

STATA の使い方

0. 基本事項

(1) 事前準備

Wooldridge の Stata 用データ（拡張子が .dta となっているファイル）をダウンロード。

Stata から読み込みやすいフォルダに保存しておく。ここでは /home0X/zzXXXXX/statafiles というディレクトリにファイルが存在するものとする。

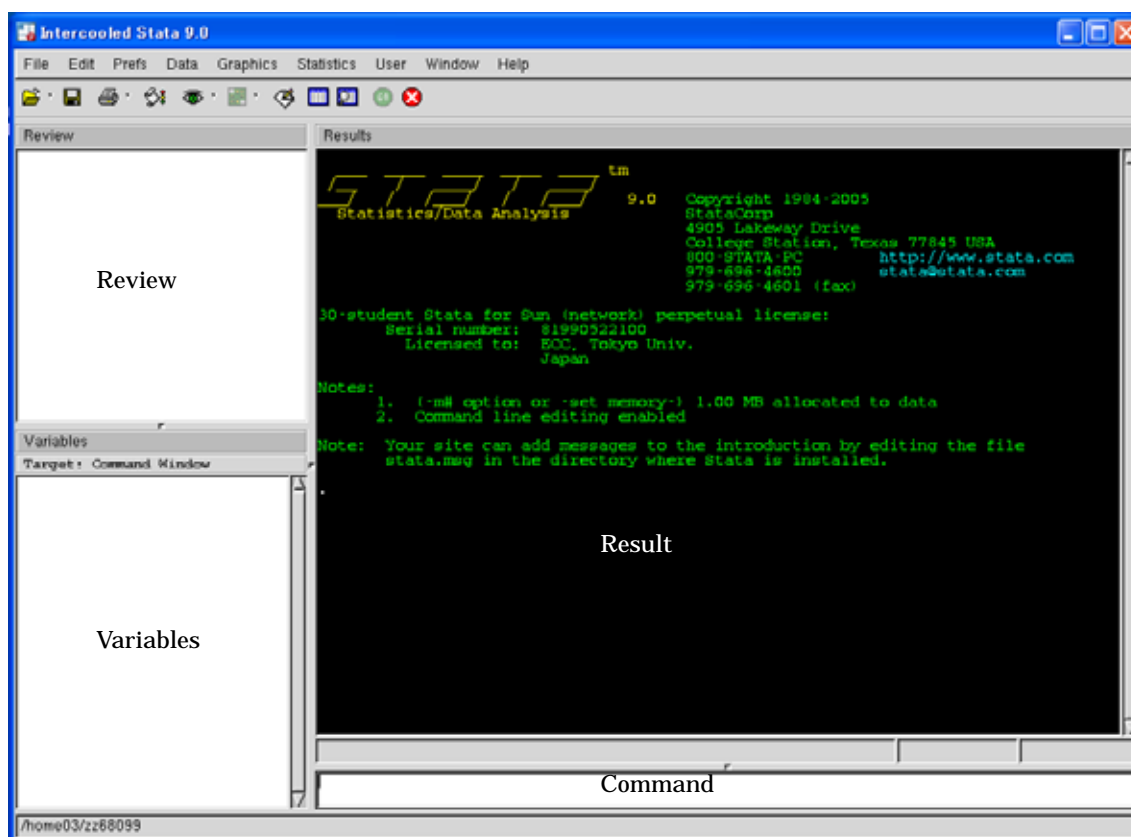
(2) Stata の起動

- ・大学の iMac 環境

画面下のドックから Stata をクリックすれば起動する。

- ・自分の PC を通じて大学の Stata を使用する方法

青木理音さん（昨年の TA）作成のファイル“学外からの Stata 利用法”（ブログにアップします）を参考にしてください



- ・インターフェイスについて

Result アウトプットを表示する

Command コマンドを入力する

Review 実行したコマンドを表示する、クリックすると に再び表示される。

Variables 読み込んだデータの変数が表示される、クリックすると に表示される。

Stata の作業はコマンドを入力して実行する方法とメニューから実行する方法という 2 つの方法があり、当然のことですがどちらも同じです。以下では主にコマンド入力の作業を説明します。GUI (Graphical User Interface) しか使ったことがなくて CUI に不慣れな人はコマンド入力よりもメニューから実行するほうがよいかもしれません。

(3) 作業に取り掛かる前に

・作業ディレクトリの指定

大学のコンピューターから Stata を立ち上げると、/home0X/zzXXXXXX/ が作業ディレクトリとなっているはずですが、

```
pwd [enter]
```

