

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

取り込みデータをStata形式で保存する方法 (方法1)

取り込みデータをStata形式で保存する方法 (方法2)

save C:\保存ファイル名\保存ファイル名.dta, オプションコマンド

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

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…

Statal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

67

ログファイルの作成: Do-file Editor

log using logファイル名
ログ記録開始のコマンド

この間の作業結果を記録する

log close
ログ記録終了のコマンド

68

ログファイルの作成: Do-file Editor

translate ファイル名 .smcl ファイル名 .txt

text形式でのファイル保存コマンド

StataのViewerで読み取れるファイルコマンド

69

ログファイルの作成: Result window

ローカルディスク(C:)のdataフォルダに保存された
Smcl形式で保存された

テキスト形式でもログファイルが保存された

70

ログファイルの保存

デフォルトでは、使用データファイル(.dta)と同じフォルダ的に、ログファイルは保存されます。
(既定先フォルダ=作業フォルダ)

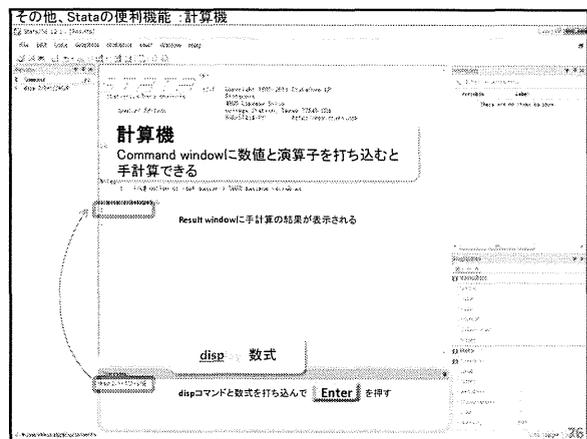
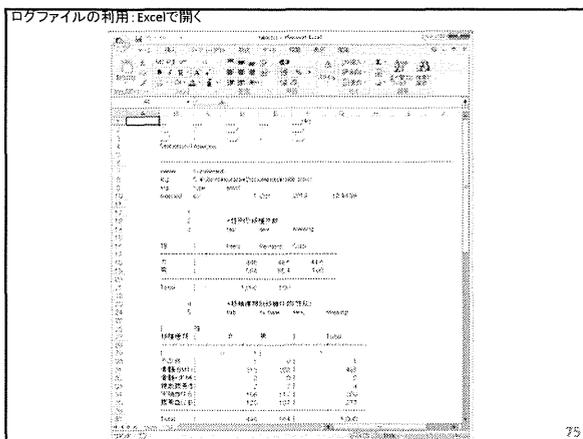
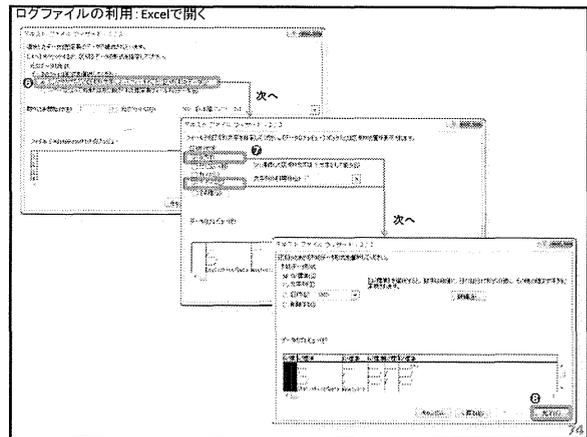
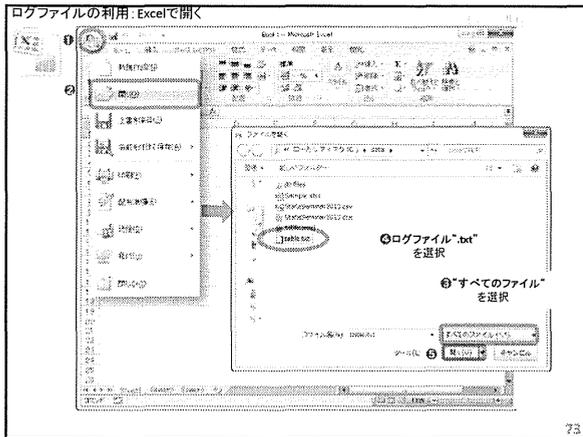
71

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

72



```

1
2 *****
3 * データを取り込む前の準備 *
4 *****
5
6 ///* データセットの前準備 *///
7 * ExcelまたはCSVのファイル名を英語名にしてください。(TRUMP2013 など)
8 * 項目が英語名のデータセットを使用してください。
9 * 2010年度データセットに関しては、JSHCT HPに英語項目名変換用のエクセルマクロを使用してください。
  (JSHCTホームページよりダウンロード可能)
10 * 最終行に入力率が入っている場合はこれを削除してください。
11 * WGデータセット配布時に設定されている開封パスワードを解除し、保存したデータを使用してください。
12
13
14 *****
15 * データの取り込み *
16 *****
17
18 ///* Excelファイルからの取り込み *///
19 * Stata12からはエクセルファイルのまま取り込みが可能です。
20 * Excelのファイルサイズが40メガバイトを超える場合には、Excelファイルからの取り込みはできません。
21
22 ////* CSVファイルからの取り込み *///
23 *
  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。
24 * →CSVファイルから取り込んでください。
25
26
27 *****データ取り込みコマンド*****
28
29 * insheet using C:¥データファイルの保存先フォルダ名¥データファイル名, オプションコマンド
30
31 ***** 以下に記載するコマンドの解説 *****
32 * 取り込みコマンドのusing以降はファイルの場所(directory)を示します。
33 *
  directoryは、あらかじめローカルディスクCの中に"data"というフォルダを作成し、元データを保存してある場所です。
34 * 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。
35 * ちなみに、オプションコマンドとは、カンマ以降のコマンドのことです。
36
37 ***** ミニ知識 *****
38 * 米印が先頭に来る場合、あるいは/*と*/で囲んだところは、メモ書きとしてStataは認識します。
39
40
41 ***** CSVファイルからの取り込みコマンド *****
42 *for Windows
43 insheet using C:¥data¥TRUMPSeminar2013.csv, name
44
45 *for Mac (※OSのversionによって取り込み方法が異なる場合があります)

