





■データフレームについて(復習) □
手入力によるデータフレームの作成 □テキストファイルからのデータの読み込み □EXCELファイルからのデータの読み込み データ加T・データハンドリング □データへのアクセス,データの抽出・加工 ■データフレームを使った統計解析の例 □回帰分析 □ 生存時間解析

第3回のメニュー



データフレームについて(復習) データ加工・データハンドリング データフレームを使った統計解析の例





データフレームとは



数値ベクトルや文字ベク トル,因子ベクトル(文 字型ベクトル)などの異 なる型のデータをまとめ てもつ変数

> 各列の要素の型はバラ バラでも構わない

- 外見は行列と同じ
- データフレームの各行・
 列はラベルを必ず持ち、
 ラベルによる操作が可能

SEX	HEIGHT	WEIGHT
F	158	51
F	162	55
М	177	72
М	173	57
М	166	64



ベクトル(や行列,リストなど)からデータフレームを作成する

「性別」「身長」「体重」データをベクトル で用意して,それらを関数 <u>data.frame()</u>で 1つのデータフレームに変換する

ファイルにあるデータを読み込んでデータフ
 レームを作成する

関数 <u>read.table()</u> などでファイルから データを読み込む



SEX	HEIGHT	WEIGHT		SEX	HEIGHT	WEIGHT
F	158	51		F	158	51
F	162	55	data.frame()	F	162	55
М	177	72		М	177	72
М	173	57		М	173	57
М	166	64		М	166	64

sex <- c("F", "F", "M", "M", "M")
height <- c(158, 162, 177, 173, 166)
weight <- c(51, 55, 72, 57, 64)</pre>



データフレームを生成すると・・・
 「データフレームの『性別』」や「データフレームの『体重』」としてデータを取り出すことが出来る
 取り出す方法は「データフレーム名¥\$ 列名」などとすればよい.

HEIGHT WEIGHT SEX # 身長 x\$HEIGHT F 158 51 [1] 158 162 177 173 166 F 162 55 # 体重の平均 mean(x\$WEIGHT) 177 72 М [1] 59.8 173 57 М # 2 列目を表示 x[,2] М 166 64 [1] 158 162 177 173 166



国数 summary() を使うことでデータフレームの列ごとの特徴を見ることが出来る。

summary(x)

SEX	HEIGHT	WEIGHT
F:2	Min. :158.0	Min. :51.0
M:3	1st Qu.:162.0	1st Qu.:55.0
	Median :166.0	Median :57.0
	Mean :167.2	Mean :59.8
	3rd Qu.:173.0	3rd Qu.:64.0
	Max. :177.0	Max. :72.0

SEX	HEIGHT	WEIGHT
F	158	51
F	162	55
М	177	72
М	173	57
М	166	64



- ファイルからデータを読み込むには・・・
 - □ まず, データがあるディレクトリ(フォルダ)に作業ディレクト リを変更する
 - □ 次に, 関数 read.table() などでファイルからデータを読み込む
- ファイルからデータやプログラムを読み込んだり,ファ イルにデータを書き出したりする場所を作業ディレクト リという
- 指定したディレクトリに指定した作業ディレクトリに データがセーブされたり, R 用エディタなどが保存され るようになる

getwd() [1] "c:/usr"

setwd("c:/usr") # 作業ディレクトリを変更 # 現在のディレクトリを確認

作業ディレクトリの変更(Windows)

RGui	
ファイル 編集 その他 パック	ージ ウインドウ ヘルプ
Bコードのソース 新しいスクリプト	🕶 🖨
スクリプトを開い ファイルの表示	
作業スペースの読み込み… 作業スペースの保存…	
履歴の読み込み 履歴の保存	
ディレクトリの変更	
印刷 ファイルを保存	
終了	
ァイル] の [デ- を译	ィレクトリの _{辈択}

Change directory
作業ディレクトリの変更
C:¥Program Files¥R¥rw2010pat Browse
OK Cancel

[Browse] をクリックして, 変更先のディレクトリを選択



[その他] の 作業[ディレクトリの変更] を選択

000 新しい作業ディレクトリを選択 4 1 〒スクトップ \$ RR Network 10 グラブ Macintosh HD R 〒 デスクトッフ 1 x 🗛 アプリケー 書類 ムービー ミュージック ビクチャ (キャンセル

