



PHISEM
Program for Human Integrative Science
and Education of Mind



東京大学
THE UNIVERSITY OF TOKYO



UTIDAHM
東京大学こころの多様性と適応の統合的研究機構
UTokyo Institute for Diversity & Adaptation of Human Mind

Introduction of R

2020.2.6

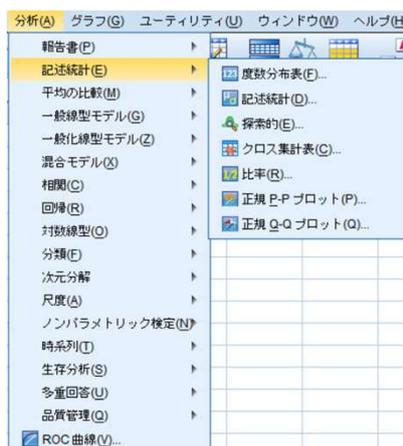
東京大学こころの多様性と適応の統合的研究機構(UTIDAHM)
小池進介

1

Statistical softwares			
	SPSS	R	Excel
配布元	IBM	The R project	Microsoft
価格	Base ¥115,560 Options ¥63,936 / per 1 opt (学生パック ¥10,000-20,000)	Free!	¥15,000-30,000
OS	Win, Mac, Linux	Win, Mac, Linux	Win, Mac
何ができるか?	基本から応用まで (Optionsによる!)	基本から応用まで (Install packages)	基本
利点	<ul style="list-style-type: none"> • GUI base • 心理学データにマッチ • スクリプトが簡単 (GUIから吐き出せる) • 疑問はGoogleで解決 	<ul style="list-style-type: none"> • All free • カスタマイズが用意 • 非常に多くのパッケージ • 応用編の解析も比較的簡単 • 疑問はGoogleで解決 (ただし、難易度が高く、英語の場合もしばしば) 	<ul style="list-style-type: none"> • Datasheet base • 全員使える • 心理学データにマッチ • 疑問はGoogleで解決
欠点	<ul style="list-style-type: none"> • 非常に高価 • 応用編の解析は視覚的にできるとはとも思えない 	<ul style="list-style-type: none"> • CUI base 	<ul style="list-style-type: none"> • 信用されない • 応用解析はない • 大規模データに向かない

2

Example of SPSS



3

Example of R (Rstudio)

```

331 best.models.forcsv <- cbind(ROIlist, best.models)
332 write.table(best.models.forcsv, file = "bestmodels.csv", append = F, row.names=F, sep = ",")
333 #save best sig models
334 colnames(p.vals.sig) = c("ROIlist",ROIlist)
335 write.table(p.vals.sig, file = "bestmodels.pvals.sig.csv", append = F, col.names = F, row.names=T, sep = ",") #t(A): transverse matrix
336 best.sig.models.forcsv <- cbind(ROIlist, best.sig.models)
337 write.table(best.sig.models.forcsv, file = "bestmodels.sig.csv", append = F, row.names=F, sep = ",")
338 setwd("./")
339 x <- x.backup #back up original x
340 #delete data
341 rm(best.models)
342 rm(best.models.forcsv)
343 rm(best.sig.models)
344 rm(best.sig.models.forcsv)
345 rm(model.comparison)
346 rm(p.val)
347 rm(p.vals)
348 rm(p.vals.sig)
349
350 #only patients group
351 x = x[x$diag==0 & !is.na(x$diag),] # only for patients group
352 dir.create("compare.lmerfit.onlypatients")
353 setwd("./compare.lmerfit.onlypatients")
354 for (i in ROIlist){
355   compare.lmerfit(1,"uhr+schizophreniabipolar+sex+age+JART+performance+ #Depression as reference
356     c.uhr:c.age:c.bipolar:c.sex","performance","JART")
357 }
358 #save best models
359 colnames(p.vals) = c("ROIlist",ROIlist)
360 write.table(p.vals, file = "bestmodels.pvals.csv", append = F, col.names = F, row.names=T, sep = ",") #t(A): transverse matrix
361 best.models.forcsv <- cbind(ROIlist, best.models)
362 write.table(best.models.forcsv, file = "bestmodels.csv", append = F, row.names=F, sep = ",")
363 #save best sig models
364 colnames(p.vals.sig) = c("ROIlist",ROIlist)
365 write.table(p.vals.sig, file = "bestmodels.pvals.sig.csv", append = F, col.names = F, row.names=T, sep = ",") #t(A): transverse matrix
366 best.sig.models.forcsv <- cbind(ROIlist, best.sig.models)
367 write.table(best.sig.models.forcsv, file = "bestmodels.sig.csv", append = F, row.names=F, sep = ",")
368 setwd("./")
369 x <- x.backup #back up original x
370 #delete data
371 rm(best.models)
372 rm(best.models.forcsv)
373 rm(best.sig.models)
374 rm(best.sig.models.forcsv)
375 rm(model.comparison)
376 rm(p.val)
377 rm(p.vals)
378 rm(p.vals.sig)
379
380
381
382 ##### Association between clinical variables and fNIRS signals #####
383 #main effect of GAF and diag x GAF interaction. Depression as ref
384 x = x[x$diag==0 & !is.na(x$diag),] # only for patients group
385 dir.create("compare.lmerfit.gaf")
386 setwd("./compare.lmerfit.gaf")

```

4

FIRST TRY RSTUDIO!

