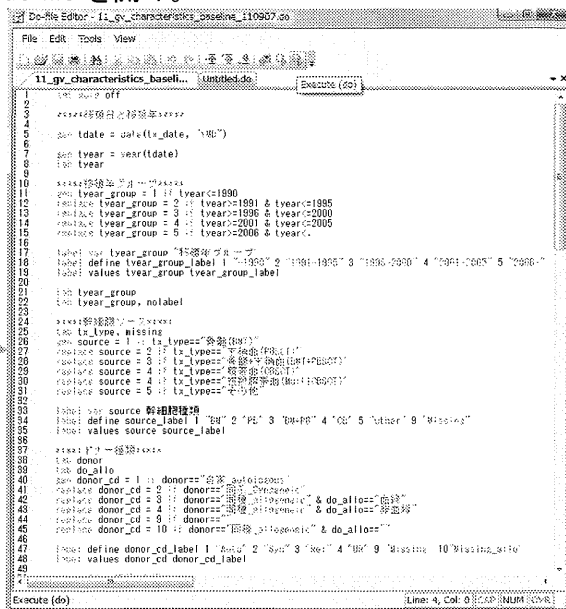
2. プログラム(do file)の読み込み



読み込む do file を選択する。




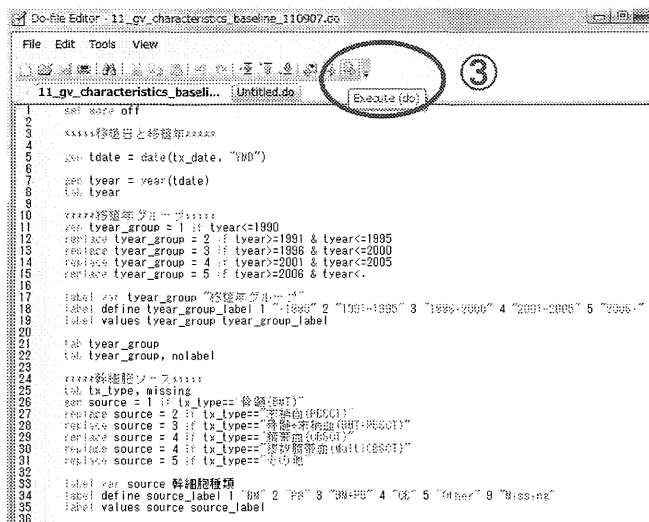
do file を開く。




3. プログラム(do file)の実行

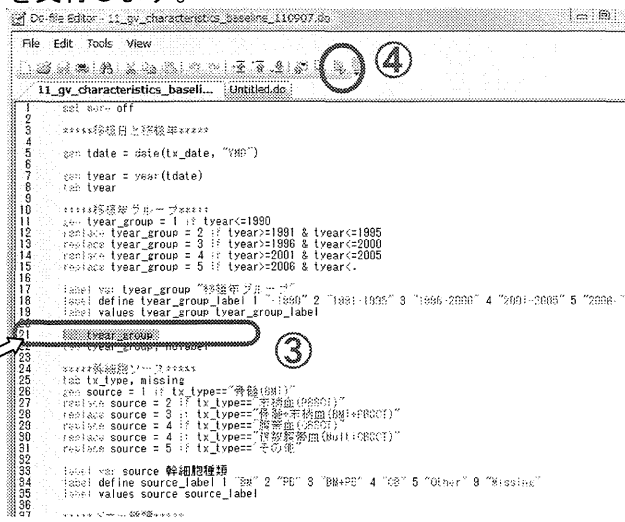
■プログラムの全内容を実行する。

右端の  のアイコンをクリックし、プログラムを実行します。



■プログラムを部分的に実行する。

マウスを使い、実行したい部分のプログラムを選択し、右端の  アイコンをクリックし、プログラムを実行します。



4. ログをとる

結果はStata Resultsウィンドウに表示されますが、結果が長くなると遡って見ることができなくなります。

ログを取ると結果を遡って見るのに便利です。例の最後のtranslateコマンドではテキストファイルでも保存されるため、メモ帳などで見ることができ便利です。

**log using ファイル名
log close**

例 :

```
log using ログファイル名
xxx
log off
yyy
log on
zzz
log close
translate ログファイル名.smcl ログファイル名.txt
```

