

```

151 * 3) 移植年カテゴリーの作成
152 *****
153 *Excel 1900年1月1日がシリアル値「1」
154 *Stata 1960年1月1日がシリアル値「1」
155 *Excel
    をStataにデータを取り込んだ場合には、自動的にStata日付データに変換されるため、下記のコマンドは不要になります。
156 ** 1960/1/1を「1」とした5桁の変数「tdata」を作成します。
157 ** 10/20/2013と入力されたデータの場合は"MDY"と指定する必要があります。
158
159 generate tdate= date(tx_date, "YMD")
160 gen tyear=year(tdate)
161 *5桁のStata日付データから「年」が抽出され、新たな変数tyearを作成
162 tab tyear
163
164 **3) 移植年のcategorical dataを作る [宿題]**
165 gen tyear_cat = 1 if tyear>=2000 & tyear<=2005
166 *あるいは 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
168
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
172
173 tab tyear_cat tyear, missing
174 *tab tyear_cat tyear, missing nolabel
175
176 /**** 背景表(Table 1)の作成 *****/
177 tab mismatch, missing
178 sum age if mismatch==0, detail
179 sum age if mismatch==1, detail
180 ranksum age, by (mismatch) /*Wilcoxon Rank Sum Test*/
181
182 tab pt_sex mismatch, chi2 colum
183 tab diagnosis mismatch, chi2 colum
184 ** tab agecat mismatch, chi2
185 tab tyear_cat mismatch, chi2 colum
186
187 /**** Survival dataの作成 *****/
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
    /*データの確認、生存日が移植日より前のデータなどないか？など。詳細はTRUMPセッションで。*/
191
192 /*生死データ*/
193 tab survival, missing nolabel
194 gen event_os =1 if survival == "死亡"
195 replace event_os =0 if survival == "生存"
196 *tab survival event_os, missing