```

```
46 insheet using /c/data/TRUMPSeminar2013.csv, name
47
48 *読み込みに多少時間はかかります、あせらないでくださいね！！
49
50 ***** Excelファイル(xlsファイル)からの取り込みコマンド *****
51 import excel using C:\data\TRUMPSeminar2013.xls, firstrow
52
53 ***** エクセルファイル(xlsxファイル)からの取り込みコマンド *****
54 import excel using C:\data\TRUMPSeminar2013.xlsx, firstrow
55
56
57 *****
58 * 変数にラベルを付ける *
59 *****
60
61 *取り込んだ変数の内容がわかるように変数ラベルを付けると便利です。
62
63
64 ///* ラベルの付け方の一例 *///
65 /*変数survivalに日本語ラベル"生存状況"を付けます。
66 label var survival "生存状況"
67 *変数sexに日本語ラベル"性"を付けます。
68 label var sex "性"
69 *変数raceに日本語ラベル"人種"を付けます。
70 label var race "人種"
71 */
72
73 *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できます
74
75
76 ***** doファイルを使って、ラベルを付けるコマンドを実行する *****
77 *配布したDoファイル[01_label_variable_wg17-21_XXXXXX.do]をDo-file Editorにて開き、実行してください。
78
79
80 *****
81 * 取り込んだデータの確認 *
82 *****
83
84 ***** 取り込んだデータの確認方法 *****
85
86 * 取り込んだ内容を直接確認する
87 * Data Editor(Browse)を開いて全内容を見ることができます
88
89 * 標本数と変数の属性を確認する (各変数のタイプ、ラベル名、観測値、変数の数、ファイルサイズなど)
90 *使用するコマンド：describeを用いる (例：変数survival,no,tx_type)
91 de survival no tx_type
92
93 * 連続変数の要約値を確認する (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)
```

```
94 *使用するコマンド：summarize (例：連続変数age_month)
95   sum age_month
96 *オプションコマンドdetailを加えると、中央値、パーセンタイル値、歪度、尖度を算出することができます。
97   sum age_month, detail
98
99 * カテゴリー変数の要約値を確認する (各カテゴリーの頻度、割合)
100 *使用するコマンド：tabulate (例：カテゴリー変数tx_type)
101   tab tx_type
102 *オプションコマンドmissingを加えると、欠損値のデータ数を算出することができます。
103   tab tx_type, missing
104 *tabコマンドの後ろに2つのカテゴリー変数を指定するとクロス集計を行うことができます。
105   tab tx_type no
106
107 * データのサンプル数(観測値)を確認する
108 *使用するコマンド：count
109   count
110
111 *リストを作成してデータの内容を確認する
112 *使用するコマンド：list
113 *指定した範囲のリストを確認する場合：if 条件式
114   **(例：移植日が2000年、2001年の移植例の移植種類と移植日)
115   list tx_type tx_date if tx_date>="2000/1/1" & tx_date<="2001/12/31"
116   /*
117   result windowに表示される内容が多い場合、result windowに-more-の文字が表示されます。
118   -more-の文字上をクリックするか、EnterキーまたはSpaceキーを押すと結果の続きが表示されます。
119   */
120
121
122 *****
123 * Stata形式でのデータの保存 *
124 *****
125
126 *****Stata形式でのデータ保存コマンド*****
127 **取り込みデータファイル(ExcelまたはCSV)と同じ場所(ローカルディスクCのdataフォルダ)に同名で保存する。(
  半角英数字)
128 save C:¥data¥TRUMPSeminar2013.dta
129
130 /*
131 **既存のStata形式データに上書き保存する場合
132 **オプションコマンド replaceを付ける
133 save C:¥data¥TRUMPSeminar2013.dta, replace
134 */
135
136
137 *****
138 * Stata形式データの読み込み *
139 *****
140
141 ***** Stata形式で保存した.dtaファイルは以下の方法で読み込む *****
142
```

```
143 * 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
144
145 * .dtaファイルの読み込みコマンドにて開く
146 *use C:\%data%\TRUMPSeminar2013.dta, clear
147 /* オプションコマンドのclearを付けると、既に読み込んであるデータをStataのメモリから消して、
148 同時に新たなデータを読み込むことができます。
149 (開いているStataを閉じることなく、新たな保存データ.dtaを読み込む)
150 */
151
152
153 *****
154 * ログファイルの作成と利用 *
155 *****
156
157 *ログファイルとは、Stataの作業結果を記録したファイルのことです。
158
159 ///  
160 *作業結果はResult windowに表示されますが、いくつかの作業を行い結果が長くなると遡って表示されません。
161 *ログファイルを作成すると、すべての作業結果を保存することができ便利です。
162
163 ///  
164 *ログをtext形式で保存すると、メモ帳などで開くこともできます。
165
166 *ログをtext形式で保存し、Excelファイルから開くと、結果tableを作成する際に便利です。
167
168 ***** ログファイルの作成方法 *****
169 /*
170 log useing ログファイル名
171 xxx
172 log off
173 yyy
174 log on
175 zzz
176 log close
177 translate ログファイル名.smcl ログファイル名.txt
178 */
179
180 *ログファイルに「xxx」「zzz」の部分の記録はされますが、「yyy」部分は記録されません。
181
182
183 *ログファイルを作成し、Excelファイルでtableを作成
184 *例:データの要約値をログファイル[result]に記録して、tableを作成する。
185 log using table
186
187 *性別の移植件数
188 tab sex, missing
189 *移植種別移植件数(性別)
190 tab tx_type sex, missing
191
192 *患者性別、年齢 (Mann-Whitney検定)
```

```
193 ranksum age, by(sex)
194
195 *患者性別、移植種類 (chi-square検定)
196 tab sex do_sex, chi2
197
198 log close
199 translate table.smcl table.txt
200
201 ***** textファイルで保存されたlogファイルをExcelで開く方法 *****
202 /* Excel
   開く→textファイル開く→[カンマやタブなどの区切り文字によってフィールドごとに区切られたデータ]を選択
   →次へ→[タブ] [スペース]にチェックを入れる→次へ→完了
203
204 */
205
```

変数の作成の基本的事項

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

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
鎌塚八千代

配布資料

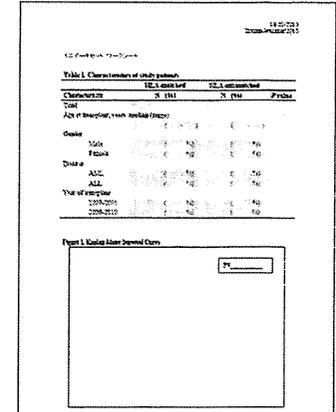
- ① Minidataset_2013 (Stata Dataset)
- ② Minidataset_2013 (Excel)
- ③ S1_minidata_2013
(Stata Do-file・プリント)
- ④ ミニデータワークシート_2013
(Excel・プリント)

minidataset_2013.xls

1	2	3	4	5	6	7	8	9	10	11	12
sex	test_date	tx_date	age	donor	dx	diagnosis	disease_s	seromis	Sabd2		
1	男	2001/8/18	2001/1/26	11	同種_alog	急性リン	2	1	0		
2	男	2011/7/11	2002/8/9	27	同種_alog	急性骨髄	1	4	2		
3	女	2007/7/21	2007/3/20	5	同種_alog	急性リン	2	4	1		
4	女	2011/8/2	2009/8/19	5	同種_alog	急性リン	2	1	0		
5	男	2011/8/1	2010/10/1	10	同種_alog	急性骨髄	1	2	0		
6	男	2007/5/7	2006/5/29	16	同種_alog	急性骨髄	1	4	0		
7	男	2005/12/2	2004/5/21	10	同種_alog	急性骨髄	1	1	0		
8	男	2004/1/11	2004/10/1	5	同種_alog	急性骨髄	1	1	1		
9	女	2011/8/17	2008/1/25	21	同種_alog	急性リン	2	1	0		
10	女	2007/11/2	2007/3/6	13	同種_alog	急性骨髄	1	1	0		
11	男	2001/2/13	2000/10/2	14	同種_alog	急性骨髄	1	1	0		
12	男	2009/3/11	2007/1/03	2	同種_alog	急性リン	2	4	0		
13	男	2001/3/13	2000/9/13	15	同種_alog	急性骨髄	4	4	0		
14	女	2007/1/26	2003/10/2	3	同種_alog	急性リン	2	2	1		
15	男	2007/8/13	2003/9/10	7	同種_alog	急性リン	2	4	0		
16	男	2007/8/13	2003/8/18	20	同種_alog	急性リン	2	4	1		
17	男	2008/5/12	2006/3/16	18	同種_alog	急性リン	2	1	1		
18	男	2004/10/1	2004/3/31	1	同種_alog	急性リン	2	4	0		
19	男	2002/2/21	2002/1/19	3	同種_alog	急性リン	2	1	0		
20	女	2011/4/8	2010/12/2	14	同種_alog	急性リン	2	1	1		

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



minidataset_2013.dta

The screenshot shows the Stata command window and data editor. The command window shows the command 'use minidataset_2013.dta'. The data editor shows the same data as the Excel file, with columns for 'survival', 'sex', 'test_date', 'tx_date', 'donor', and 'dx'.