ログファイルには「xxx」と「zzz」の部分のログは記録されますが、「yyy」の部分は記録されません。

ログファイル名はdo fileと同じファイル名にしておくと便利です。

プログラム中にコメントを書いておくと、あとから見直しやすくなります。

* コメントの書き方…

Stataは「*」で囲まれた部分をプログラムとして認識しないため、コメントの行頭に「*」をつけます。コメントが複数行に亘り長文になる場合には、「/*」「*/」でコメントを挟みます。

コマンド

1. 基本操作に用いるコマンド

コマンド	説明	書式
use ~ ,clear	ファイルを読み込んで解析をおこなう	use ファイル名,clear
log	全ての解析結果のログをとる	log using ファイル名
help	ヘルプを参照する	help 探すコマンド

2. 変数の作成・変更・条件文に用いるコマンド

コマンド	説明	書式
generate (gen)	新しい変数を作成し、その全てに値を代入する	gen 新変数名 = 値
replace	既存の変数の値を変更する	replace 既存の変数 = 新しい値
rename (ren)	既存の変数の名前を変更する	rename 既存の変数名 新変数名
if	条件文	コマンド if 条件
destring	文字列を数値に変換して、既存の変数を置き換える	destring 既存の変数, replace
	文字列を数値に変換した新しい変数を作る	destring 既存の変数, gen(新変数)
tostring	数値を文字列に変換して、既存の変数を置き換える	tostring 既存の変数, replace
	数値を文字列に変換した新しい変数を作る	tostring 既存の変数, gen(新変数)
label	変数にラベルをつける	label var 変数 "ラベルの内容"
	値と対応するラベルセットを設定する	label define セット名 値 1 "ラベル 1" 値 2 "ラベル 2" 値 3 "ラベル 3"
	ラベルセットを変数に割り付ける	label value 変数 セット名

()は省略形

* Excel ファイル[変数一覧]の label について説明します。

	A	B	C	D	E	F	G	H
	R-Variables	R-Values	コメント	Variables	Value	Value label	コメント	
1								
2								
3	移植日	移植日	文字列のまま	tdate	移植日付			
4								
5	移植年	SCT.Year	数値	tyear	移植年			
6								
7	診断から移植までの日数	Dx.ToSCT	数値	dx_to_sct				
8								
9	診断から移植までの日数グループ	Dx.ToSCTGroup	0 <=90 1 <=180 2 >180	dx_to_sct_group	0 1 2	<=90 90 to 180 >180		
10								
11								
12								
13	移植年グループ	SCT.YearGroup	-1990 文字列 1991-1995 1996-2000 2001-2005 2006-	tyear_group	1 2 3 4 5	-1990 1991-1995 1996-2000 2001-2005 2006-		
14								
15								
16								
17								
18								
19	幹細胞種類	Source	BM PB BM+PB CB Others	source	1 2 3 4 5	BM PB BM+PB CB Other		
20								
21								
22								
23								
24							Missing	

Value label とは…

項目の値に“ラベル”がついている、ということです。

Variable は数値でコード化してありますが、コードの意味がわからなくなないようにラベルとなっています。

実際の入力値に関しては以下のように確認ができます。

<項目 “sct_type” の場合>

tab sct_type,missing

tab sct_type,nolabel missing ←これでラベルなしのコードが出ます。

“sct_type_label” というラベルを用いていますので、

label list sct_type_label でコードとラベルの対応もわかります

if 以下構文で用いる演算子です。

以下のコマンドを用いると、数式・論理式・等号不等号などの各種演算子を表示することができます。

help operator

コマンド	説明
+	等しい
-	差
*	積
/	商
^	べき乗
&	かつ
	または
!	否定
~	否定
>	より大きい
<	より小さい
>=	以上
<=	以下
==	等しい
!=	等しくない
~=	等しくない