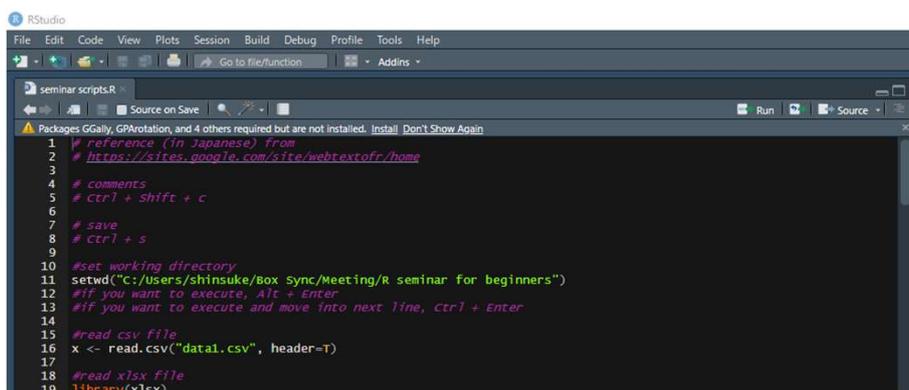
5

RStudioの起動

- **./Rintro/seminar script.R をクリック**
- **RStudioを起動後、ファイルを選択**

6

RStudioの起動



```

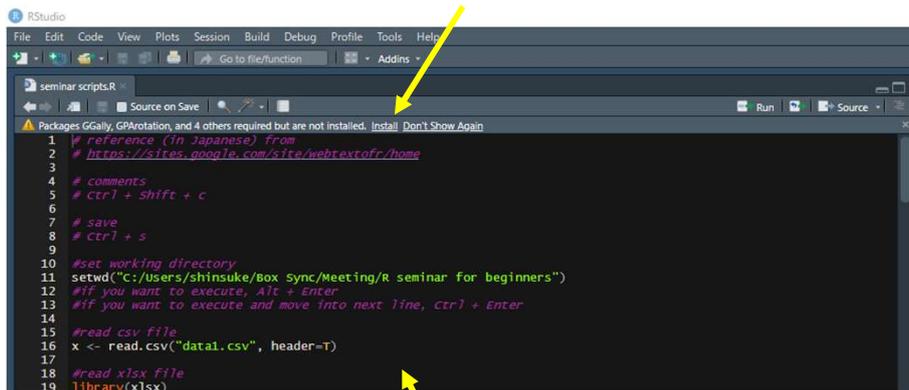
1 # reference (in Japanese) from
2 # https://sites.google.com/site/webtextofr/home
3
4 # comments
5 # Ctrl + Shift + c
6
7 # save
8 # Ctrl + s
9
10 #set working directory
11 setwd("C:/Users/shinsuke/Box Sync/Meeting/R seminar for beginners")
12 #if you want to execute, Alt + Enter
13 #if you want to execute and move into next line, Ctrl + Enter
14
15 #read csv file
16 x <- read.csv("data1.csv", header=T)
17
18 #read xlsx file
19 library(xlsx)

```

7

RStudioの起動

まだインストールされていないlibraryが表示されます
ここでクリックしたらインストールされます



```

1 # reference (in Japanese) from
2 # https://sites.google.com/site/webtextofr/home
3
4 # comments
5 # Ctrl + Shift + c
6
7 # save
8 # Ctrl + s
9
10 #set working directory
11 setwd("C:/Users/shinsuke/Box Sync/Meeting/R seminar for beginners")
12 #if you want to execute, Alt + Enter
13 #if you want to execute and move into next line, Ctrl + Enter
14
15 #read csv file
16 x <- read.csv("data1.csv", header=T)
17
18 #read xlsx file
19 library(xlsx)

```

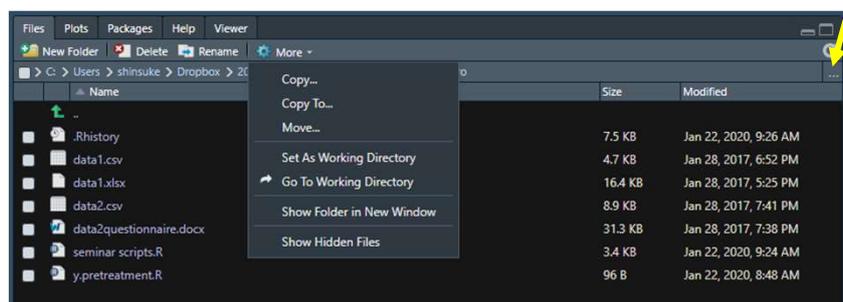
色やフォントが違います
Tools→Global Options→Appearance