S1_minidata_2013.do Stata Do-file & 印刷資料

```

1 *** 印刷資料 ***
2 *** 印刷資料 ***
3 *** 印刷資料 ***
4 *** 印刷資料 ***
5 *** 印刷資料 ***
6 *** 印刷資料 ***
7 *** 印刷資料 ***
8 *** 印刷資料 ***
9 *** 印刷資料 ***
10 *** 印刷資料 ***
11 *** 印刷資料 ***
12 *** 印刷資料 ***
13 *** 印刷資料 ***
14 *** 印刷資料 ***
15 *** 印刷資料 ***
16 *** 印刷資料 ***
17 *** 印刷資料 ***
18 *** 印刷資料 ***
19 *** 印刷資料 ***
20 *** 印刷資料 ***
21 *** 印刷資料 ***
22 *** 印刷資料 ***
23 *** 印刷資料 ***
24 *** 印刷資料 ***
25 *** 印刷資料 ***

```

Stata起動とエクセルデータの取り込み

- Stata を起動
- ログファイルを作成
– 方法は問いません
- minidataset_2013.xls をStata に取り込み
– 方法は問いません

参考: TRUMP統計セミナー基礎編資料
S1_Minidata_2013 do file 7-9行目

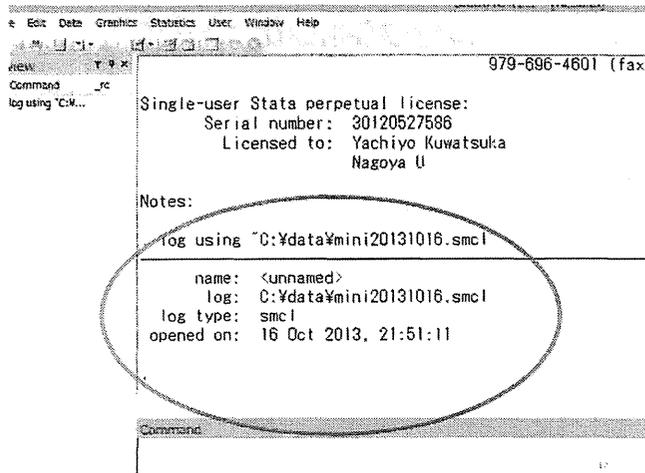
ログファイルの作成 ①

配布資料: S1¥Minidata_2013: 8行目

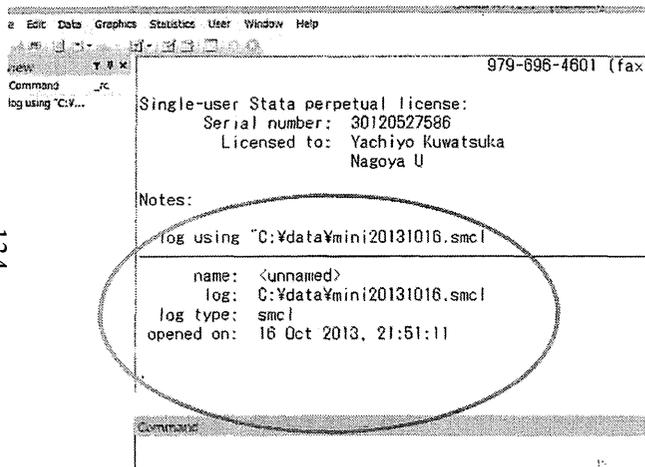
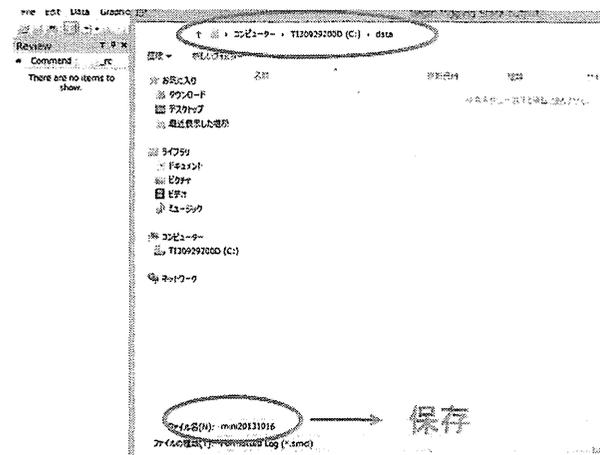
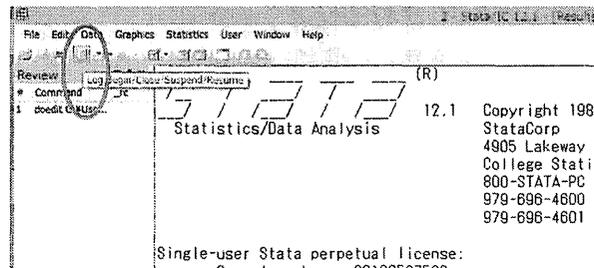
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

Single-user Stata perpetual license:
Serial num
Licensec
suka

log using "C:\¥data¥mini20131016.smcl"

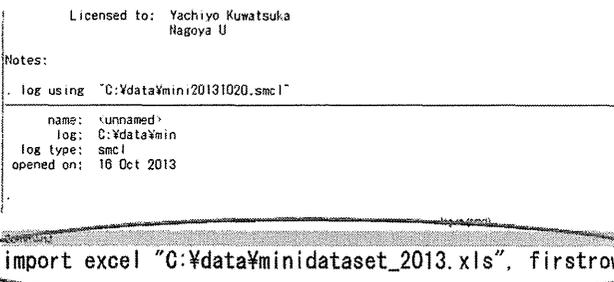


ログファイルの作成 ②



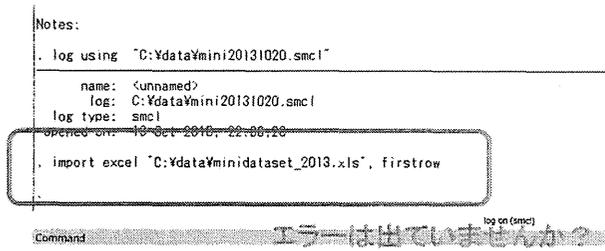
エクセルデータ取り込み (コマンド使用)

S1\Miniadata_2013: 7行目



エクセルデータ取り込み (コマンド使用)