以下のコマンドを実行すると、数式の一覧を表示することができます。

使用頻度は低いと思いますが参考まで。

help mathfun

コマンド	説明
abs	絶対値
acos	アークコサイン(数値の逆余弦)
acosh	アークハイパボリックコサイン(逆双曲線余弦)
asin	アークサイン(数値の逆正弦)
asinh	アークハイパボリックサイン(逆双曲線正弦)
atan	アークタンジェント(数値の逆正接)
atan2	アークタンジェント 2(xy 座標の逆正接)
atanh	アークハイパボリックタンジェント(逆双曲線正接)
ceil	小数点以下を切り上げ整数
cloglog	連結関数
comb	コム関数 $n! / \{k!(n - k)!\}$
cos	コサイン(余弦関数)
cosh	ハイパボリックコサイン(双曲線余弦関数)
digamma	ディガンマ関数(ガンマ関数の対数の一階微分関数)
exp	指数関数 e^x
floor	小数点以下を切り捨て $n < x < n + 1$ を満たす整数
int	小数点以下を切り捨てる

コマンド	説明
ln	自然対数
lnfactorial	自然対数 n の階乗
lngamma	ガンマ関数の自然対数
log	自然対数
log10	常用対数
logit	ロジット関数 (log of the odds ratio of x)
max	最大値
min	最小値
mod	MOD 関数(剰余を求める)
reldif	"relative" difference $ x-y /(y +1)$
round	四捨五入
sign	サイン関数(数値項目の符号を返す)
sin	サイン(正弦関数)
sinh	ハイパボリックサイン(双曲線正弦関数)
sqrt	平方根
tan	タンジェント(正接関数)
tanh	ハイパボリックタンジェント(双曲線正接関数)

invcloglog	連結関数の逆関数
invlogit	ロジット関数の逆関数

trigamma	トリガンマ関数(ガンマ関数の対数の二階微分関数)
trunc	小数点以下を切り捨てて整数または指定された桁数に数値を変換

3. データの要約に用いるコマンド

コマンド	説明	書式
tab	カテゴリ変数の要約値を表示する	tab 変数
	オプションで missing をつけると欠損データ数を算出することができる	tab 変数,missing
summarize(sum)	連続変数の要約値を計算する	sum 変数
	オプションで detail をつけると総数・平均値・標準偏差・中央値・パーセンタイル値・最小値・最大値・分散・歪度・尖度を算出することができる	sum 変数,detail
ci	平均値・標準偏差・信頼区間を計算する	ci 変数
sort	サンプルを並び替える	sort 変数
list	データのリスト作成	list 変数 変数
count	変数が欠損値でない、サンプル数を表示する	count

4. 基本の検定に用いるコマンド

コマンド	説明	書式
ttest	t 検定	ttest 変数,by (グループ変数)
	1 群の t 検定	ttest 変数==数値
	paired t 検定	ttest 変数 1==変数 2
	two-sample t 検定	ttest 変数 1==変数 2,unpair
ranksum	Mann-Whitney 検定	ranksum 変数,by(グループ変数)
tab~ ~,chi2	chi-squared 検定	tab 変数 1 変数 2,chi2
tab~ ~,exact	Fisher の正確検定	tab 変数 1 変数 2,exact

5. 生存解析に用いる基本コマンド

コマンド	説明	書式
stset	エンドポイントを表す変数=イベントの値	stset 時間変数,failure(エンドポイントを表す変数==イベント値)
sts graph	生存曲線を描く	sts graph, title(グラフのタイトル) ytitle(y 軸のタイトル) xtitle(x 軸のタイトル)
sts list	生存解析結果を表示する	sts list ,by(グループ変数) at(時間時点リスト)
sts test	Log-rank 検定	sts test (グループ変数)
	Wilcoxon 検定	sts test (グループ変数),wilcoxon

6. 多変量解析に用いる基本コマンド

コマンド	説明	書式
stcox	コックス比例ハザードモデル	stcox 説明変数リスト,オプション
stcrreg	競合リスク回帰モデル	stcrreg 変数,compete(crvar==numlist)オプション

グラフの作成

1. グラフの作成に用いるコマンド

コマンド	説明	書式
histogram	正規性の検証(正規曲線)	histogram 変数,オプション
scatter	散布図	scatter 変数 1 変数 2
graph bar	棒グラフ	graph bar 変数 1 変数 2,オプション
graph box	箱ひげ図	graph box 変数,オプション
twoway	折れ線グラフ	twoway (line 変数 1,オプション)
sts graph	Kaplan-Meier 曲線	sts graph, by(グループ変数) 他のオプション

2. グラフの保存

グラフの保存には以下のコマンドを実行します。

```
graph save グラフ名,オプション
```

例：生存曲線を描く場合

```
sts graph if (条件) & (条件) & (条件), by(群分けグループ変数) title(グラフのタイトル) ytitle(y 軸のタイトル) xtitle(x 軸のタイトル)
```