変更先のディレクトリを選択



(1) 列名がなく、データ間がスペースで区切ら れている場合 R が勝手に列名を決めている

x <- read.table("data01.txt")</pre>

5 M 166 64

🍩 C:¥			
F	158	51	
F	162	55	
М	177	72	
М	173	57	
М	166	64	

data01.txt



(2) 列名があり,データ間がスペースで区切ら れている場合

X	<- r	ead.tab	ole("data0	2.txt",	🇢 C:¥			<u>- 0 ×</u>
			×	header=T)	sex	height	weight	
	sex	height	weight	· · · ·	F	158	51	
1	Б	150	лотунц Б1		F	162	55	
I	Г	100	51		М	177	72	
2	F	162	55		М	173	57	
3	М	177	72		М	166	64	
4	М	173	57					
5	М	166	64					

data02.txt



(3)1行目にコメント,2行目に列名があり, データ間がスペースで区切られている場合

х -	<- re	ad tabl	e("data0)3.txt".		≪ C:¥			
			hea	<u>der=T, ski</u>	<u>ip=1</u>)	### (data03	txt	
Ś	sex h	eiaht w	eiaht			sex f	neight	weight	
	_	4=0				F	158	51	
1	F	158	51			F	162	55	
2	F	162	55			М	177	72	
3	М	177	70			М	173	57	
5	IVI	177	12			М	166	64	
4	М	173	57						
5	М	166	64						

data03.txt



(4) 列名があり,データ間がコンマで区切ら れている場合

x	<- re	ad.tab	ole("data0	4.txt".	🆘 C:¥	
	sex h	eight	heade weight	<u>er=T</u> , <u>sep=","</u>)	sex,height,weight F,158,51	
1	F	158	51		F,162,55 M,177,72	
2	F M	162	55 72		M,173,57	
3 4	M	173	57		M, 100,04	
5	М	166	64			

data04.txt

データフレームの作成(EXCEL)



- ■目的は関数 read.csv() で読み込める形式に すること(前節の data04.txt の状態)
 - ■まず, EXCEL ファイルを開き,メニューの [ファイル]の[開く]から,[名前をつけて保存] を選択する
 - □保存する名前をつけた後,次に[ファイルの種類] から[CSV カンマ区切り]を選択して保存する



■Windows版Rの場合

🔀 Micı	rosoft Excel – datafr	ame.xls	
77	ァイル(E) 編集(E) 表示(☑ 挿入① 書式(<u>の</u> ツール(T)
	新規作成(<u>N</u>)	Ctrl+N	ר - 🍓 Σ
	; 開((<u>O</u>)	Ctrl+O	
	上書き保存(S)	Ctrl+S	D
1	名前を付けて保存(<u>A</u>)		
2	ページ設定(U)		
4	印刷範囲(工)	+	
5)印刷(<u>P</u>)	Ctrl+P	
6	送信(<u>D</u>)	•	
8	<u>1</u> ¥Documents and Sett	ing¥rand.csv	
9	2 ¥Documents and Sett	ings¥X¥1.csv	
10	<u>3</u> ¥Documents and Sett	ing¥1.csv.xls	
11	<u>4</u> ¥Documents and S¥	dataframe.csv	
12	×		
13			

別名で保存



CSV (カンマ区切り) で保存

データフレームの作成(EXCEL)