S1\Miniadata_2013: 7行目



- 134 -

新たな変数作成の練習 (準備運動)

S1\Miniadata_2013: 11~28行目

generate, replace

新たな変数の作成・データの置換

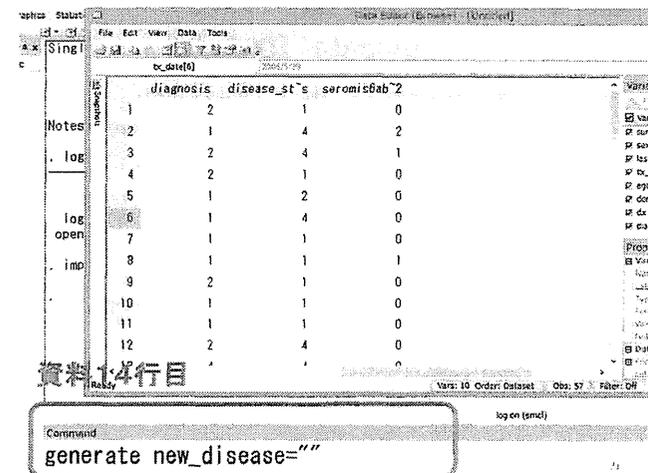
drop, keep

変数・症例の削除

新たな変数作成の練習

S1\Miniadata_2013: 11~22行目

```
11 //////////////**新たな変数作成の練習**/////////////////
12 * 準備運動 *
13 /**generate, replace*/ 文字のデータ
14 generate new_disease=""
15 *新たな変数new_diseaseが作成された
16 replace new_disease=dx
17 replace new_disease="AML" if diagnosis==1
18 replace new_disease="ALL" if diagnosis==2
19
20 generate new_disease2="" 数字のデータ
21 replace new_disease2=1 if new_disease=="AML"
22 replace new_disease2=2 if new_disease=="ALL"
```



dx	diagnosis	disease_st's	seromisabab'2	new_disease
1	2	1	0	
2	1	4	2	
3	2	4	1	
4	2	1	0	
5	1	2	0	
6	1	4	0	
7	1	1	0	
8	1	1	1	

```
.generate new_disease=
(57 missing values generated)
```

```
Command
```

dx	diagnosis	disease_st's	seromisabab'2	new_disease	
1	急性リンパ性白血病(ALL)	2	1	0	
2	急性骨髄性白血病(AML)	1	4	2	
3	急性リンパ性白血病(ALL)	2	4	1	
4	急性リンパ性白血病(ALL)	2	1	0	
5	急性骨髄性白血病(AML)	1	2	0	
6	急性骨髄性白血病(AML)	1	4	0	
7	急性骨髄性白血病(AML)	1	1	0	
8	急性骨髄性白血病(AML)	1	1	1	

```
.generate new_disease=""
(57 missing values generated)
```

```
Command
```

```
replace new_disease=dx 資料16行目
```

dx	diagnosis	disease_st's	seromisabab'2	new_disease	
19	急性リンパ性白血病(ALL)	2	1	0	急性リンパ性白血病(ALL)
20	急性リンパ性白血病(ALL)	2	1	1	急性リンパ性白血病(ALL)
21	急性骨髄性白血病(AML)	1	2	0	急性骨髄性白血病(AML)
22	急性リンパ性白血病(ALL)	2	4	0	急性リンパ性白血病(ALL)
23	慢性骨髄性白血病(CML)	4	4	1	慢性骨髄性白血病(CML)
24	急性リンパ性白血病(ALL)	2	1	0	急性リンパ性白血病(ALL)
25	急性骨髄性白血病(AML)	1	1	2	急性骨髄性白血病(AML)
26	急性リンパ性白血病(ALL)	2	3	0	急性リンパ性白血病(ALL)

```
.replace new_disease=dx
new_disease was str1 now str23
(57 real changes made)
```

```
Command
```

dx	diagnosis	disease_st's	seromisabab'2	new_disease	
19	急性リンパ性白血病(ALL)	2	1	0	急性リンパ性白血病
20	急性リンパ性白血病(ALL)	2	1	0	急性リンパ性白血病
21	急性骨髄性白血病(AML)	1	2	0	急性骨髄性白血病
22	急性リンパ性白血病(ALL)	2	4	0	急性リンパ性白血病
23	慢性骨髄性白血病(CML)	4	4	1	慢性骨髄性白血病
24	急性リンパ性白血病(ALL)	2	1	0	急性リンパ性白血病
25	急性骨髄性白血病(AML)	1	2	2	急性骨髄性白血病
26	急性リンパ性白血病(ALL)	2	3	0	急性リンパ性白血病
27	急性骨髄性白血病(AML)	1	0	0	急性骨髄性白血病
28	急性リンパ性白血病(ALL)	2	4	0	急性リンパ性白血病
29	急性骨髄性白血病(AML)	1	4	1	急性骨髄性白血病
30	急性骨髄性白血病(AML)	1	1	0	急性骨髄性白血病
31	急性骨髄性白血病(AML)	1	2	0	急性骨髄性白血病

```
replace new_disease="AML" if diagnosis==1
```

資料17行目

dx	diagnosis	disease_st's	seromisabab'2	new_disease	
19	急性リンパ性白血病(ALL)	2	1	0	急性リンパ性白血病(ALL)
20	急性リンパ性白血病(ALL)	2	1	0	急性リンパ性白血病(ALL)
21	急性骨髄性白血病(AML)	1	2	0	AML
22	急性リンパ性白血病(ALL)	2	4	0	急性リンパ性白血病(ALL)
23	慢性骨髄性白血病(CML)	4	4	1	慢性骨髄性白血病(CML)
24	急性リンパ性白血病(ALL)	2	1	0	急性リンパ性白血病(ALL)
25	急性骨髄性白血病(AML)	1	1	2	AML
26	急性リンパ性白血病(ALL)	2	3	0	急性リンパ性白血病(ALL)
27	急性骨髄性白血病(AML)	1	1	0	AML

```
replace new_disease="AML" if diagnosis==1
(24 real changes made)
```

```
Command
```

```
replace new_disease="ALL" if diagnosis==2
```

disease_st's	seromisabab'2	new_disease	new_disease2
20	1	ALL	
21	2	AML	
22	4	ALL	
23	4	慢性骨髄性白血病(CML)	
24	1	ALL	
25	1	AML	
26	3	ALL	

```
.generate new_disease2=
(57 missing values generated)
```

```
Command
```

-135-

dx	diagnosis	disease_st's	seromisabab'2	new_disease
3	急性リンパ性白血病(ALL)	2	1	ALL
1	急性骨髄性白血病(AML)	1	2	AML
2	急性リンパ性白血病(ALL)	2	0	ALL
3	慢性骨髄性白血病(CML)	4	1	慢性骨髄性白血病(CML)
4	急性リンパ性白血病(ALL)	2	1	ALL
5	急性骨髄性白血病(AML)	1	1	AML
3	急性リンパ性白血病(ALL)	2	3	ALL
7	急性骨髄性白血病(AML)	1	1	AML
8	急性リンパ性白血病(ALL)	2	4	ALL

```
replace new_disease="ALL" if diagnosis==2
(30 real changes made)
```

'diagnosis' が2の症例の 'new_disease' は"ALL"に置き換わりましたか?