で確認できます。

/home0X/zzXXXXXX/statafiles というディレクトリに存在するファイルを使う場合には

```
cd statafiles [enter]
```

としてディレクトリを変えておくほうが便利です。

このディレクトリにあるファイルを見なければ、

```
ls [enter]
```

とします。

・ファイルの読み込みと保存

.dta 形式のデータファイルの場合

.dta は Stata 独自のファイル形式です。例えば wage1.dta というファイルを読み込む場合には、

```
use wage1.dta [enter]
```

とします。ディレクトリの設定をしていない場合は、パス名も入れなくてはならず、

```
use /home0X/zzXXXXXX/statafiles/wage1.dta [enter]
```

とします。なお、iMac 環境では同じ文字でも大文字・小文字で区別して認識されるので注意してください。

.csv 形式のデータファイルの場合

統計データの多くは csv 形式でインターネット上からダウンロードされます。例えば wage1.csv というファイルを読み込む場合には、

```
insheet using wage1.csv [enter]
```

とします。

既にデータを読み込んでいて、新たに別のデータを読み込みたい場合
読み込んでいるデータを Stata のメモリ上から消去する場合は

```
clear [enter]
```

とする。さらに clear と同時に新たなデータファイルを読み込む場合は、

```
use wage1.dta, clear [enter]
```

とします。

編集したデータを上書き保存したい場合

```
save wage1.dta, replace [enter]
```

・作業メモリの設定

パネルデータなど大きなデータを扱う場合はデフォルトのメモリ量では対応できないこ
ともあります。例えばメモリ量を 30MB に増やしたいときは、

```
set memory 30m [enter]
```

と入力します。

・Result の表示を切り替える

Result ウィンドウは、デフォルトでは自動スクロールされず、

```
--- more ---
```

という表示がでて、その都度 [enter] をしなくてはなりません。アウトプットを一度に表
示させたい場合は

```
set more off [enter]
```

とします (逆の場合は set more on [enter] です)。

・作業記録を残す

作業記録を残す場合は.log ファイルを作成します。例えば.log ファイルを loghw01 という
名前で保存したければ、

```
log using loghw01.log [enter]
```

とします。コマンド末尾の拡張子.log がない場合は.smcl という形式で保存される。秀丸
や Notepad といったテキストエディタで開きたい場合は.log でないときれいに表示されな
い。

作業が終了し、ログをとるのを終えるときは

```
log close [enter]
```

とします。(ログファイルは自動的に作業フォルダに保存されます)

(4) 参考文献など

- ・ Stata の Help や Search コマンド
- ・ 筒井他 (2007) 『Stata で計量経済学入門』 (総合図書館にあり)
- ・ Lawrence C. Hamilton, “Statistics with Stata” (Stata version8 のものは総合図書館と経済学部図書館にあり。現在の最新版は version9 のもの)
- ・ 「ネコでもわかる Stata 入門」 (左のタイトルで検索してみてください)
- ・ 「経済分析のための Stata 入門」 (同上)

(5) 基本コマンド

- ・ 上述のコマンドのまとめ
コマンドに追加されるオプションについては適宜 help や search を見ましょう。

コマンド	説明
cd	changes the working directory
mkdir	creates a new directory (folder)
pwd	displays the path of the current working directory
dir	list the names of files in the specified directory
ls	list the names of files in the specified directory
use	loads a Stata-format dataset (.dta)
insheet	reads text (ASCII) files
save	stores the dataset currently in memory
exit	exit Stata and exit from a program or do-file
set more off	not to display the --more-- message
set memory #[b k m g]	setting the size of memory
log	make a full record of your Stata session
help	when you know the name of the Stata command on which you want informat
search	when you need to find out the name of the command