■Mac OS X 版 R の場合

 ▲ A 新規作成 第人 第人	0	プロジェクト ギャラリー	🗘 ፝ ж Р 📄	2016-03	K17 :		
2 60 DN Cuil D	A y	新規作成 聞く	жN жO		(四書類		
4 63 5 (R存 365 72 (R存 365 72 (R) (R) 8 66 (R) 7 68 (R) 8 63 (R) 9 65 (R) 0 64 (R) (R) 7 (R) (R) 2 (R) (R) 7 (R) (R) 8 (R) (R) 7 (R) (R) 8 (R) (R) 7 (R) (R) 8 (R) (R) 9 (T) (R) 7 (R) (R) 8 (R) (R) 9 (T) (R) 1 (R) (R) 2 (R) (R) 3 (R) (R) 4 (T) (R) 7 (R) (R) 8 (R) (R) 9 (T) (R) <td>60 59</td> <td>閉じる</td> <td>жw</td> <td>Macintash HD</td> <td> サイト デスクトップ </td> <td>P dataframe. P dataframe.</td> <td>cia.</td>	60 59	閉じる	жw	Macintash HD	 サイト デスクトップ 	P dataframe. P dataframe.	cia.
66 加名 C+455 68 Web ページとして保存 63 作業状態の保存 63 ブラウザでブレビュー 64 ブラウザでブレビュー 1 ページ設定 7リント範囲 ムービー 2 ブリント範囲 7リント 第日 2 ブリント 7リント 第日 2 ブリント 7 ブリント 2 ブリント 7 アンワント 2 ブリント 7 アント 2 アンリント 3 ブリント 4 ブリント 7 第日 3 ブロパティ 4 アロパティー 5 ブロパティ 5 ブロパティ 5 ブロパティ 5 アンパーマット 5 アンパーマット 5	63 72	保存	жs	R ±	③ パブリック 目 ビクチャ	Excelition Excelition Instructure	7x1,45
63 作業状態の保存 65 ブラウザでプレビュー 64 ブラウザでプレビュー グリント範囲 ムービー ブリント ブレビュー ムービー ブリント ジレビュー エージック ブリント 第6 ブロパティ 第6 ブロパティ 第7 1 Macintosh HD:Users:x:D:dataframe.xls 2 2 Macintosh HD:Users:x:idataframe.csv 7	66 68	がれてはない。 Webページとして保存		デスクトップ	■ ミュージック ■ ムーピー	 KR-menu t Microsoft 3 	d Lーザーデータ
64 ブラウザでプレビュー アリントジン 第二 アリントジン 第二 アリントジン アリントジン アリントジン アリントジン アリントジン アリント Re-100 アリント Re-21.0.txt Re-27.10.txt Re-7.2.3.b/h ブリント ブリント ブロパティ 第 ア ア ア Re-100 Re-100 Re-21.0.txt Re-7.2.3.b/h ブロパティ 第 ア ア ア ア Re-100 Re-100 Re-100 Re-100 Re-100 Re-21.0.txt Re-21.0.txt Re-7.2.3.b/h Re-100 Re	63 65	作業状態の保存		1 ×	2-72B	pdf 7 ≥ √)	を出力する,ht
ページ設定 プリント範囲 ▲ービー ■ R-2.1.0.dmg プリント プレビュー ■ アージ酸ル ■ R-2.1.0.txt ■ R-2.1.0.txt プリント プレビュー ■ アージ酸ル ■ R-7-ジ酸ル ■ R-7-ジ酸ル プリント 第P ■ ビッチャ ■ Rep 送信 ■ アーマット ■ V Excel ブック プロパティ ■ 1 Macintosh HD:Users:x:D:dataframe.xls ■ 2 × / Littict Ret To ALP ■ CSV (力ンマ区切り) XML スプレッドシート × ML スプレッドシート	64	ブラウザでプレビュー		「青樹	國際的	P 〒 R - 行利や	ルーマーストJitu デーロで出力.ht
プリント プレビュー プリント プレビュー プリント パレビュー ガリント プリント 第P 送信 アロパティ プロパティ 1 Macintosh HD:Users:x:D:dataframe.xls 2 Macintosh HD:Users:x::dataframe.csv		ページ設定 プリント範囲		₩ 4-ビ-		⊒ R-2.1.0.dn ≥ R-2.1.0.txt	9
プリント 第P 送信 フォーマット: プロパティ プロパティ 1 Macintosh HD:Users:x:D:dataframe.xls 2 Macintosh HD:Users:x::dataframe.csy		プリント プレビュー		€ 21-599		 R-アーダ語 ■ Rconsole 	9达-9-txt
送信 フォーマット: v Excel ブック ガロパティ プロパティ 1 Macintosh HD:Users:x:D:dataframe.xls 2 Macintosh HD:Users:x::dataframe.csv 2 Macintosh HD:Users:x::dataframe.csv		プリント	ЖР		10		
プロパティ Pacel 2004 for Mac の営業形式です。 - 2004 for Mac の営業形式です。 テンプレート 2004 for Mac でも使用されていま テンプレート CSV (カンマ区切り) XML スプレッドシート		送信	•	201	r-マット: ✓ Excel ブック	2	- 11
1 Macintosh HD:Users:x:D:dataframe.xls 2 Macintosh HD:Users:x::dataframe.csv 2 Macintosh HD:Users:x::dataframe.csv		プロパティ		Excel 2004 for Mac D ~ 2004 for Mac D 68	#学形式です。 用されていま テンプレー	京 -ト	Excel 9
2 Macintosh HD:Users:x::dataframe.csv		1 Macintosh HD:Users:x:D:dat	aframe.xls	ファイル形式に関するヘル	CSV (カン XML スプ	/マ区切り) (レッドシート	
✓ 拡張子を追加する Web ページ (HTML)		2 Macintosh HD:Users:x::data	frame.csv	☑拡張子を追加する	Web ベー	ジ (HTML)	