8

Working directoryの変更

- setwd (“directory path”)
- 右下の枠内にある設定で
– More→Set As Working Directory

どこを表示しているか
見当たらない方は
ここをクリック



9

R (RStudio) の基本ルール1

- 大文字と小文字は異なります
– view (x) or View (x)
- 代入は <- or =
- 改行で実行
- ()は閉じない限り実行されない
– 視覚的には横に長くなりすぎるなら改行
- スペースは空けても空けなくても良い

```
> 1 * (2 +
+ 3
+ )
[1] 5
>
```

10

R (RStudio) の基本ルール2

- #以降はコメントです(実行されません)
 - プログラムの右側にコメントを書くことも可能

- 該当部分をコメントにしたい場合

```
4 # comments
5 # Ctrl + Shift + c
6
```

Macは Command + Shift + c です

- 該当箇所を実行したい場合(2種類)

```
10 #set working directory
11 setwd("C:/Users/shinsuke/Desktop/rintro")
12 #if you want to execute, Alt + Enter
13 #if you want to execute and move into next line, Ctrl + Enter
14
```

Macは Option + Return, Command + Return です

11

データファイルの読み込み

```
14
15 #read csv file
16 x <- read.csv("data1.csv", header=T)
17
```

- File -> Import Dataset -> From text (base)

The screenshot shows the 'Import Dataset' dialog box in RStudio. The 'Name' field is set to 'x'. The 'Input File' field contains a long path: 'C:/Users/shinsuke/Desktop/R_seminar_for_beginners/data1.csv'. The 'Encoding' is set to 'Automatic', 'Header' is checked, and 'Raw names' is set to 'Automatic'. The 'Data Frame' preview shows a table with columns 'id', 'sex', 'vars1', 'vars2', 'vars3', 'vars4', 'vars5', 'vars6', 'vars7', and 'vars8'. The terminal window shows the command 'x <- read.csv(\"C:/Users/shinsuke/Desktop/R_seminar_for_beginners/data1.csv\")' and the output 'View(x)'.

12

data1.csv

- **id**
 - 50 participants
- **sex**
 - 25:25
- **var1~var30**
 - Randomly input 1~100

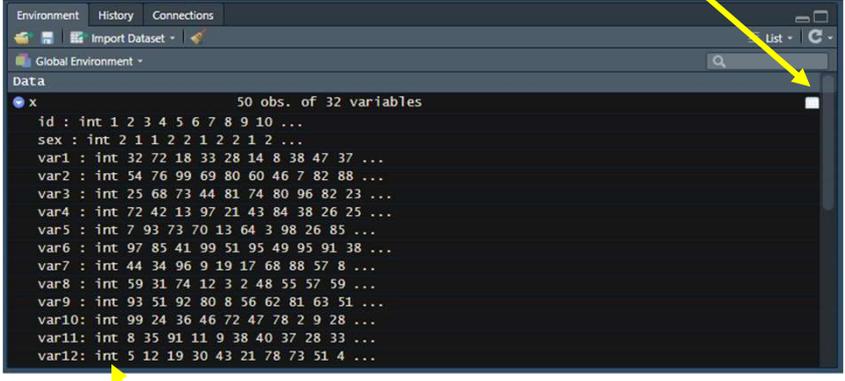
15

簡便なデータチェックの方法

- **summary (x)**
- **head (x)**
 - head (x, n=10)

16

View(x)
(最初は大文字)



```

Environment  History  Connections
Global Environment
Data
x
50 obs. of 32 variables
id : int 1 2 3 4 5 6 7 8 9 10 ...
sex : int 2 1 1 2 2 1 2 2 1 2 ...
var1 : int 32 72 18 33 28 14 8 38 47 37 ...
var2 : int 54 76 99 69 80 60 46 7 82 88 ...
var3 : int 25 68 73 44 81 74 80 96 82 23 ...
var4 : int 72 42 13 97 21 43 84 38 26 25 ...
var5 : int 7 93 73 70 13 64 3 98 26 85 ...
var6 : int 97 85 41 99 51 95 49 95 91 38 ...
var7 : int 44 34 96 9 19 17 68 88 57 8 ...
var8 : int 59 31 74 12 3 2 48 55 57 59 ...
var9 : int 93 51 92 80 8 56 62 81 63 51 ...
var10 : int 99 24 36 46 72 47 78 2 9 28 ...
var11 : int 8 35 91 11 9 38 40 37 28 33 ...
var12 : int 5 12 19 30 43 21 78 73 51 4 ...