・ 各種演算子

コマンドに続けて if オプションで条件式を書く際になどに使用します。

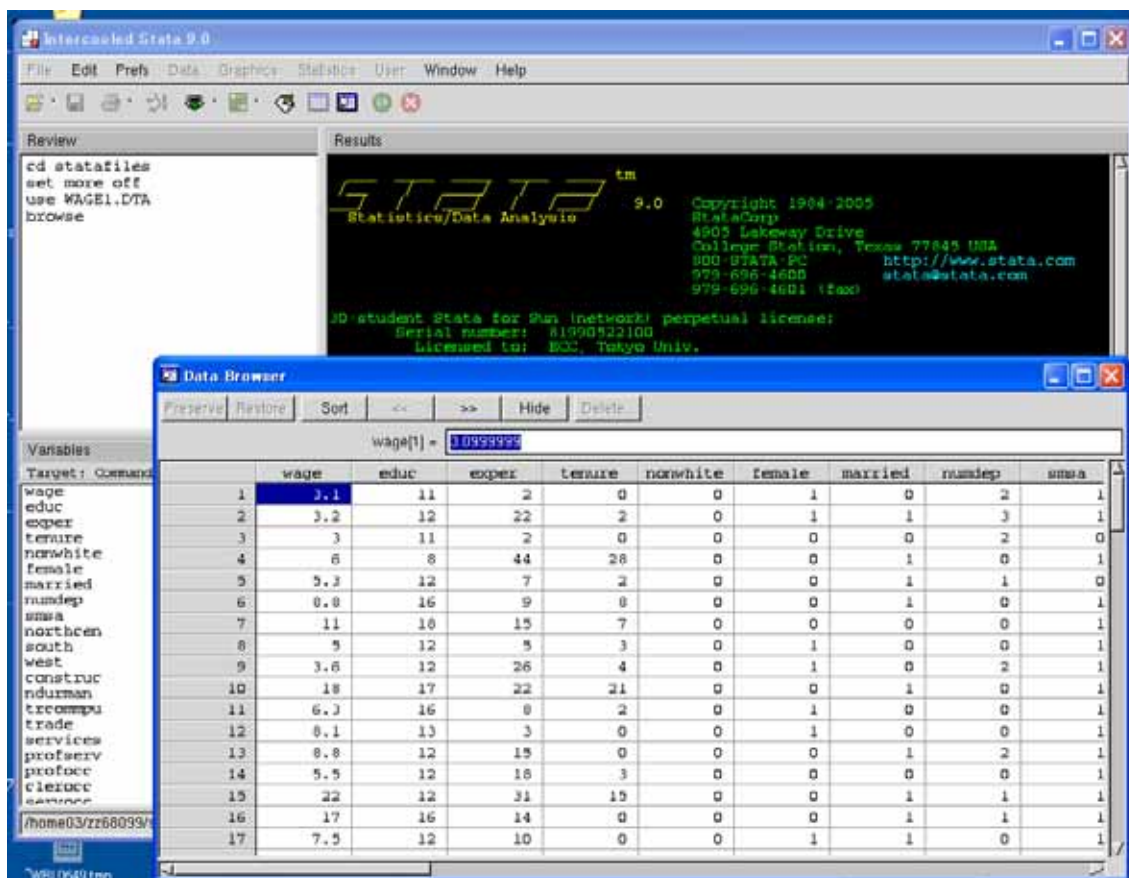
数式	論理式	等号・不等号
+ addition	~ not	> greater than
- subtraction	! not	< less than
* multiplication	or	>= > or equal
/ division	& and	<= < or equal
^ power		== equal
		~= not equal
		!= not equal

1. データと基本統計量

(1) データの確認

- Data Browser でのデータの確認

wage1.dta を例とします。データファイルを読み込んだら、コマンドウィンドウに browse と入力してみましょう。すると以下のように Data Browser が立ち上がります。Data Browser では、Excel などのスプレッドシートのように直接データを加工することはできません。



また

```
browse wage educ exper [enter]
```

と入力すると、wage・educ・exper のみのデータが表示されます。

- Result ウィンドウ上でのデータの確認

list 変数の内容が表示される

```
list wage educ exper [enter]
```

describe 標本数、変数の属性などが表示される

(2) 基本統計量など

- summarize 変数の基本統計量が表示されます

```
sum wage educ exper [enter]
```

とすると、

```
. sum wage educ exper
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	526	5.896103	3.693086	.53	24.98
educ	526	12.56274	2.769022	0	18
exper	526	17.01711	13.57216	1	51

それぞれの変数について標本数・平均・標準偏差・最小値・最大値が表示されます。

より詳しい統計量が必要な場合は detail オプションをつけることで、

```
sum wage educ exper, detail [enter]
```

分散・歪度・尖度・分位点などが得られます。

・ table 度数を表示します。例えば

```
table nonwhite [enter]
```

とすると、次のようなアウトプットとなります。

```
. table nonwhite
```

=1 if nonwhite	Freq.
0	472
1	54

同様のことですが、nonwhite の数のみを知りたい場合には、

```
count if nonwhite==1 [enter]
```

としても値が得られます。

・ tabulate 度数分布表が表示されます。変数を 1 つで実行すると、