```

```
101 *Data editorを確認。pt_sexの項目がFemale, Maleとラベルされています
102
103 **** 3) 新しく作成した変数を確認
104 *tab sex, missing
105 tab pt_sex, missing
106 tab pt_sex, missing nolabel
107 tab sex pt_sex, missing
108
109 *****
110 * HLA mismatch あり1, なし0 のdataを作成
111 * STEP
112 * 1) mismatch locus 数:0個を mismatch=0
113 * mismatch locus 数: 1個以上がmismatch=1 の変数を作成
114 * 2) それに"match" "mismatch"のラベルを付けます
115 * 3) 作成した変数の確認 (終了して、余裕のある方は次の年齢のcategorical data作成へ。)
116 *****
117 tab seromis6abdr2, missing
118 generate mismatch=0 if seromis6abdr2==0
119 replace mismatch=1 if seromis6abdr2==1 | seromis6abdr2==2
120
121 ** 欠損値がある場合は、欠損値のカテゴリも作る。
122 ** replace mismatch=99 if seromis6abdr2==.
123
124 label define mismatch_label 0 "match" 1 "mismatch"
125 label values mismatch mismatch_label
126
127 tab mismatch seromis6abdr2, missing
128 tab mismatch seromis6abdr2, missing nolabel
129
130 *****
131 * 年齢のcategorical dataの作成
132 * 0-9歳、10-19歳、20-29歳の変数を作成
133 * ラベルをつける
134 *****
135 * summarize age
136 * sum age, detail
137 * gen agecat=1 if age>=0 & age<10
138 * replace agecat=2 if age>=10 & age<20
139 * replace agecat=3 if age>=20 & age<30
140 *Stataの場合欠損値は"large positive values"として認識。使用する統計ソフトによって設定が異なる。
141 * label define agecat_label 1 "0-9" 2 "10-19" 3 "20-30"
142 * label values agecat agecat_label
143 * tab agecat, missing
144 * tab agecat, missing nolabel
145
146 *****
147 * 日付と日数の計算
148 * STEP
149 * 1) stata日付データの作成
150 * 2) 移植日から移植年の作成
```

```
51 * STEP
52 * 1) AML, ALL, CMLの中からCMLを削除
53 * 2) survival dataが欠損の症例を削除
54 * last_date (最終観察日)が欠損の症例を削除
55 *****
56 **** 1) AML, ALL, CMLの中からCMLを削除 ****
57 tab dx, missing
58 tab diagnosis, missing
59 tab diagnosis, missing nolabel
60 /*2x2 tableを作る*/
61 tab dx diagnosis, missing
62 /*AMLとALLのみをkeep */
63 * if以下の"等しい"は、"=="
64 * 文字列の値は、"" で囲む。数値は囲まない
65 * & (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 でも同じ結果が得られます
68 count
69
70 **** 2) survival dataとlast_dateが欠損の症例を削除 ****
71 tab survival, missing
72 tab last_date, missing
73 drop if survival=="" | last_date==""
74 * keep if survival!="" | last_date!=""
75 count
76
77 /**** dataの加工 ****/
78 *****
79 * 性別dataの作成
80 * 「男」「女」のdataでは後に多変量解析ができないため、数値による変数を作成する必要があります
81 *
82 * STEP
83 * 1) 「男」「女」というデータから、女性が0 男性が1の新たな変数を作り
84 * 2) それに"Female" "Male"のラベルを付けます
85 * 3) 作成した変数の確認
86 *****
87
88 **** 1) sexの「男」「女」というデータから、女性が0 男性が1の新たな変数"pt_sex"を作成
89 tab sex, missing
90 gen pt_sex = 0 if sex=="女"
91 *変数 "pt_sex" を作成。sex=="女" の症例を0とする
92 replace pt_sex = 1 if sex=="男"
93 *Data Editorを確認。新たな変数 "pt_sex"が作成されている
94
95 ** 項目にラベルをつけることができる (公開スクリプト01を参照)
96 ** label var pt_sex "患者性別"
97
98 **** 2) コードの意味を忘れてしまうこともあるが、以下のラベルが便利
99 label define sex_label 0 "Female" 1 "Male"
100 label values pt_sex sex_label
```

```
1  /**** 解析データセットの作成 ****/  
2  
3  *Stataを起動して、ファイルの中にある、minidataset_2013.xls を取り込んでください。  
4  *余裕があれば、log fileを作成してみてください。  
5  *方法は問いません。  
6  
7      *** import excel "C:¥data¥minidataset_2013.xls", firststrow  
8      *** log using "C:¥data¥mini20131020.smcl"  
9      *** あるいは、新たなログファイルを"Log Begin"のタブから作成  
10  
11  //////////****新たな変数作成の練習****/////////  
12  * 準備運動 *  
13  /**generate, replace**/  
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=.  
21  replace new_disease2=1 if new_disease=="AML"  
22  replace new_disease2=2 if new_disease=="ALL"  
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。戻らないので注意!  
27  drop if diagnosis==1  
28  keep if diagnosis==4  
29  
30  *一度Stataを閉じてください。  
31  * あるいは、clearコマンドを使用。データはすべて消えてしまいます。  
32  
33  //////////****変数作成の練習****/////////  
34  *配布資料：ミニデータセット ワークシートの作成*  
35  
36  *新たに配布DVD/USB内の "Minidata_2013"をダブルクリックして開く  
37      *** use "C:¥data¥Minidata_2013.dta", clear のコマンド使用も可能  
38  
39  /****データセットをData Editorで確認 ****/  
40  *DataEditorを使って、データのシートを開いてください。  
41      *Variables, Properties, Data editor (Browse)  
42      *Label, Value Label  
43  /**変数ラベルの概念についての復習：変数「sex」におけるラベル「性」**/  
44  /**項目ラベルの概念について：変数「diagnosis」の中のデータ「1」に対するラベル「AML」**/  
45  
46  /**症例数確認*/  
47  count  
48  
49  /****不要データの削除****/  
50  *****/
```

```
197
198 stset lday, failure (event_os==1)
199 sts graph
200 sts graph if diagnosis==1, by (mismatch) censored(multiple)
201
202 generate lyear=lday/365.25
203 stset lyear, failure (event_os==1)
204 sts graph, by (mismatch) title (Overall Survival) censored (multiple) /*
205 */ ytitle (Probability) xtitle (Years after transplant) xlabel (0 (2) 12)
206
207 ** Log rank 検定 **
208 sts test mismatch
209
210 ** 生存解析の結果の表示 (点推定値) **
211 sts list
212 sts list, by (mismatch)
213 sts list, by (mismatch) at (1, 3, 5)
214
```