```

変数タイプ (後述)

17

Q1

- 男女差があるか検討してみましょう

`t.test (var1 ~ sex, var.equal = T)`

- var1~var30 のどの変数が男女差があるか検討してみましょう

18

R (RStudio) の基本ルール3

```
> #t.test
> t.test(var1~sex)
eval(predvars, data, env) でエラー: オブジェクト 'var1' がありません
> |
```

- 定義されたオブジェクトは”x”であって、”sex”等ではない
- 選択するためには、
 - x\$sex 等と\$でつなぐ
t.test(x\$var1~x\$sex)
 - 行列番号を指定する
t.test(x[,3]~x[,2]) →
 - attach(x) を実行する
attach(x)
t.test(var1~sex)

行列の指定方法
 x[,1]: 1行目のベクトル
 x[,2]: 2列目のベクトル
 x[1,2]: 1行目2列目のスカラー

19

sexを独立変数、var1を従属変数としたt検定(等分散仮定せず)

```
> # attach variables in x
> attach(x)
> #t.test
> t.test(var1~sex)

welch Two Sample t-test

data: var1 by sex
t = 1.1197, df = 45.779, p-value = 0.2687
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-6.511244 22.831244
sample estimates:
mean in group 1 mean in group 2
    54.56         46.40
> |
```

attachしたオブジェクトは、detach()で解除されます

20

for 構文

- 先に "t.test.sex ()" 関数を定義して繰り返す

```

33
34 # for syntax
35 # define t.test.sex function
36 t.test.sex <- function(i){
37   print(t.test(x[,i] ~ x$sex, var.equal = T))
38 }
39
40 # repeat testing
41 for (i in 3:32) t.test.sex(i)
42

```

t.test.sex(3) を実行すれば
print(t.test(x[,3] ~ ...
が実行される関数を作成

t.test.sex(i) を i = 3~32 で実行

- 文字列の結合(paste関数)と文字列の実行(eval(parse)関数)を組み合わせる

```

43 # for syntax and eval(parse(paste(text=))) command set
44 for (i in 1:30){
45   eval(parse(text=paste("print(t.test(x$var", i, " ~ x$sex, var.equal = T))", sep = "")))
46 }
47

```

eval(parse(text=))関数セット
文字列を実行関数とすることができる

paste関数: 文字列の結合、例えば i = 1 なら、
print(t.test(x\$var と 1 と ~x\$sex, var.equal = T)) という 3 つの文字列を、
間に何も挟まず (sep = "") 結合する

なお、Rの for 構文は癖があるようで、ほかのプログラミング言語とは少し違うようです

21

見づらい、記録に残せない

```

Console Terminal Jobs
C:\Users\shinsuke\Desktop\R seminar for beginners/
> t.test(x[,3] ~ x$sex, var.equal = T)
-23.63317 10.59317
sample estimates:
mean in group 1 mean in group 2
43.72 50.24

Two Sample t-test

data: x[, i] by x$sex
t = 0.01528, df = 48, p-value = 0.9908
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-14.03278 13.87278
sample estimates:
mean in group 1 mean in group 2
50.16 50.24

Two Sample t-test

data: x[, i] by x$sex
t = -0.46666, df = 48, p-value = 0.6429
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-19.53565 12.47565
sample estimates:
mean in group 1 mean in group 2
44.24 47.92
>

```

22

R (RStudio) の基本ルール4

- sink関数: 出力結果の保存先を指定
- 49行目を実行後、再びfor構文を走らせる
- 解除は sink ()

```
47
48 # save results to a file
49 sink(file="results_ttest_data1.txt")
50 sink()
51
```

- 出力されたファイルを見て、var1~var30 のどの変数が男女差があるか検討してみましょう

23

Multiple testing

- もし、独立した変数を用いて繰り返し検定したとき、5%有意水準を満たす変数が発生する確率は以下の通りになります
 - 2変数: $1 - (1 - 0.05)^2 = 0.0975$
 - 3変数: $1 - (1 - 0.05)^3 = 0.143$
 - ...
 - 20変数: $1 - (1 - 0.05)^{20} = 0.642$
- そのため、Bonferroni, FWER, FDR等の方法で有意水準を補正したほうが良いでしょう

24





SPSSと対応させたRによるデータ解析チュートリアル

2020.2.6

東京大学こころの多様性と適応の統合的研究機構(UTIDAHM)
小池進介

25

data2.csv

- どのような質問紙かは、data2questionnaire.docx を参照してください
- 230 participants
- Demographics
 - age
 - sex
 - iq
- Longitudinal data (0, 1, and 2 years later)
 - WHO5: wellbeing
 - GHQ12: depressive symptoms
 - SSQ (A, B): social support (number, satisfaction)
- Some missing values
 - blank: No response
 - -1: Missing values

26

欠損値が2種類ある

```
> head(y,10)
  id age sex iq who5_0 ghq12_0 ssqa_0 ssqb_0 who5_1 ghq12_1 ssqa_1 ssqb_1 who5_2 ghq12_2 ssqa_2 ssqb_2
1  1  19  2 108   23     0     6     6    25     0     6     6    25     0     6     6
2  2  21  1 110   15     6     3     5     9     6     3     5     5     5     3     5
3  3  23  2 105   15     4     5     6    16     1     4     5    14     5     6     6
4  4  21  1 108   13     9     4     5    12     7     3     4     5    11     5     5
5  5  25  2 112   21     0     5     1     8    10     4     3    16     6     6    -1
6  6  21  2 118   13    -1     5     6    15     5     6     6    15     6     4     6
7  7  18  1 115   17     4     3     4    NA    NA    NA    NA    16     7     3     4
8  8  19  2 103   11     4     6     6    19     1     6     6    21     0     5     6
9  9  20  2 100     9     8     3     4    10     5     2     4     9     7     3     4
10 10  20  1  97     8     8     1    -1     1    12     0    -1    NA    NA    NA    NA
```

blank: No response
-1: Missing values

27

R (RStudio) の基本ルール5

- データ抽出には比較演算子を用いる

```
==   =
!=   ≠
ほか <=, >=, <, >
また、is.null(), is.na() 等
```