dx	diagnosis	disease_st's	seromisabab'2	new_disease
3	急性リンパ性白血病(ALL)	2	1	ALL
1	急性骨髄性白血病(AML)	1	2	AML
2	急性リンパ性白血病(ALL)	2	0	ALL
3	慢性骨髄性白血病(CML)	4	1	慢性骨髄性白血病(CML)
4	急性リンパ性白血病(ALL)	2	1	ALL
5	急性骨髄性白血病(AML)	1	1	AML
3	急性リンパ性白血病(ALL)	2	3	ALL
7	急性骨髄性白血病(AML)	1	1	AML
8	急性リンパ性白血病(ALL)	2	4	ALL

```
.generate new_disease=""
(57 missing values generated)
```

```
.replace new_disease=dx
new_disease was str1 now str23
(57 real changes made)
```

```
.replace new_disease="AML" if diagnosis==1
(24 real changes made)
```

```
.replace new_disease="ALL" if diagnosis==2
(30 real changes made)
```

```
Command
```

```
generate new_disease2= 資料20行目
```

disease_st's	seromisabab'2	new_disease	new_disease2
20	1	ALL	
21	2	AML	
22	4	ALL	
23	4	慢性骨髄性白血病(CML)	
24	1	ALL	
25	1	AML	
26	3	ALL	

```
.generate new_disease2=
(57 missing values generated)
```

```
Command
```

Stata/IC 12.1 - [Results]

Obs	disease_st's	seromis6ab~2	new_disease	new_disease2
20	1	1	ALL	.
21	2	0	AML	.
22	4	0	ALL	.
23	4	1	慢性骨髄性白血病 (CML)	.
24	1	0	ALL	.
25	1	2	AML	.
26	3	0	ALL	.
27	1	0	AML	.
28	4	0	ALL	.
29	4	1	AML	.

replace new_disease2=1 if new_disease=="AML"

資料21行目

新たな変数作成の練習

配布資料: S1¥Minidata_201323~28行目

```

23 ***症例削除の練習***
24 **drop, keep**/
25 drop new_disease
26 * 変数new_diseaseが削除される。戻らないの
27 drop if diagnosis==1          で注意！
28 keep if diagnosis==4

```

136

Stata/IC 12.1 - [Results]

dx	diagnosis	disease_st's	seromis6ab~2	new_di
1	急性リンパ性白血病 (ALL)	2	1	0
2	急性骨髄性白血病 (AML)	1	4	2
3	急性リンパ性白血病 (ALL)	2	4	1
4	急性リンパ性白血病 (ALL)	2	1	0
5	急性骨髄性白血病 (AML)	1	2	0
6	急性骨髄性白血病 (AML)	1	4	0
7	急性骨髄性白血病 (AML)	1	1	0
8	急性骨髄性白血病 (AML)	1	1	1
9	急性リンパ性白血病 (ALL)	2	1	0

drop if diagnosis==1

資料27行目

```

(57 real changes made)
. replace new_disease="AML" if diagnosis==1
(24 real changes made)

. replace new_disease="ALL" if diagnosis==2
(30 real changes made)

. generate new_disease2=.
(57 missing values generated)

. replace new_disease2=1 if new_disease=="AML"
(24 real changes made)

```

replace new_disease2=2 if new_disease=="ALL"

資料22行目

Stata/IC 12.1 - [Results]

Obs	disease_st's	seromis6ab~2	new_disease	new_disease2
11	1	1	AML	1
12	4	0	ALL	2
13	4	0	慢性骨髄性白血病 (CML)	.
14	2	0	ALL	2
15	4	0	ALL	2
16	4	0	ALL	2
17	1	0	ALL	2
18	4	0	ALL	2
19	1	0	ALL	2

drop new_disease

資料25行目

Stata/IC 12.1 - [Results]

dx	diagnosis	disease_st's	seromis6ab~2	new_c
1	急性リンパ性白血病 (ALL)	2	1	0
2	急性リンパ性白血病 (ALL)	2	4	1
3	急性リンパ性白血病 (ALL)	2	1	0
4	急性リンパ性白血病 (ALL)	2	1	0
5	急性リンパ性白血病 (ALL)	2	4	0
6	慢性骨髄性白血病 (CML)	4	4	0
7	急性リンパ性白血病 (ALL)	2	2	1

. drop if diagnosis==1
(24 observations deleted)

Stata/IC 12.1 - [Results]

disease_st's	seromis6ab~2	new_disease	new_disease2
1	1	ALL	2
2	0	AML	1
4	0	ALL	2
4	1	慢性骨髄性白血病 (CML)	.
1	0	ALL	2
1	2	AML	1
3	0	ALL	2
1	0	AML	1

replace new_disease2=1 if new_disease=="AML"
1 changes made)

Stata/IC 12.1 - [Results]

dx	diagnosis	disease_st's	seromis6ab~2	new_disease2
1	急性リンパ性白血病 (ALL)	2	1	0
2	急性骨髄性白血病 (AML)	1	4	2
3	急性リンパ性白血病 (ALL)	2	4	1
4	急性リンパ性白血病 (ALL)	2	1	0
5	急性骨髄性白血病 (AML)	1	2	0
6	急性骨髄性白血病 (AML)	1	4	0
7	急性骨髄性白血病 (AML)	1	1	0
8	急性骨髄性白血病 (AML)	1	1	1
9	急性リンパ性白血病 (ALL)	2	1	0

drop new_disease

Stata/IC 12.1 - [Results]

dx	diagnosis	disease_st's	seromis6ab~2
3	急性リンパ性白血病 (ALL)	2	1
4	急性リンパ性白血病 (ALL)	2	1
5	急性リンパ性白血病 (ALL)	2	4
6	慢性骨髄性白血病 (CML)	4	4
7	急性リンパ性白血病 (ALL)	2	2
8	急性リンパ性白血病 (ALL)	2	4
9	急性リンパ性白血病 (ALL)	2	4
10	急性リンパ性白血病 (ALL)	2	1
11	急性リンパ性白血病 (ALL)	2	4

keep if diagnosis==4

資料28行目: CMLのみをデータ内に残す

keep if diagnosis==4
(30 observations deleted)

clear

資料31行目:準備運動はここまで。

変数作成の練習
資料33行目～

変数作成の練習: 資料36行目～

- 配布資料 Minidataset_2013.dta (Stata dataset) を取り込んでください。
- クリックで開くも可ですが、今まで使用とは別の Stataプロジェクトが開きます。

データセットをData Editorで確認

データセットをData Editorで確認

変数に対するLabelと各項目に対する Value Label

- 137 -

```
. use "C:\%data%\Minidata_2013.dta", clear
file C:\%data%\Minidata_2013.dta not found
r(601);
use "C:\%data%\Minidataset_2013.dta", clear
```

S1\%Minidata_2013 49行目～

```
***不要データの削除***
*****
* STEP
* 1) "diagnosis"がAML, ALL, CMLの中からCMLを削除
* 2) survival dataが欠損の症例を削除
* last_date (最終観察日)が欠損の症例を削除
*****
```

drop, keepを使用。

順番に疾患データの中身を確認

57. tab dx, missing (欠損値も表に含める)

```
. tab dx, missing
```

主疾患名	Freq.	Percent	Cum.
急性リンパ性白血病(ALL)	30	52.63	52.63
急性骨髄性白血病(AML)	24	42.11	94.74
慢性骨髄性白血病(CML)	3	5.26	100.00
Total	57	100.00	