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

H25年度 TRUMP統計セミナー
基礎編

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・統計解析の実際

担当: 熱田由子

変数の決定と変数リストの作成

- 変数の種類
 - 患者背景変数
 - 患者年齢・性別
 - 疾患情報
 - 移植関連情報
 - アウトカム変数
 - イベント変数
 - 時間変数
- 研究計画が定まったら、解析に必要な変数リストを作成する

例1: ABO mismatch

データを覗いてみると...

	pt_abo	do_abo	abo_mach
1519.	0	0	適合
1520.			
1521.			
1522.			
1523.	B	B	適合
1524.			
1525.	A		
1526.	0	0	
1527.			
1528.			
1529.	0	0	
1530.			
1531.			
1532.	A	A	適合
1533.	A	A	

変数の作成

エクセル上で作業をしない！！

- 150-200例未満程度の場合は目視・手入力でも出来るかもしれないが..
- ソートを繰り返していたらデータが無茶苦茶..
- 気付いたらどこかで一行ずれてしまっていた..
- 変数作成のアルゴリズムを忘れてしまうことも..

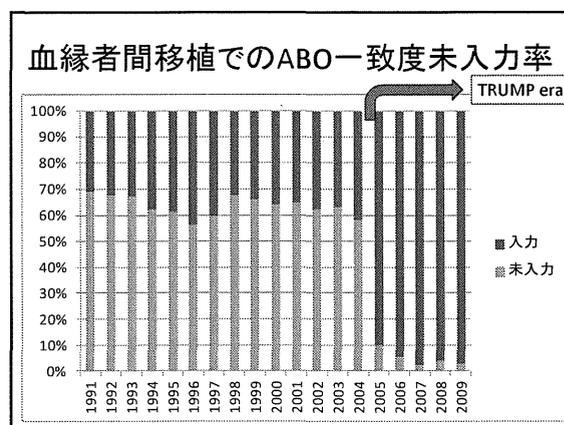
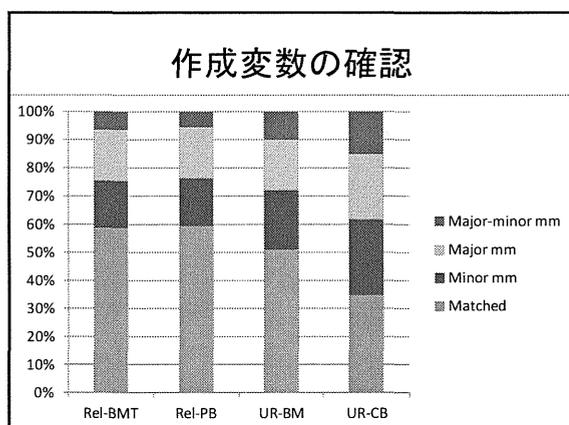
“.....”だらけ

Stata-EZR共通の変数リスト、スクリプトファイルがダウンロード可能

変数一覧(基本変数リスト)
(Stata / EZR共有スクリプト対応)
ダウンロード可能!!