グラフを保存する

```
translate @Graph (グラフ名).emf, trans(Graph2emf)
```

graph名にはスペース、<>等は使用できません。

emfの形式にてデータ(.dtaファイル)保存先のフォルダに保存されます。

様々な設定でのサブグループで抽出する際には、by()グループ変数名を入力します。

打ち切りイベントが発生した時点に縦線をつける場合には、オプション箇所に censored(multiple)を加えます。

tied-event の例数が多い場合には、censored(single)を用いてください。

詳細については、「プログラムファイル(do.file)の内容」の章をご参照ください。

プログラムファイル(do.file)の内容

1. 配布 do.file

ファイル名	プログラム内容
0_insheet_XXXXXX.do	データの取り込み方法
01_label_variable_XXXXXX.do	変数ラベルの作成(TRUMP の日本語変数名を読み込みます)
02_dataset_making_XXXXXX.do	解析データセットの作成方法
03_merge_XXXXXX.do	共通変数をもつ2つの異なるデータセットを結合する方法
11_gv_characteristics_XXXXXX.do	基本情報(移植情報, 患者/ドナー情報)、疾患分類、移植時病期、疾患リスク分類、移植前処置、GVHD 予防、HLA マッチングの変数作成
12_gv_outcome_XXXXXX.do	生存解析のイベント、競合リスクのイベント(再発, 好中球回復, 血小板回復, 急性 GVHD, 慢性 GVHD)の変数の作成
21_characteristic_XXXXXX.do	データ要約、検定に用いるコマンドの説明
31_univariate_OS_XXXXXXX.do	単変量解析の解説(Kaplan-meier 生存解析、生存曲線の描き方、検定)

* ファイル名の末尾の6桁の数字は、ファイルの作成または更新の日付を示します。ファイル内容の改訂毎に日付は更新されます。
ファイルが最新であるかをご確認のうえ、ご使用ください。

次に、do.file 内容と書き方について説明します。

2. 配布 do.file の内容とプログラムの書き方

0_insheet_XXXXXX.do

excel ファイルまたは、csv ファイルからのデータの取り込み方法の説明です。

```

1  **データの取り込み
2  **$atal2からはエクセルファイルのまま取り込みが可能
3  **項目名のみ英語に差し替えておく
4  **以下はcsvファイルからの取り込み
5
6  **読み込みコマンド、using以降はファイルの場所(directory)
7  **directory, 以下は、Cの中'にdata'というフォルダを作成し、元データを入れてあります
8  **オプションコマンドのnameは、第一行のデータが項目名であることを示す。
9  **オプションコマンドとは、カンマ以降のコマンド
10
11  insheet using C:\data\TRUMP2010_2.csv, name
12

```

csv ファイルからのデータ取り込みの際のコマンドです。

▶01_label_variable_XXXXXX.do

取り込んだデータの変数名に日本語の項目ラベルを付けるプログラムです。

```

Do-file Editor - 01_label_variable.do
File Edit Tools View
01_label_variable.do
1  **日本語での項目名を、項目ラベルとしてつけます。
2
3  label var tid "匿名ID"
4  label var survival "生存状況"
5  label var sex "性"
6  label var race "人種"
7  label var race_o "人種_その他"
8  label var last_date "生死最終確認日"
9  label var cod_main "死因_主"
10 label var cod_main_o "死因_主_その他"
11 label var cod_sub "死因_副"
12 label var cod_sub_o "死因_副_その他"
13 label var fu_termination "通院終了"
14 label var ic "移植登録調査研究への同意"
15 label var ps_last "最終観察日のPS"
16 label var no "移植回数"
17 label var tx_date "移植日"

```

データを取り込んだ段階では変数に名前はついていても、ラベルはついていません。作業の際に便利であるため、TRUMPの質問項目と同じに日本語のラベルをつけます。

label variables 変数“ラベルの内容”