- 例

- sex が 2 の行を抜き出す
y[y\$sex == 2,]
- sex が 2 かつ iq が 100 以上の行を抜き出す
y[y\$sex == 2 & y\$iq >= 100,]

- 今回は -1 のものをすべてNAと置き換えたいので、次のように実行します

```
61
62 # define NA
63 y[y == -1] <- NA
64
```

28

R (RStudio) の基本ルール6

・変数の定義が重要です

データ型	例	適用可能データ	構造検査	構造変換
論理値	TRUE, FALSE	スカラー～行列	is.logical()	as.logical()
文字列	"a", "b", "c"	スカラー以上	is.character()	as.character()
整数	-1, 0, 1	スカラー～行列	is.integer()	as.integer()
実数	-1, 0.5, 1.23	スカラー～行列	is.numeric()	as.numeric()
複素数	1 + 2i	スカラー～行列	is.complex()	as.complex()
順序なし因子	"Boy", "Girl"	スカラー～行列	is.factor()	as.factor()
ベクトル	c(1, 2, 3)	ベクトル以上	is.vector()	as.vector()
行列	matrix(1:6, nrow = 2, ncol = 3)	ベクトル以上	is.matrix()	as.matrix()
データフレーム	1つ以上のベクトル	ベクトル以上	is.data.frame()	as.data.frame()
リスト	1つ以上のデータ セット	ベクトル以上	is.list()	as.list()

29

R (RStudio) の基本ルール6

・変数の定義が重要です

データ型	例	適用可能データ	構造検査	構造変換
論理値	TRUE, FALSE	スカラー～行列	is.logical()	as.logical()
文字列	"a", "b", "c"	スカラー以上	is.character()	as.character()
整数	-1, 0, 1	スカラー～行列	is.integer()	as.integer()
実数	-1, 0.5, 1.23	スカラー～行列	is.numeric()	as.numeric()
複素数	1 + 2i	スカラー～行列	is.complex()	as.complex()
順序なし因子	"Boy", "Girl"	スカラー～行列	is.factor()	as.factor()
			is.vector()	as.vector()

```

64
65 # change variable type
66 y$id <- as.character(y$id)
67 y$sex <- as.factor(y$sex)
68 summary(y)
69 class(y)
70 class(y$sex)
71 is.factor(y$sex)
72 is.numeric(y$sex)
73

```

Attention!
Rではfactor変数が鬼門です

is.data.frame()	as.data.frame()
is.list()	as.list()

30

Stack point

すいません、2月5日（水）に変更したので、それ以前にダウンロードした方は上の方にあるこちらを実行してから75行目以降を試してみてください

```

73
74 # stack point in factor variables
75 y <- read.csv("data2.csv", header=T)
76 table(y$ssqb_0)
77 y$test <- as.factor(y$ssqb_0)
78 summary(y$test)
79 y$test2 <- as.numeric(y$test)
80 table(y$test2)
81

```

31

factorは自動判別されて読み込まれることが多々あります

回避策

- **stringsAsFactors = F とする**
`read.csv("aaa.csv", header = T,
stringsAsFactors = F)`
- **最初からデータ型をすべて指定する**
`read.csv("aaa.csv", header = T,
colClasses = c("character", "numeric",
"factor", rep("numeric", 15)))`

32

Q2

- 2年後のうつ症状 (GHQ12) は、ベースラインと1年後のうつ症状どちらから影響を受けているでしょう

```
lm (ghq12_2 ~ ghq12_0 + ghq12_1, data = y)
```

- 2年後のうつ症状は、ベースラインのうつ症状と Wellbeingどちらから影響を受けているでしょう

33

主な統計解析手法

- t test, U test
- Paired t test
- Pearson' s r, Spearman' s rho
- ANOVA, ANCOVA
- Repeated measures ANOVA
- General linear model (GLM)
- Chi square test
- Multiple regression
- Logistic regression model
- Generalized linear model
- Generalized linear mixed model
- Structure equation model

34

主な統計手法の分類

検討したい従属変数、独立変数(交絡変数)の種類がわかれば、検定方法は自ずと決まります

				従属変数 (目的変数)		
				名義変数	順序変数もしくは量的変数	
				non-parametric		parametric
独立変数 (説明変数)	名義変数	2つ	反復測定なし	カイ二乗検定	Mann-Whitney U検定	t 検定
			反復測定あり	McNemar検定	Wilcoxon順位和検定	paired-t検定
		3つ以上	反復測定なし	カイ二乗検定	Kruskai-Walis検定	ANOVA
			反復測定あり	CochranのQ検定	Friedman検定	repeated-measures ANOVA
	順序変数 もしくは 量的変数	方向性なし (相関)		NA	Spearman's rho	Pearson's r, 偏相関
		方向性あり		ロジスティック回帰分析	多項ロジスティック分析	重回帰分析
			拡張モデル 一般化線系モデル (Generalized linear model) 一般化線型混合モデル 共分散構造分析 ベイズ統計学 ...			