変数作成

- ABO matching
 - Matched
 - Minor mismatch
 - Major mismatch
 - Major-minor mismatch
- ABO major mismatch
 - (0) Matched or minor mismatch
 - (1) Major (major-minor) mismatch



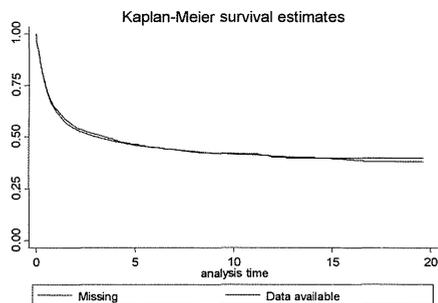
どうしよう?

- 研究課題
 - ①. 「血縁者間移植においてABO一致度が生存に与える影響」
 - ②. 「血縁者間移植におけるBMとPBの比較」

研究①「ABO一致度の影響」の場合

- 2005年以降の移植に限定した解析を行う
 - 観察期間が短い
- 入力があるものを採用する。その場合、2004年以前に関しては対象に入るのは3-4割程度。
 - 施設でのTRUMP以降の自主的入力
 - 未入力の場合と大きな差が(なんらかのバイアスが)ないか確認しておく必要がある。
- 2004年以前移植に関しては、入力率が7割以上の施設のデータのみを対象

1991-2004移植での確認



研究②「RBM vs. RPB」の場合

- “ABO matching”変数の使用目的
 - RBM vs. RPBの比較の際の補正因子
 - (主アウトカムは全生存率)
- 補正に用いる因子の候補
 - 検討するアウトカムに有意に影響することが過去の研究で既に明らかになっているもの
 - 比較群間に偏りがあるもの
- 選択
 - 補正因子として用いない
 - 補正因子として用いる

Missing dataの取り扱い -補正因子として用いる場合-

- 3つの方法
 - 1. 欠損値を解析対象からはずす
 - 2. 欠損値をひとつのカテゴリとする
 - 3. imputation: 似た患者の情報の平均値を入れる
- Klein JP says (Klein et al. BMT 2001):*
- 欠損値が少ないとき $\leq 10\%$ (EBMT guideline, 5%)
 - 1番 ただし、はずした場合とはずさなかった場合で結果が変わっていないかを確認する
- 欠損値が比較的多いとき $> 10\%$
 - 2番・3番 データ欠損患者をその他と比較することも可能

欠損値を一つのカテゴリとする?

	HR	p value	
Stem Cell Source			
EM	1.00	67%	29%
PB	1.14	0.007	
ABO matching			
ABO matched or minor	1.00		
ABO major	0.99	0.923	
ABO missing	1.08	0.145	
Disease stage			
Standard	1.00		
Advanced	2.91	<0.0001	
Patient age			
age<40	1.00	100%	100%
Age>40	1.48	<0.0001	
		3,327例	2,255例

TRUMP 統計解析セミナー（基礎編）出席者名簿

1	新井 康之	京都大学大学院医学研究科
2	青木 淳	駒込病院
3	青木 智広	名古屋第二赤十字病院
4	有馬 靖佳	田附興風会医学研究所北野病院
5	伊豆津 宏二	虎の門病院
6	井上 政弥	島根大学医学部附属病院
7	植木 俊充	長野赤十字病院
8	梅田 雄嗣	京都大学医学部附属病院
9	緒方 正男	大分大学医学部
10	冲中 敬二	国立がん研究センター中央病院
11	籠尾 壽哉	国立病院機構東京医療センター
12	加藤 格	京都大学医学部附属病院
13	加藤 剛二	名古屋第一赤十字病院
14	加藤 実穂	名古屋第一赤十字病院
15	亀井 美智	名古屋市立大学
16	鴨下 園子	名古屋赤十字病院
17	河北 敏郎	国立病院機構熊本医療センター
18	川島 直実	名古屋第一赤十字病院
19	川瀬 有美	京都大学医学部附属病院
20	木田 理子	NTT 東日本関東病院
21	栗田 尚樹	筑波大学
22	近藤 健	北海道大学病院
23	佐々木 裕哉	小倉記念病院
24	住 昌彦	長野赤十字病院
25	杉盛 千春	石川県立中央病院
26	瀬戸 愛花	岡崎市民病院
27	妹尾 紀子	信州大学医学部附属病院
28	高野 久仁子	大分大学医学部附属病院
29	竹田 淳恵	神戸医療センター中央市民病院
30	千原 大	愛知県がんセンター研究所
31	中田 佳世	大阪府立成人病センター
32	長谷川 大一郎	兵庫県立こども病院
33	林 邦雄	明和病院
34	半田 寛	群馬大学医学部附属病院
35	平瀬 敏志	神戸大学医学部附属病院
36	深野 玲司	九州がんセンター
37	藤盛 好啓	兵庫医科大学