▶02_dataset_making_XXXXXX.do

解析データセットを作成する方法です。

```

Do-file Editor - 02_dataset_making_120508.do
File Edit Tools View
02_dataset_making_120508.do Untitled.do
1  **解析対象のデータセットを作成する。
2
3
4  *条件に合う変数以外の変数を全て除いたデータセットを作成する場合
5  //(**コマンド書式**)
6  *keep if 条件式
7
8  *移植年2001年以降の初回自家移植のみのデータを作成する場合
9  keep if tyear_group>=4 & no==1 & donor=="自家_autologous"
10
11
12 *加工したデータの保存
13 //(**コマンド書式**)
14 *save ファイル名
15 save Auto
16
17
18 *条件に合う変数を全て除いたデータを作成する場合
19 //(**コマンド書式**)
20 *drop if 条件式
21
22 *生存状態または最終確認日が不明の不備データを除外する場合
23 drop if survival==" " | last_date==" "
24 *result windowにて除外数の確認ができます
25
26 *除外後の例数確認
27 count
28
29 *加工したデータの保存
30 //(**コマンド書式**)
31 *save ファイル名
32 save TRUMP2012_2
33

```

上記のコマンドにて除かれた変数は復元することはできませんが、解析に不要な変数がデータに含まれる場合、処理速度が遅くなるので、keepやdropなど使い、データをsaveしておくとう便利です。

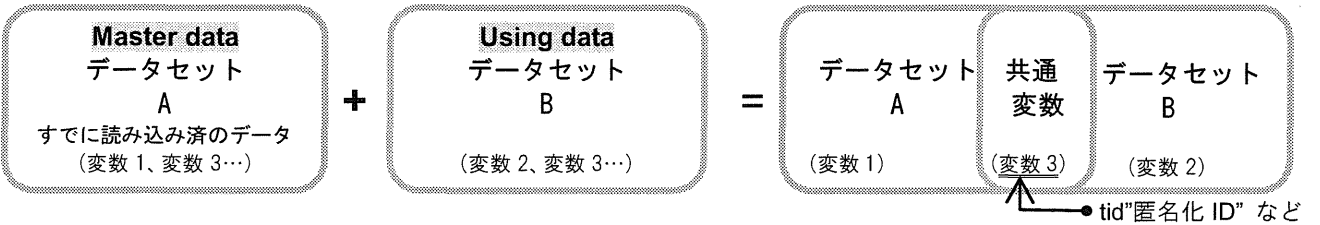
▶03_merge_XXXXXX.do

一部共通する変数をもつ2つの異なるデータセットにて、片方のサンプルにない変数を追加する方法です。

```

1  **2つの異なるデータを結合させる。
2
3  *一部共通する変数をもつ2つの異なるデータにて、片方のサンプルにはない変数を追加する。
4
5  //**コマンド書式**//
6
7  *master dataとusing dataの観測データが1対1で対応する場合 ①
8  *merge 1:1 変数リスト using ファイル名
9
10
11 *上記の変数リストは双方のデータセットに存在する変数で、キーとなり結合データが作成されます。
12 *ただし、mergeする前に双方のデータセットにて「変数リスト」でsortされていないと、mergeコマンドによる結合が行われないので注意が必要です。
13
14
15 *TRUMP IDを用いてTRUMP2010データセット(もともとのデータセット)にAddのデータを結合する場合
16 *データセット名は半角英数字にしてください。
17 use C:\ydata\TRUMP2010.dta
18 sort tid
19 merge 1:1 tid using C:\ydata\Add.data
20 *両方のデータセットに存在する匿名化ID(変数名 tid)がキーとなり結合されます。
21
22 **注意**
23 *追加するデータセットの変数名は、もともとのデータセットの変数名と別名にしてください。
24

```



上記の場合、データセットAとデータセットBの共通する変数(変数3)がキーとなり結合データが作成されます。

どちらか一方にしか値がない場合、値がないものは欠損値となります。

merge コマンドを実行する際の注意点...

- ※merge する前にデータセットAおよびデータセットBの両方にて変数リスト(共通する変数)で sort されていなければなりません。
 - デスク上にある using data のデータセットBが sort されていない場合には、まずデータセットBを読み込んで sort し、保存しておく必要があります。
- ※master data と using data の両方に同じ変数名があると、using data のほうは取り込まれません。
 - rename コマンドにて変数名を変えておく必要があります。

①master data と using data の観測データが '1' 対 '1' で対応する場合

```
merge 1:1 変数リスト using ファイル名
```

<上記の例の場合>

```
merge 1:1 変数3 using データセット3
```

merge を実行すると、新しい変数「_merge」が作成され、1, 2, 3の値が格納されます。

- 1,2,3の値は、どちらのデータセットに属していたかを意味します。
 - 1: もともとのデータセット(master data)にあるが、読み込んだデータセット(using data)にはない
 - 2: もともとのデータセット(master data)にはないが、読み込んだデータセット(using data)にのみある
 - 3: もともとのデータセット(master data)と、読み込んだデータセット(using data)の両方にある

merge 作業を続ける場合の注意点...