```
tabulate nonwhite [enter]
```

で度数・相対度数・累積相対度数が表示されます。

また変数を 2 つで実行すると、

```
tabulate female married [enter]
```

で 2 次元の度数分布表が得られます。

```
. tabulate nonwhite
```

=1 if nonwhite	Freq.	Percent	Cum.
0	472	89.73	89.73
1	54	10.27	100.00
Total	526	100.00	

```
. tabulate female married
```

=1 if female	=1 if married		Total
	0	1	
0	86	188	274
1	120	132	252
Total	206	320	526

- Stata ではある作業（推定や検定など）を行ったあとに、主な統計量がメモリに保存された状態になっています（次の作業を行うとメモリから消去されます）。return list あるいは ereturn list（回帰分析などの結果のリストを表示するとき）コマンドで表示をすることができます。

sum wage [enter]

return list [enter]

```

sum wage
-----+-----
Variable |      Obs      Mean    Std. Dev.    Min    Max
-----+-----
wage     |      526    5.896103    3.693086     .53    24.98

return list
Scalars:
      r(N) = 526
      r(sum_w) = 526
      r(mean) = 5.896102674787035
      r(var) = 13.63888436377501
      r(sd) = 3.693086021713413
      r(min) = .5299999713897705
      r(max) = 24.97999954223633
      r(sum) = 3101.350006937981

```

- `display` 計算などの結果を Results ウィンドウに表示します。

disp 5*4 [enter]

```

. disp 5*4
20

```

sum wage [enter]

disp r(mean)+3 [enter]

```

. sum wage
-----+-----
Variable |      Obs      Mean    Std. Dev.    Min    Max
-----+-----
wage     |      526    5.896103    3.693086     .53    24.98

. disp r(mean)+3
8.8961027

```

(3) グラフ

Stata では様々なグラフを作成することができます。その一部について触れます。グラフについては Hamilton の本に詳しいです。

- ヒストグラム

histogram wage, option [enter]

option には、frequency: 縦軸は頻度を表示、percent: 縦軸はパーセント表示、norm: 正規分布曲線を表示（上記の場合、平均・標準偏差は wage に依るもの）、などがあります。

- 2次元のグラフ

graph twoway option wage educ [enter]

option には、scatter: 散布図、line: 折れ線図、などがあります。

(4) データの編集

edit [enter]

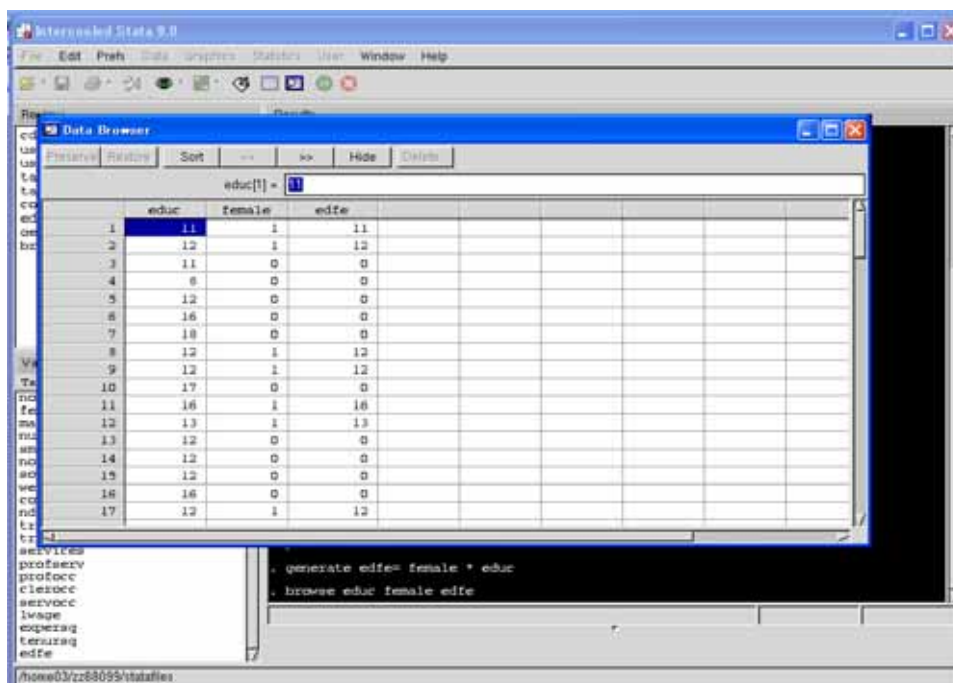
で Data Edit Window が開きます。Data Edit Window 上でデータを変更することができますが、あまり使用する機会はないと思います。

(5) 変数の加工

・generate 新たな変数を作るときや、変数を加工する場合に使用します。例えば、変数 educ とダミー変数 female を掛け合わせた変数 edfe (交差項、cross term) を作りたいとします。この場合は、