35

例題

1. 男女 (Binary)間で抑うつ症状得点 (continuous) の差を知りたいとき	
2. 男女 (Binary)間で抑うつの有無 (Binary)の差を知りたいとき	3. 年齢層 (若中老の3 category)間で抑うつ症状得点 (continuous)の差を知りたいとき
4. 年齢 (continuous)と抑うつ症状得点 (continuous)の相関関係を知りたいとき	5. 年齢 (continuous)が抑うつ症状得点 (continuous)に影響を与えるという因果関係を知りたいとき
6. 年齢 (continuous)が抑うつの有無 (binary)に影響を与えるという因果関係を知りたいとき	

				従属変数 (目的変数)		
				名義変数	順序変数もしくは量的変数	
				non-parametric		parametric
独立変数 (説明変数)	名義変数	2つ	反復測定なし	カイ二乗検定	Mann-Whitney U検定	t 検定
			反復測定あり	McNemar検定	Wilcoxon順位和検定	paired-t検定
		3つ以上	反復測定なし	カイ二乗検定	Kruskai-Walis検定	ANOVA
			反復測定あり	CochranのQ検定	Friedman検定	repeated-measures ANOVA
	順序変数 もしくは 量的変数	方向性なし (相関)		NA	Spearman's rho	Pearson's r, 偏相関
		方向性あり		ロジスティック回帰分析	多項ロジスティック分析	重回帰分析
			拡張モデル 一般化線系モデル (Generalized linear model) 一般化線型混合モデル 共分散構造分析 ベイズ統計学 ...			

36

回答1

				従属変数 (目的変数)		
				名義変数	順序変数もしくは量的変数	
				non-parametric		parametric
独立変数 (説明変数)	名義変数	2つ	反復測定なし	カイ二乗検定	Mann-Whitney U検定	t検定
			反復測定あり	McNemar検定	Wilcoxon順位和検定	paired-t検定
	3つ以上	反復測定なし	カイ二乗検定	Kruskal-Wallis検定	ANOVA	
		反復測定あり	CochranのQ検定	Friedman検定	repeated-measures ANOVA	
	順序変数もしくは量的変数	方向性なし (相関)	NA	Spearman's rho	Pearson's r, 偏相関	
			方向性あり	ロジスティック回帰分析	多項ロジスティック分析	重回帰分析
		拡張モデル	一般化線形モデル (Generalized linear model) 一般化線形混合モデル 共分散構造分析 ベイズ統計学 ...			

- 男女 (Binary) 間で抑うつ症状得点 (continuous) の差を知りたいとき
- 男女 (Binary) 間で抑うつの有無 (Binary) の差を知りたいとき
- 年齢層 (若中老の3 category) 間で抑うつ症状得点 (continuous) の差を知りたいとき
- 年齢 (continuous) と抑うつ症状得点 (continuous) の相関関係を知りたいとき
- 年齢 (continuous) が抑うつ症状得点 (continuous) に影響を与えるという因果関係を知りたいとき
- 年齢 (continuous) が抑うつの有無 (binary) に影響を与えるという因果関係を知りたいとき

37

回答2

				従属変数 (目的変数)		
				名義変数	順序変数もしくは量的変数	
				non-parametric		parametric
独立変数 (説明変数)	名義変数	2つ	反復測定なし	カイ二乗検定	Mann-Whitney U検定	t検定
			反復測定あり	McNemar検定	Wilcoxon順位和検定	paired-t検定
	3つ以上	反復測定なし	カイ二乗検定	Kruskal-Wallis検定	ANOVA	
		反復測定あり	CochranのQ検定	Friedman検定	repeated-measures ANOVA	
	順序変数もしくは量的変数	方向性なし (相関)	NA	Spearman's rho	Pearson's r, 偏相関	
			方向性あり	ロジスティック回帰分析	多項ロジスティック分析	重回帰分析
		拡張モデル	一般化線形モデル (Generalized linear model) 一般化線形混合モデル 共分散構造分析 ベイズ統計学 ...			