- ※作成された_merge 変数はデータベース上に残るため、削除しない限り新たな_merge を定義できなくなります。
 - 次々と追加で merge を行う場合には drop コマンドで削除を行ってください。

▶11_gv_characteristics_baseline_XXXXXX.do

基本情報(移植情報、患者/ドナー情報、疾患分類、移植時病期、移植前処置、GVHD 予防、HLA マッチング)の変数を作成するプログラムです。

```

1  **** Baseline ****
2  set more off ①
3
4  ****移植日と移植年****
5
6  gen tdate = date(tx_date, "Yd") ②
7
8  /*Excelファイルからデータを取り込んだ場合、Excel日付データはdate日付データに自動変換されているため、上記の日付コマンド実行は不要です。
9  /*以下のコマンドを実行してください。
10 /**/間のコマンドを実行する場合は、【/*】【*/】を取ってください
11  gen tdate = tx_date
12
13  */
14
15  gen tyear = year(tdate) ③
16  gen tyear
17
18  ****移植年グループ****
19  gen tyear_group = 1 if tyear<=1990 ④
20  replace tyear_group = 2 if tyear>=1991 & tyear<=1995
21  replace tyear_group = 3 if tyear>=1996 & tyear<=2000 ⑤
22  replace tyear_group = 4 if tyear>=2001 & tyear<=2005
23  replace tyear_group = 5 if tyear>=2006 & tyear<.
24
25  label var tyear_group "移植年グループ"
26  label define tyear_group_label 1 "1990" 2 "1991-1995" 3 "1996-2000" 4 "2001-2005" 5 "2006-" ⑥⑦⑧
27  label values tyear_group tyear_group_label
28
29  tab tyear_group
30  tab tyear_group, nlabel ⑨⑩
31
32  ****診断から移植までの日数****
33  gen ddate = date(dx_date, "Yd")
34
35  /*Excelファイルからデータを取り込んだ場合、Excel日付データはdate日付データに自動変換されているため、上記の日付コマンド実行は不要です。
36  /*以下のコマンドを実行してください。
37  /**/間のコマンドを実行する場合は、【/*】【*/】を取ってください
38  gen ddate = dx_date
39
40  */
41  gen dx_to_sct = tdate - ddate
42  replace dx_to_sct = . if dx_to_sct<0
43
44  ****診断から移植までの日数グループ****
45  gen dx_to_sct_group = 0 if dx_to_sct<=90
46  replace dx_to_sct_group = 1 if dx_to_sct>90 & dx_to_sct<=180
47  replace dx_to_sct_group = 2 if dx_to_sct>180 & dx_to_sct<.

```

<作成する変数>

Baseline	患者性別	MAST/RIST
移植日と移植年	ドナー性別	TBI
診断から移植までの日数	性別マッチ/ミスマッチ	移植前処置
診断から移植までの日数グループ		GVHD 予防
移植年グループ	患者 ABO 型	CSA/TAC のし死亡
幹細胞種類	ドナー ABO 型	体内 T 細胞除去
ドナー種類	ABO メジャーマッチ/ミスマッチ	
ドナー種類_血縁	ABO ミスマッチ	HLA マッチング
移植種類		
Auto/Allo	疾患名	
移植時患者年齢	疾患分類	
移植時患者年齢グループ		
年齢 50 歳以上	移植時病期	
成人/小児コード	疾患リスク分類	

① set more off

Stata Result ウィンドウでは、出力が長くなるときには出力が止まり、result ウィンドウに"-- more --" という表示が出ます。以降の結果表示を確認するメッセージです。

-- more --の表示上をクリックするか、Enter キーや Space キーを押さない限り続きの結果は表示されません。

長いプログラムを最後まで一度に実行させたい場合には、このコマンドは便利です。