```
generate edfe = educ * female [enter]
```

とすれば新しい変数 edfe が追加されます。



・egen : generate コマンドには使えないいくつかの関数を使うことができるコマンド。

```
egen newvar = function(variable) [enter]
```

function のところには、関数を書き込みます。利用できる関数の主なものとしては、

mean 平均値

sum 合計

max 最大値

min 最小値

例えば

```
egen avg = mean(educ) [enter]
```

とすると、

	educ	avg		
1	11	12.56274		
2	12	12.56274		
3	11	12.56274		
4	8	12.56274		
5	12	12.56274		
6	16	12.56274		
7	18	12.56274		
8	12	12.56274		
9	12	12.56274		
10	17	12.56274		
11	16	12.56274		

のようになります。

- `replace` すでに存在する変数の値などを書き換える時に使用します。例えば `construc` という変数に欠損値があってセルが “.” となっているものがあるとします。この欠損値を 0 という値に変えたい場合は、

```
replace construc = 0 if construc ==. [enter]
```

とします。また、10 行目の人のデータの `construc` という変数の値を 1 に変えたい場合には、

```
replace construc = 1 in 10 [enter]
```

とします。

- `label` 新しく作った変数などにラベルをつけるときなどに使用します。先ほど `generate` コマンドを使って `edfe` という変数を新しく作りましたが、このような新しく作った変数の混同を防ぐために `label` をつけることができます。

```
label variable edfe "cross term of educ and female" [enter]
```

で、`describe` コマンドを実行すると、

```
. label variable edfe "cross term of educ and female"
. describe edfe
variable name      storage  display  value  variable label
type              format  label
-----
edfe              float   %9.0g   cross term of educ and female
```

と表示されます。

- `preserve` および `restore` あくまで一時的にデータを変形させて作業を行いたいときに使うコマンド。`save` コマンドでデータを上書きするよりも気軽に使えます。

```
preserve [enter]
```

```
replace construc = 0 if construc ==. [enter]
```

```
browse [enter]
```

```
restore [enter]
```

(6) データセット、random variables の作成の一例

- Stata でデータセット自体を一から作成する場合、まず観測値の数を設定する必要があります。もし 500 の観測値を設定しようとするなら、以下のコマンドを用います。

```
set obs 500 [enter]
```

- `uniform()` Stata でランダムサンプルを抽出する際に観測値を作るためによく使われる function。[0, 1) の一様分布からサンプルを抽出します。

```
set obs 100 [enter]
```

```
gen randnum = uniform() [enter]
```

と入力すると、[0, 1) の範囲から 100 の観測値をランダムに抽出し、それを `randnum` と定義することができます。

- (標準) 正規分布からランダムサンプルを抽出する作業をしばしば行うことがあります。その際には、標準正規分布の値を返す際に使用する function である `invnorm()` と、上記の `uniform()` を組み合わせて使います。

```
set obs 100 [enter]
```

```
gen var1 = 1 + invnorm(uniform())*2 [enter]
```

と入力することで、 $N(1, 2^2)$ から 100 個のランダムサンプルを抽出し、それを `var1` と定義することができます。

(7) その他関数など

- `Binominal(n,k,p)` 成功確率が p であるベルヌーイ試行を n 回行うときに k 回以上成功する確率を計算します。

- `norm(z)`、`invnorm(p)` 標準正規分布に関する関数。`norm(z)` では標準正規分布の中での任意の値 z における累積確率 p を、`invnorm(p)` は累積確率が p となる値 z を示します。

```
disp norm(0) [enter]
```

```
disp invnorm(0.5) [enter]
```

```
. disp norm(0)
.5
. disp invnorm(0.5)
0
```

・ $t_{tail}(n,t)$ 、 $inv_{ttail}(n,p)$ t 分布に関する関数。 $t_{tail}(n,t)$ は自由度 n の t 分布の中での任意の値 t における右片側の確率 p を、 $inv_{ttail}(n,p)$ は右片側の確率 p となる値 t を示します。

・ $F_{tail}(n1,n2,f)$ 、 $invF_{tail}(n1,n2,p)$ F 分布に関する関数。 $F_{tail}(n1,n2,f)$ は自由度($n1,n2$)の F 分布の中での任意の値 f における右片側の確率 p を、 $invF_{tail}(n1,n2,p)$ は右片側の確率 p となる値 f を示します。