順番に疾患データの中身を確認

58. tab diagnosis, missing (欠損値も表に含める)

```
tab diagnosis, missing
```

疾患名1	Freq.	Percent	Cum.
AML	24	42.11	42.11
ALL	30	52.63	94.74
CML	3	5.26	100.00
Total	57	100.00	

Labelが付いた状態

順番に疾患データの中身を確認
59. tab diagnosis, missing nolabel (ラベルを付けない)

tab diagnosis, missing

疾患名1	Freq.	Percent	
AML	24	42.11	4
ALL	30	52.63	9
CML	3	5.26	10
Total	57	100.00	

tab diagnosis, missing nolabel

疾患名1	Freq.	Percent	Cum.
1	24	42.11	42.11
2	30	52.63	94.74
4	3	5.26	100.00
Total	57	100.00	

S1¥Minidata_2013 62行目~

/*AMLとALLのみをkeep*/

* if以下の"等しい"は、"=="

* 文字列の値は、"" で囲む。数値は囲まない

* & (and かつ),

| (or または)

== (equal 等しい)

!= (not equal 等しくない)

66. keep if diagnosis==1 | diagnosis==2

67. drop if diagnosis==4

どちらかを使用して、CML症例をデータから削除

68. count

71. tab survival, missing

(3 observations deleted)

```
count
54
```

```
tab survival, missing
```

生存状況	Freq.	Percent	Cum.
死亡	2	3.70	3.70
生存	25	46.30	50.00
Total	54	100.00	

2x2 tableを作る

61. tab dx diagnosis, missing

66. keep if diagnosis==1 | diagnosis==2

68. count

dx	diagnosis	disease_st's	seromis6ab*2
急性リンパ性白血病(ALL)	ALL	CR1	0
急性骨髄性白血病(AML)	AML	HR	2
慢性骨髄性白血病(CML)	ALL	HR	1
急性リンパ性白血病(ALL)	ALL	CR1	0
急性骨髄性白血病(AML)	AML	CR2	0
慢性骨髄性白血病(CML)	AML	HR	0
急性リンパ性白血病(ALL)	AML	CR1	0
慢性骨髄性白血病(CML)	AML	CR1	1

```
. keep if diagnosis==1 | diagnosis==2
(3 observations deleted)

. count
54
```

72. tab last_date, missing

```
tab last_date, missing
```

生死最終確認日	Freq.	Percent	Cum.
2001/2/13	4	7.41	7.41
2001/8/18	1	1.85	9.26
2001/9/24	1	1.85	11.11
2002/11/3	1	1.85	12.96
2002/2/21	1	1.85	14.81
2002/8/2	1	1.85	16.67
2002/3/9	1	1.85	18.52
2004/10/17	1	1.85	20.37
2004/11/1	1	1.85	22.22
2004/3/18	1	1.85	24.07
	1	1.85	25.93

スペースキーなどで下を見られます。

2x2 tableを作る

61. tab dx diagnosis, missing

```
tab dx diagnosis, missing
```

主疾患名	疾患名1			Total
	AML	ALL	CML	
急性リンパ性白血病(ALL)	0	30	0	30
急性骨髄性白血病(AML)	24	0	0	24
慢性骨髄性白血病(CML)	0	0	3	3
Total	24	30	3	57

S1¥Minidata_2013 70行目~

survival (生存状況)のデータと、last_date(最終確認日)が欠損の症例を削除

まずは、データの確認

71. tab survival, missing

72. tab last_date, missing

次に、73か74 どちらかで欠損症例を削除

73. drop if survival=="" | last_date==""

74. keep if survival!="" | last_date!=""

75. count

73. drop if survival=="" | last_date==""

75. count

```
drop if survival=="" | last_date==""
(4 observations deleted)

. count
50
```

S1¥Minidata_2013 77行目~

/**** dataの加工 ****/

* 性別データの作成

*「男」「女」のdataでは後に多変量解析ができないため、
数値(0, 1, 2,)の変数を作成する必要があります

STEP

- 1) sex の「男」「女」というデータをもとに、
女性が0 男性が1の新たな変数"pt_sex"を作成
Generate, replaceを使います
- 2) それに"Female" "Male"のラベルを付けます
- 3) 作成変数の確認

変数 "pt_sex" のデータを、
sex=="男"の症例は「1」に変更する

92. replace pt_sex = 1 if sex=="男"

diagnosis	disease_st's	seromis6ab'2	pt_sex
ALL	CR1	0	1
AML	HR	2	0
ALL	HR	1	0
ALL	CR1	0	0
AML	CR2	0	1
AML	HR	0	1
AML	CR1	0	1

性別データの確認

89. tab sex, missing

性	Freq.	Percent	Cum.
女	22	44.00	44.00
男	28	56.00	100.00
Total	50	100.00	

新たな変数 "pt_sex" を作成、

sex=="女"の症例は「0」

90. gen pt_sex = 0 if sex=="女"

diagnosis	disease_st's	seromis6ab'2	pt_sex
1	ALL	CR1	0
2	AML	HR	2
3	ALL	HR	0
4	ALL	CR1	0
5	AML	CR2	0
6	AML	HR	0

「sex_label」という名前のラベルの定義をする

99. label define sex_label 0 "Female" 1 "Male"

diagnosis	disease_st's	seromis6ab'2	pt_sex
ALL	CR1	0	1
AML	HR	2	0
ALL	HR	1	0
ALL	CR1	0	0
AML	CR2	0	1
AML	HR	0	1
AML	CR1	0	1
AML	CR1	1	1

変数 "pt_sex" にラベル「sex_label」を
関連付ける(名札をつける作業)

100. label values pt_sex sex_label

diagnosis	disease_st's	seromis6ab'2	pt_sex
ALL	CR1	0	Male
AML	HR	2	Female
ALL	HR	1	Female
ALL	CR1	0	Female
AML	CR2	0	Male
AML	HR	0	Male
AML	CR1	0	Male
AML	CR1	1	Male

- 139 -

2x2 tableを作って、間違えていないか確認

107. tab sex pt_sex, missing

	男	0	28	28
Total		22	28	50

性	pt_sex		Total
	Female	Male	
女	22	0	22
男	0	28	28
Total	22	28	50

S1¥Minidata_2013 109行目~ ミスマッチ変数の作成

Characteristic	N (%)	N (%)	P-value
Total			
Age at transplant, years, median (range)			
Gender			
Male	(%)	(%)	
Female	(%)	(%)	
Disease			
AML	(%)	(%)	
ALL	(%)	(%)	
Year of transplant			
2006-2009	(%)	(%)	
2004-2010	(%)	(%)	

S1¥Minidata_2013 109行目~

tab seromis6abdr2, missing 117行目

r2	Freq.	Percent	Cum.
0	29	58.00	58.00
1	17	34.00	92.00
2	4	8.00	100.00
Total	50	100.00	

HLA ミスマッチあり1, なし0の変数を作成

- 1) seromis6abdr2 でミスマッチ数 0個を "mismatch"=0
seromis6abdr2 でミスマッチ数 1個以上を "mismatch"=1
- 2) 0 "match" 1 "mismatch"のラベルを付けます
- 3) 作成した変数の確認