- 男女 (Binary) 間で抑うつ症状得点 (continuous) の差を知りたいとき
- 男女 (Binary) 間で抑うつの有無 (Binary) の差を知りたいとき
- 年齢層 (若中老の3 category) 間で抑うつ症状得点 (continuous) の差を知りたいとき
- 年齢 (continuous) と抑うつ症状得点 (continuous) の相関関係を知りたいとき
- 年齢 (continuous) が抑うつ症状得点 (continuous) に影響を与えるという因果関係を知りたいとき
- 年齢 (continuous) が抑うつの有無 (binary) に影響を与えるという因果関係を知りたいとき

38

回答3

				従属変数 (目的変数)		
				名義変数	順序変数もしくは量的変数	
				non-parametric		parametric
1.	男女 (Binary)間で抑うつ症状得点 (continuous) の差を知りたいとき					t 検定
2.	男女 (Binary)間で抑うつの有無 (Binary)の差を知りたいとき	名義変数	2つ	反復測定なし カイ二乗検定	Mann-Whitney U検定	t 検定
			反復測定あり	McNemar検定	Wilcoxon順位和検定	paired-t検定
3.	年齢層 (若中老の3 category)間で抑うつ症状得点 (continuous)の差を知りたいとき	名義変数	3つ以上	反復測定なし カイ二乗検定	Kruskal-Wallis検定	ANOVA
			反復測定あり	CochranのQ検定	Friedman検定	repeated-measures ANOVA
4.	年齢 (continuous)と抑うつ症状得点 (continuous)の相関関係を知りたいとき	独立変数 (説明変数)		方向性なし (相関)	NA	Spearman's rho Pearson's r / 偏相関
						ロジスティック回帰分析 多項ロジスティック分析 重回帰分析
5.	年齢 (continuous)が抑うつ症状得点 (continuous)に影響を与えるという因果関係を知りたいとき	順序変数もしくは量的変数	方向性あり	拡張モデル	一般化線形モデル (Generalized linear model) 一般化線形混合モデル 共分散構造分析 ベイズ統計学 ...	
6.	年齢 (continuous)が抑うつの有無 (binary)に影響を与えるという因果関係を知りたいとき					

39

R, SPSS対応表

SPSS	R
重複ケースの特定	
ケースの並び替え Sort cases by id.	X[order(x\$id,decreasing=F),]
変数の並び替え	
行と列の入れ替え	t(x)
ファイルの結合 (変数の追加)	x <- cbind(x, new.var)
ファイルの結合 (ケースの追加)	x <- rbind(x, new.case)
再構成	
ファイルの分割	x.male <- x[x\$sex==1,] x.female <- x[x\$sex==2,]
ケースの選択	
ケースの重み付け	table()
他の変数への値の再割り当て recode age (0 thru 19=1)(20 thru 49=2)(50 thru 99=3)(else=sysmis). recode age (-1=sysmis)(0 thru 19=1)(20 thru 49=2)(50 thru hi=3) into agecategory.	expss::recode(x\$age) = c(0:19 ~ 1, 20:49 ~ 2, 50:99 ~ 3, other ~ NA) x\$agecategory = recode(x\$age, -1 ~ NA, 0 %thru% 19 ~ 1, 20 %thru% 49 ~ 2, 50 %thru% 99 ~ 3, other ~ copy) recode(x\$age, -1 ~ NA, 0 %thru% 19 ~ 1, 20 %thru% 49 ~ 2, 50 %thru% hi ~ 3) %into% x\$agecategory
連続変数のカテゴリ化	x\$agecategory <- ifelse(x\$age>20,1,0)

* SPSSの場合、大文字小文字の区別はない。最後にピリオド(.)で締める必要がある。

40

R, SPSS対応表

SPSS	R
記述統計：度数分布表	
記述統計：クロス集計表	<code>chisq.test(x\$sex,x\$category)</code> <code>chisq.test(x\$sex,x\$category)\$residuals # post-hoc residual analysis</code>
平均の比較：グループの平均	<code>mean(x\$age, na.rm = T)</code>
平均の比較：1 サンプルの t 検定	
平均の比較：独立したサンプルの t 検定	<code>t.test(x\$age ~ x\$sex, var.equal = T)</code>
平均の比較：対応のあるサンプルの t 検定	<code>t.test(age1, age2, paired=T)</code>
平均の比較：一元配置分散分析	<code>summary(aov(x\$age ~ x\$3categories))</code> <code>summary(aov(age ~ 3categories, data = x))</code> <code>anova.result <- aov(age ~ 3categories, data = x)</code> <code>summary(anova.result)</code> <code>TukeyHSD(anova.result)</code>
一般線形モデル：1 変量	<code>lm()</code>
一般線形モデル：多変量	
一般線形モデル：反復測定	<code>aov(var1 ~ age * sex + Error(ID), data = x)</code> OR https://www.wantedly.com/companies/diligence/post_articles/74556

41

R, SPSS対応表

SPSS	R
一般化線形モデル	<code>glm</code>
混合モデル：線形	<code>lmer</code>
混合モデル：一般化線形	<code>glmer</code> or <code>nlme</code> packages
相関：2 変量	<code>cor.test(x\$age,x\$iq)</code> <code>psych::corr.test(x) # correlation matrix</code>
相関：偏相関	
回帰：線形	<code>lm(sex ~ age + iq, data=x)</code>
回帰：二項ロジスティック	<code>glm(sex ~ age + iq, data=x, family="binomial")</code>
回帰：多項ロジスティック	
次元分解：因子分析	See page below
ROC曲線	