38	藤原 弘	愛媛大学病院
39	増成 太郎	中国中央病院
40	町田 真一郎	東海大学医学部附属病院
41	水谷 元紀	愛知医科大学医学部
42	三橋 健次郎	東京女子医科大学
43	宮尾 康太郎	安城厚生病院
44	宮村 能子	大阪大学医学部附属病院
45	森 健	神戸大学医学部附属病院
46	森下 喬允	名古屋大学大学院医学系研究科
47	矢内 友子	兵庫県立こども病院
48	横畠 絵美	名古屋赤十字病院
49	吉満 誠	鹿児島大学病院
50	渡邊 健一郎	京都大学大学研究科

TRUMP 統計セミナー資料集 (応用編)

*変数表は省略する

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)
「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：応用編

日時： 平成 25 年 11 月 23 日（土） 10：00～15：30

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

I. 我流 Stata の使い方：自分なりの方法を確立しよう

(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
准教授 鈴木 律朗)

10:30-12:30

II. 多変量解析に臨む前に知っておくこと

(九州大学大学院医学研究院 予防医学分野 教授 松尾 恵太郎)

12:30-13:30

昼食

13:30-14:30

III. 競合リスクイベントの扱いの解説と演習

(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)

14:30-15：30

IV. 時間依存性変数の扱いの解説と演習

(自治医科大学附属さいたま医療センター 血液科 諫田 淳也)

我流 Stata の使い方：自分なりの方法を確立しよう

2013年11月23日(祝)
名古屋大学大幸キャンパス

我流Stataの使い方： 自分なりの方法を確立しよう

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

飛行機人間 と グライダー人間

- 飛行機人間：
自分で空を飛ぶことができる
自分で問題を設定して、解決して行く
大学教育。
研究活動。
- グライダー人間：
空は飛べるが、助けが必要
自力で飛ぶことができない
他人が作った問題を解くのが上手
高校生(大学受験)まで

外山滋比古「思考の整理学」

TRUMPセミナー応用編にあたって

自分なりの「飛び方」を考えてください。

- 我々の言っていることが、必ずしも正しいとは限りません。
- ネットで調べるなどして、適切な手法か状況判断して下さい。
(自己責任)
- 質問・相談は歓迎しますが、高度な内容になると、回答者によって考え方が異なります。
- 論文執筆になると、その方向性を左右することにもなります。
- 「ドクター・ショッピング」にならないように。
最後は自己責任ということ、忘れずに。

我流 Stata 使用法

Stata が少し使えるようになったら・・・

自分なりのメモ・使い方を、残しましょう。

ノートでもよいです。でも、紛失したり散逸したりするので、コンピューターのファイルがよいと思います。

記憶力に頼るのは、年齢的に危険です。

私は、Excelで自分が使えるようになったコマンドを管理して、日々追加しています。

No.	コマンド	用途	備考
1	検索・編集	edit	
2	ヘルプ	help	
3	変数のリストを管理	var list	
4	変数のリストを管理	var list	
5	変数のリストを管理	var list	
6	変数のリストを管理	var list	
7	変数のリストを管理	var list	
8	変数のリストを管理	var list	
9	変数のリストを管理	var list	
10	変数のリストを管理	var list	
11	変数のリストを管理	var list	
12	変数のリストを管理	var list	
13	変数のリストを管理	var list	
14	変数のリストを管理	var list	