新規フォルダ

Sheet2 Sheet3

別名で保存

CSV(カンマ区切り)で保存

Excel 5.0/95 ブック

Excel 4.0 ワークシート Excel 3.1 ワークシート

Excel 2.2 ワークシート

Excel 4.0 ブック

保存.

OCAPS ON



データフレームの作成(read.csv)

(4') 列名があり,データ間がコンマで区切ら れている場合

x <- read.csv("data04.csv")</pre>

sex	height	weight
••••	··•	

1	F	158	51
2	F	162	55
3	М	177	72
4	М	173	57
5	М	166	64

🧇 C:¥	
sex,height,weight F,158,51 F,162,55 M,177,72 M,173,57 M,166,64	

data04.csv



データフレームの作成(read.csv)

(5) 列名がなく,データ間がコンマで区切ら れている場合

myr	name	<- c("SE>	<","HEIGHT","WEIGHT")
Χ <	<- re	ead.csv("c	lata05.csv",	
		<u>header=F</u>	, <u>col.names=myname</u>)
Se	ex he	eight weig	ght	
1	F	158	51	
2	F	162	55	
3	М	177	72	
4	М	173	57	
5	М	166	64	

🗢 C:¥	
F. 158.51	
F,162,55	
M, 177, 72	
M, 173, 57	
M,166,64	

data05.csv



■Windows 版の場合は,列名をコピーしても しなくてもよい

🔣 Mi	crosoft	Exc	el – datafra	me.xls			
1 🔊 🗆	マイル(E)	編	集(E) 表示(⊻)	挿入仰	書式(0)	ッール①	
] 🗅 (i 🚽 着	<mark>ا</mark>	元に戻す(世) 則	的付け	Ctrl+Z	- 🍓 Σ	
	A7	Đ	⊐ピ–©)		Ctrl+C		
	A		形式を選択して	て貼り付け	(S)	D	
1		_	クリア(<u>A</u>)		,		
2			 削除(<u>D</u>)				
4			シートの削除(」)			
5		åå,	検索(F)		Ctrl+F		л Б Ц
6				÷			# % J
7	x		У	ļ		-	
8		60	62				X <-
9		59	65				
10		63	73				
11		72	59				л Ті
12		66	69				# 9 J
13		68	68				
14		63	75				X <-
15		65	67				
16		64	74	ļ			

列名をコピーした場合

< read.delim("clipboard", header= T)</pre>

列名をコピーしなかった場合

x <- read.delim("clipboard", header= F)



|データフレームの作成 (EXCELのセルをコピー&ペースト)

Mac OS X 版の場合は,列名をコピーして はいけない

}

Ć	Exce	1 ファ	イル	編集	表示	挿入	書式	ツーノ	レ データ
0	00			元に	戻せま	せん			жz [
\diamond	A	В	С	繰り)返しで	きませ/	h		ЖY
1	x	У		+					
2	60	62		779	' Γ				ЖX
3	59	65							ЖC
4	63	73		スク	ラップ	ブック	EDR-	-	ΰжс
5	72	59		<u>∧</u> °−	・スト				жv
6	66	69		スク	ラップ	ブックフ	からペー	-スト	企業V
7	68	68		形式	た選択	してペー	-スト		
8	63	75		11-1	パーリ	ンクとし	してペー	-スト	
9	65	67							
10	64	74		フィ	ル				•
11				消去	ŧ				
12				削除	È				
13				×)-	トの削	涂			
14				3/-	トの税	小 計≠ ナ-1	+ フピ_	_	
15				-	1.07/3/	町みたい	a		_
16				検索					ЖF
17			- 1	置披	4				
18			_	= 15	~/-				
19				21					
20				リン	クの設定	定			-
21				オフ	ジェク	5			-
22			_	12	2 2 2 2				
23				Color States	Contraction in the second second	Contraction of the local division of the loc	Sector Sector Sector		and a second

```
excel.mac <- function(...) {
  args <- c(...)
  temp <- matrix(scan(""), byrow=TRUE,
  ncol=length(args))
  data <- data.frame(temp)
  colnames(data) <- args
  return(data)</pre>
```

```
excel.mac("X", "Y") # 列名を入力
1: # ペーストする
```