42

R, SPSS対応表

SPSS	R
グラフ：棒	plot(x\$sex)
グラフ：折れ線	
グラフ：円	
グラフ：箱ひげ図	boxplot(x\$age) boxplot(x\$age ~ x\$sex) boxplot(age ~ sex, data = x) plot(age ~ sex, data = x) plot(age ~ sex + categories, data = x)
グラフ：エラーバー	
グラフ：散布図	plot(x\$ghq12 ~ x\$age)
グラフ：ヒストグラム	hist(x\$age)

43

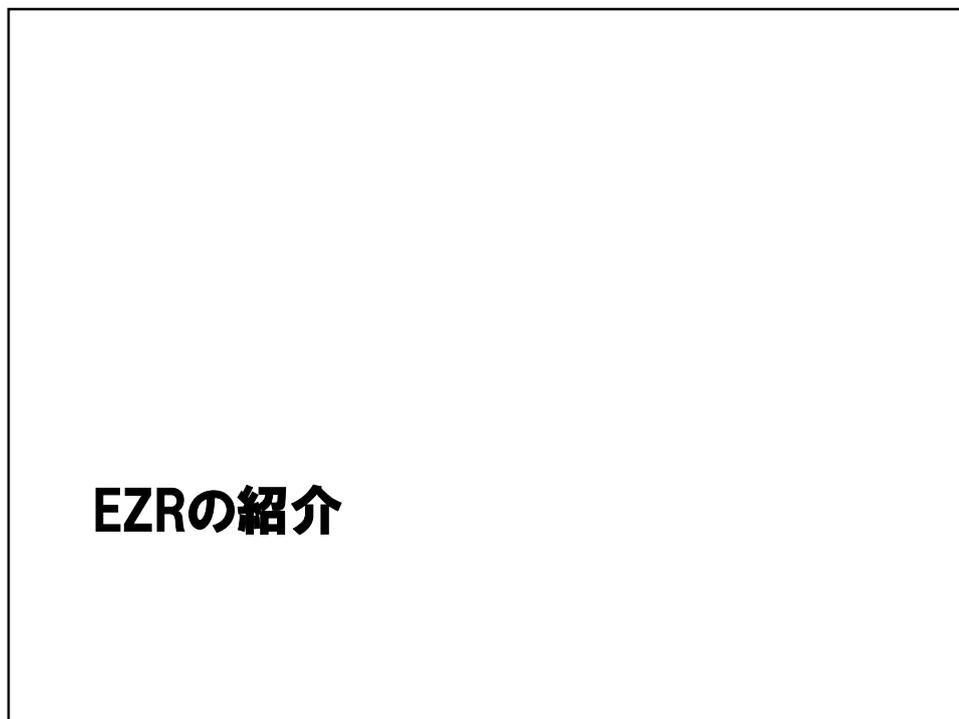
Excel, SPSS対応表

Excel	R
countif	expss::count_if(1,x\$sex) expss::count_col_if(gt(20),x\$age) expss::count_row_if(le(0),x)
sumif	expss::sum_if(gt(20),x\$age)
vlookup	expss::vlookup(x\$id,newlist,2) expss::vlookup(x\$id,newlist,"new.va r")

<1 lt(1)
 <=1 le(1)
 <>1 ne(1)
 =1 eq(1)
 >=1 ge(1)
 >1 gt(1)

First column of the 'newlist' should be the lookup variable (i.e. id).

44



45

自治医科大学附属さいたま医療センター
血液科のホームページへようこそ。

自治医科大学附属さいたま医療センター
Jichi Medical University Saitama Medical Center

***** Click here for English version *****

2019年12月1日 EZR version 1.41公開
 2014年11月初心者向けマニュアル刊行
 2015年4月EZRマニュアル第2版刊行
 (1クリックする立ち読みすることができます)
 2016年10月みんなのEBMと臨床研究刊行

Bone Marrow Transplantation (2013) 48, 452–458
 © 2013 Macmillan Publishers Limited All rights reserved 0268-3369/13
 www.nature.com/bmt

TECHNICAL REPORT
 Investigation of the freely available easy-to-use software
 ‘EZR’ for medical statistics

Y Kanda

Although there are many commercially available statistical software packages, only a few implement a competing risk analysis or a proportional hazards regression model with time-dependent covariates, which are necessary in studies on hematopoietic SCT. In addition, most packages are not clinician friendly, as they require that commands be written based on statistical languages. This report describes the statistical software ‘EZR’ (Easy R), which is based on R and R commander. EZR enables the application of statistical functions that are frequently used in clinical studies, such as survival analyses, including competing risk analyses and the use of time-dependent covariates, receiver operating characteristics analyses, meta-analyses, sample size calculation and so on, by point-and-click access. EZR is freely available on our website (<http://www.jichi.ac.jp/saitama-sct/SaitamaHP/files/statmed.html>) and runs on both Windows (Microsoft Corporation, USA) and Mac OS X (Apple, USA). This report provides instructions for the installation and operation of EZR.

Bone Marrow Transplantation (2013) 48, 452–458; doi:10.1038/bmt.2012.244; published online 3 December 2012
Keywords: statistical software; EZR; competing risk; time-dependent covariate

46