“mismatch”の作成: ミスマッチなし 0, ミスマッチあり 1

118. generate mismatch=0 if seromis6abdr2==0

119. replace mismatch=1 if

seromis6abdr2==1 | seromis6abdr2==2

disease_st's	seromis6abdr2	pt_sex	mismatch	
ALL	CR1	0	Male	0
AML	NR	2	Female	1
ALL	NR	1	Female	1
ALL	CR1	0	Female	0
AML	CR2	0	Male	0

```
generate mismatch=0 if seromis6abdr2==0
(21 missing values generated)

replace mismatch=1 if seromis6abdr2==1 | seromis6abdr2==2
(21 real changes made)
```

S1¥Minidata_2013 146行目~

last_date	tx_date
2001/8/18	2001/1/28
2011/7/11	2002/8/9
2007/7/21	2007/3/20
2011/8/2	2009/8/19
2011/8/1	2010/10/19
2007/5/7	2006/5/29
2005/12/28	2004/6/23

Excel 1900年1月1日がシリアル値「1」
2013年10月20日はシリアル値「41567」
Stata 1960年1月1日がシリアル値「1」
2013年10月20日はシリアル値「19651」

* 1960/1/1を「1」とした5桁の変数「tdate」を作成
* 10/20/2013と入力されたデータの場合は“MDY”と指定

移植年 categorical dataの作成

・ 1: 2000-2005年、2: 2006-2010年 変数 “tyear_cat”
・ ラベルをつける 1 "2000-2005" 2 "2006-2010”

・ 変数 “tyear_cat”にも 変数label(表札)をつける

162 tab tyear

165 gen tyear_cat = 1 if tyear>=2000 & tyear<=2005
あるいは

166 gen tyear_cat = 1 if tyear==2001 | tyear==2002 |
tyear==2003 | tyear==2004 | tyear==2005

167 replace tyear_cat = 2 if tyear>=2006 & tyear<=2010

変数“mismatch” にラベル「 mismatch_label」をつける
0, match; 1, mismatch

124 label define mismatch_label 0 "match" 1 "mismatch"

125 label values mismatch mismatch_label

disease_st's	seromis6abdr2	pt_sex	mismatch	
ALL	CR1	0	Male	match
AML	NR	2	Female	mismatch
ALL	NR	1	Female	mismatch
ALL	CR1	0	Female	match
AML	CR2	0	Male	match

```
label define mismatch_label 0 "match" 1 "mismatch"

label values mismatch mismatch_label
```

移植日付“tx_date”から、stata日付データ“tdate”を作成

159 generate tdate= date(tx_date, "YMD")

disease_st's	seromis6abdr2	pt_sex	mismatch	tdate
CR1	0	Male	match	15001
NR	2	Female	mismatch	15561
NR	1	Female	mismatch	17245
CR1	0	Female	match	18128
CR2	0	Male	match	18548
NR	0	Male	match	16950
CR1	0	Male	match	16212
CR1	1	Male	mismatch	16357

```
generate tdate= date(tx_date, "YMD")
```

165 gen tyear_cat = 1 if tyear>=2000 & tyear<=2005

167 replace tyear_cat = 2 if tyear>=2006 & tyear<=2010

mismatch	tdate	tyear	tyear_cat
match	15001	2001	1
mismatch	15561	2002	1
mismatch	17245	2007	2
match	18128	2009	2
match	18548	2010	2
match	16950	2006	2
match	16212	2004	1
mismatch	16357	2004	1

```
gen tyear_cat = 1 if tyear>=2000 & tyear<=2005
(25 missing values generated)

replace tyear_cat = 2 if tyear>=2006 & tyear<=2010
(25 real changes made)
```

2x2 tableで間違えないか確認

128. tab mismatch seromis6abdr2, missing nolabel

tab mismatch seromis6abdr2, missing nolabel

mismatch	seromis6abdr2			Total
	0	1	2	
0	29	0	0	29
1	0	17	4	21
Total	29	17	4	50

今作った“tdate”から、移植年変数“tyear”を作成

160 gen tyear=year(tdate)

seromis6abdr2	pt_sex	mismatch	tdate	tyear
0	Male	match	15001	2001
2	Female	mismatch	15561	2002
1	Female	mismatch	17245	2007
0	Female	match	18128	2009
0	Male	match	18548	2010
0	Male	match	16950	2006
0	Male	match	16212	2004
1	Male	mismatch	16357	2004

```
gen tyear=year(tdate)
```

169 label var tyear_cat “移植年category”

170 label define tyear_cat_label 1 "2000-2005" 2 "2006-2010"

171 label values tyear_cat tyear_cat_label

173 tab tyear_cat tyear, missing

移植年category	2000	2001	2002	2003	Total
2000-2005	3	5	5	5	25
2006-2010	0	0	0	0	25
Total	3	5	5	5	50

移植年category	2004	2005	2006	2007	Total
2000-2005	4	3	0	0	25
2006-2010	0	0	6	5	25
Total	4	3	6	5	50

移植年category	2008	2009	2010	Total
2000-2005	0	0	0	25
2006-2010	3	6	5	25

Table 1. の作成

Table 1. Characteristics of study patients

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

177 tab mismatch, missing

```
tab mismatch, missing
```

mismatch	Freq.	Percent	Cum.
match	29	58.00	58.00
mismatch	21	42.00	100.00
Total	50	100.00	

178 sum age if mismatch==0, detail

```
sum age if mismatch==0, detail
```

Percentiles		Smallest		
1%	1	1	Obs	29
5%	1		Sum of Wgt.	29
10%	2			
25%	4			
50%	10		Mean	11
		Largest	Std. Dev.	7.559289
		21		
75%	16	24	Variance	57.14286
90%	24	25	Skewness	.4397324
95%	25	26	Kurtosis	2.105173
99%	26			

179 sum age if mismatch==1, detail

```
sum age if mismatch==1, detail
```

Percentiles		Smallest		
1%	1	1	Obs	21
5%	3	3	Sum of Wgt.	21
10%	3			
25%	6			
50%	17		Mean	14.14286
		Largest	Std. Dev.	9.172163
		27		
75%	21	27	Variance	84.12857
90%	27	27	Skewness	.0578875
95%	27	27	Kurtosis	1.573266
99%	27	27		

Wilcoxon Rank Sum Test

180 ranksum age, by (mismatch)

```
ranksum age, by (mismatch)
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

mismatch	obs	rank sum	expected
match	29	675	739.5
mismatch	21	600	535.5
combined	50	1275	1275

unadjusted variance 2588.25
adjustment for ties -11.81
adjusted variance 2576.44

Ho: age (mismatch==match) = age (mismatch==mismatch)
z = -1.271
Prob > |z| = 0.2038

182 tab pt_sex mismatch, chi2 colum

```
tab pt_sex mismatch, chi2 colum
```

pt_sex	mismatch		Total
	match	mismatch	
Female	13 44.83	9 42.86	22 44.00
Male	16 55.17	12 57.14	28 56.00
Total	29 100.00	21 100.00	50 100.00

Pearson chi2 (1) = 0.0192 Pr = 0.890

183 tab diagnosis mismatch, chi2 colum

185 tab tyear_cat mismatch, chi2 colum