No.	コマンド	用途	備考
1	検索・編集	edit	
2	ヘルプ	help	
3	変数のリストを管理	var list	
4	変数のリストを管理	var list	
5	変数のリストを管理	var list	
6	変数のリストを管理	var list	
7	変数のリストを管理	var list	
8	変数のリストを管理	var list	
9	変数のリストを管理	var list	
10	変数のリストを管理	var list	
11	変数のリストを管理	var list	
12	変数のリストを管理	var list	
13	変数のリストを管理	var list	
14	変数のリストを管理	var list	

こっちの方が、TRUMP セミナー基礎編のテキストより、自分用にカスタマイズして使用することが可能です。

テキストを読んだりして、使った経験のあるコマンドを追加していくと良いと思います。

コマンドが増えれば、それだけ自分のスキルが向上したことが目で見て分かります。


```

. stx graph, by(drug) risktable legend(xing(0) position(2) rows(2))
    
```

Number at risk

drug = 0	20	15	10	5	0	0
drug = 1	20	15	14	9	5	0

By default, row titles are placed on the left of the at-risk table and are right-justified. We can illustrate this by changing the text of the row titles to have an unequal length.

凡例 (legend) を移動させるオプション。

PDFマニュアル
全20冊

- Basic Reference Manual
- Data Management Reference Manual
- Graphics Reference Manual
- Longitudinal-Data-panel-Data Reference Manual
- Math Reference Manual
- Multilevel Mixed-Effects Reference Manual
- Multiple Regression Reference Manual
- Multivariate Statistics Reference Manual
- Power and Sample Size Reference Manual
- Programming Reference Manual
- Structural Equation Modeling Reference Manual
- Survey Data Reference Manual
- Survival Analysis and Epidemiological Issues Reference Manual
- Time-Series Reference Manual
- Treatment Effects Reference Manual
- User's Guide
- The Glossary and Index
- Getting Started with Stata

基本コマンド
グラフィック
生存解析

Ver. 12 (Win), Ver. 13 (Win, Mac)
は、日本語マニュアルあり。

Light Stone

STATA 13 日本語マニュアルダウンロード

日本語マニュアル「Getting Started」

ダウンロード可能なマニュアル

Light Stone社のHPからダウンロード可能(無料)

その他のコマンド

search コマンド

help コマンドでは、コマンド名が分からないと、どうしようもない。

search + キーワードで、helpファイル中の検索ができる。

(例) Mann-Whitney 検定のコマンドを探す

```

. search mann whitney
    
```

```

Search mann whitney
-----
Keywords: mann whitney
Search: (1) official help files, cado, examples, sts, and stb
Search of official help files, cado, examples, sts, and stb

(1)  health  . . . . . Kruskal-Wallis equality-of-independence rank test
(help health)

(2)  ransum  . . . . . Equality tests on unordered data
(help ransum)

(3)  st  . . . . . Survival-time data
(help st)

FAQ  . . . . . What statistical analysis should I use?
      . . . . . Old answers: Fisher's exact test
      . . . . . http://www.afs.los.alos.al.gov/stata/faq/faqstat.htm

St-9-3  st0174  . . . . . Approx. equality of Gini-Bulterton-like Gini-Silhouette test
(help esttest if installed)
      . . . . . B. J. Good and J. Kalder
      . . . . . B. 9127454-185
    
```

set mem コマンド

Stataのパソコンメモリ使用量を指定するコマンド。
Ver. 12以降は自動化されたので、不要。

set more コマンド

画面表示のスクロール停止機能("more"と出て、止まる)を制御するコマンド。

set more off で、"more"が出なくなり最後まで一気に表示。
set more off, perm とすると、毎回立ち上げ時からスクロールは停止する。(permanent)

皆さん、もう自分で飛び立てそうでしょうか？

多変量解析に臨む前に知っておくこと