```
tab diagnosis mismatch, chi2 colum
```

疾患名	mismatch		Total
	match	mismatch	
AML	12 41.38	10 47.62	22 44.00
ALL	17 58.62	11 52.38	28 56.00
Total	29 100.00	21 100.00	50 100.00

Pearson chi2 (1) = 0.1925 Pr = 0.66

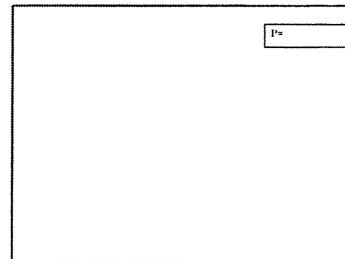
```
tab tyear_cat mismatch, chi2 colum
```

移植年 category	mismatch		Total
	match	mismatch	
2000-2005	14 48.28	11 52.38	25 50.00
2006-2010	15 51.72	10 47.62	25 50.00
Total	29 100.00	21 100.00	50 100.00

Pearson chi2 (1) = 0.0821 Pr = 0.774

Figure 1. の作成

Figure 1. Kaplan-Meier Survival Curve



S1¥Minidata_2013 187行目~

**** Survival dataの作成 ****

last_date	tx_date	Variable
2001/8/18	2001/1/28	Survival
2011/7/11	2002/8/9	Survival
2007/7/21	2007/3/20	Survival
2011/8/2	2009/8/19	Survival

日付データの取り扱い

- 1) 生死最終確認日 "last_date" から stata日付変数 "ldate" の作成
- 2) "ldate" と "tdate" から 移植日から最終観察日までの日数 "lday" を作成

188 generate ldate=date(last_date, "YMD")
Stata日付データ、5ケタの数字

189 generate lday=ldate-tdate

190 sum lday, detail
生存日が移植日より前のデータなどないか？

```

. generate ldate=date(last_date, 'YMD')
. generate lday=ldate-tdate
. sum lday, detail

```

Percentiles		Smallest		Obs	Sum of Wgt.	Mean	Std. Dev.	Variance	Skewness	Kurtosis
1%	10	10	10							
5%	33	19	19							
10%	89.5	33	33							
25%	247	47	47							
50%	554.5									
75%	1579	3217	3217							
90%	2621	3258	3258							
95%	3258	3380	3380							
99%	3847	3847	3847							

生死データの作成

193 tab survival, missing nolabel

```

. tab survival, missing nolabel

```

生存状況	Freq.	Percent	Cum.
死亡	24	48.00	48.00
生存	26	52.00	100.00
Total	50	100.00	

```

. tab survival, missing nolabel

```

survival	Freq.
1 死亡	24
2 生存	26
3	

変数作成 "event_os" 生存:0 死亡:1
 194 gen event_os=1 if survival=="死亡"
 195 replace event_os=0 if survival=="生存"

```

. tab event_os, missing

```

event_os	Freq.
0	26
1	24
Total	50

196 *tab survival event_os, missing

最終観察日, failure の宣言
 198 stset lday, failure (event_os==1)

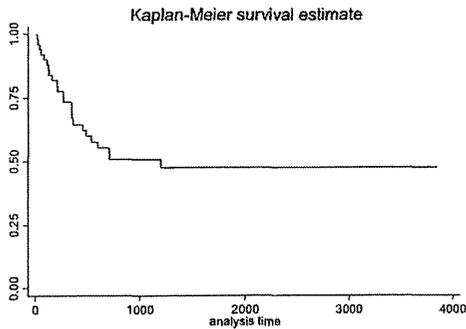
```

. stset lday, failure (event_os==1)
. failure event: event_os == 1
. obs. time interval: (0, lday]
. exit on or before: failure

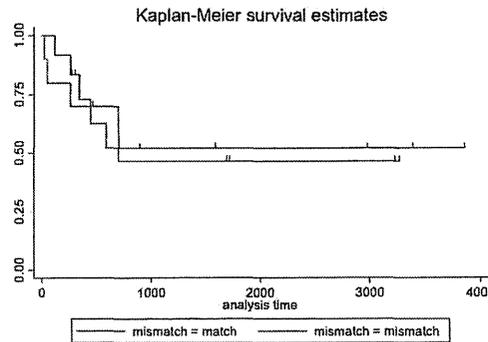
```

50	total obs.	
0	exclusions	
50	obs. remaining, representing	
24	failures in single record/single failure data	
48073	total analysis time at risk, at risk from t =	0
	earliest observed entry t =	0
	last observed exit t =	3847

199 sts graph



200 sts graph if diagnosis==1, by (mismatch) censored(multiple)



Log rank 検定

208 sts test mismatch
 . sts test mismatch

```

. failure _d: event_os == 1
. analysis time _t: lday

```

Log-rank test for equality of survivor functions

mismatch	Events observed	Events expected
match	14	14.43
mismatch	10	9.57
Total	24	24.00

chi2(1) = 0.03
 Pr>chi2 = 0.8582

生存解析の結果の表示 (点推定値)

211 sts list

```

. sts list
. failure _d: event_os == 1
. analysis time _t: lday

```

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf. Int.]
10	50	1	0	0.9800	0.0198	0.8664 0.9972
19	49	1	0	0.9600	0.0277	0.8494 0.9898
33	48	1	0	0.9400	0.0336	0.8254 0.9802
47	47	1	0	0.9200	0.0384	0.8007 0.9692
74	46	1	0	0.9000	0.0424	0.7763 0.9571
105	45	0	1	0.9000	0.0424	0.7763 0.9571
106	44	1	0	0.8795	0.0461	0.7513 0.9440
116	43	1	0	0.8591	0.0494	0.7269 0.9302
123	42	1	0	0.8386	0.0523	0.7030 0.9159
153	41	1	0	0.8182	0.0549	0.6796 0.9010
200	40	1	0	0.7977	0.0572	0.6566 0.8857
204	39	1	0	0.7773	0.0593	0.6339 0.8700

生存解析の結果の表示 (点推定値)

213 (改変) sts list, by (mismatch) at (365, 730, 1095)

```

. sts list, by (mismatch) at (365, 730, 1095)
. failure _d: event_os == 1
. analysis time _t: lday

```

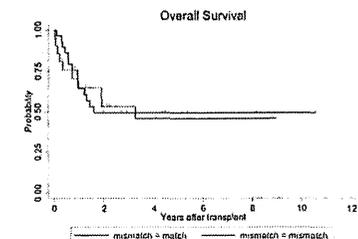
Time	Beg. Total	Fail	Survivor Function	Std. Error	[95% Conf. Int.]
match					
365	18	10	0.6448	0.0907	0.4393 0.7913
730	13	4	0.4931	0.0960	0.2974 0.6622
1095	10	0	0.4931	0.0960	0.2974 0.6622
mismatch					
365	13	7	0.6505	0.1074	0.4014 0.8166
730	10	2	0.5322	0.1160	0.2890 0.7258
1095	8	0	0.5322	0.1160	0.2890 0.7258

Note: survivor function is calculated over full data and evaluated at indicated times; it is not calculated from aggregates shown at left.

203 generate lyear=lday/365.25

204 stset lyear, failure (event_os==1)

205 sts graph, by (mismatch) title (Overall Survival) censored (multiple) ytitle (Probability) xtitle (Years after transplant) xlabel (0 (2) 12)



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