データフレームの作成 (xlsファイルを直接読み込む)

■gregmisc パッケージと ActivePerl を使う



(1) ActiveState にアクセス http://www.activestate.com/

(2) ActivePerl をインストール

(4) 関数 read.xls() を 使う







データフレームについて(復習)
 データ加工・データハンドリング
 データフレームを使った統計解析の例



データへのアクセス方法



コマンド	機能
x\$列名,x["列名"], x[["列名"]]	列データを表示
x[2], x[[2]]	2 番目の列データを表示
x[3, 2], x[[3, 2]]	3 行 2 列目のデータを表示
x[[3,"列名"]], x[[3,"列名"]]	指定した列の 3 行目のデータを表示
x[c(1, 2)]	1 列目と 2 列目のデータを表示
x[c(3, 4),]	3 行目と 4 行目のデータを表示
x[,c(T,F,T)]	論理ベクトル c(T,F,T) が TRUE と なっている列を表示
x[SEX=="F",]	性別が F(女性)である行を表示
x[,SEX=="F" & WEIGHT>50]	性別が F(女性)かつ体重が 50kg 以上である行を表示

データへのアクセス方法(1)



X	[<mark>c(1</mark> ,	3,5),]	
	sex	height	weight
1	F	158	51
3	М	177	72
5	М	166	64
X	<mark>, c(</mark> 1 sex	<mark>,3)]</mark> weight	
1	F	51	
2	F	55	
3	М	72	
-	111	. –	

64

5

М

1,3,5行目にアクセス

#1,3列目にアクセス

<u>データフレーム x</u>						
SEX	HEIGHT	WEIGHT				
F	158	51				
F	162	55				
М	177	72				
М	173	57				
М	166	64				

10	-
00	
	DY
100	

データへのアクセス方法(2)



データフレーム\$列名で指定する

x\$height # 身長データ [1] 158 162 177 173 166

x\$height <- NULL # 身長を削除

X

sex weight

- 1 F 51
- 2 F 55
- 3 M 72
- 4 M 57

5 М 64

デー	タ	フ	レー	Δ	Х

SEX	HEIGHT	WEIGHT
F	158	51
F	162	55
М	177	72
М	173	57
М	166	64



コマンド	機能
ncol(x)	x の列数(項目数)を求める
nrow(x)	x の行数(データ数)を求める
names(x)	x の列名を表示する
rbind(x,y)	xとyを縦に並べて結合する
cbind(x,y)	x と y を横に並べて結合する
data.frame(x,y)	x と y を横に並べて結合する
merge(x,y)	x と y を併合(マージ)する

データフレームのマージ(1)



W

51

55

57

55

Ε

merge(D1, D2)
 ID H W
 1 A 158 51

2 E 166 55

merge(D1, D2, all=T)

- ID H W 1 A 158 51
- 2 C 177 NA
- 3 E 166 55
- 4 B NA 55

5 D NA 57

D)1	D	2
ID	Н	ID	
А	158	А	
С	177	В	
Е	166	D	





merge(D1, D2, all=T)

	ID	VIT	Н	W
1	А	1	155	61
2	А	2	158	65
3	С	1	156	NA
4	С	2	159	NA
5	Е	1	157	64
6	Е	2	160	68
7	В	1	NA	62
8	В	2	NA	66
9	D	1	NA	63
10	D	2	NA	67

	D1				D2	
ID	VIT	Н		ID	VIT	W
А	1	155		А	1	61
С	1	156		В	1	62
ш	1	157		D	1	63
А	2	158		ш	1	64
С	2	159		А	2	65
Е	2	160		В	2	66
			-	D	2	67

Ε

2

68





S 0	rtl	ist	<-	orde	er (DS	\$W)					
D3	<-	D[\$	sort	list	t,])						
	D	Н	W								
1	А	158	51								
3	Е	166	55								
4	В	NA	55								
5	D	NA	57								
2	С	177	NA								
ro D3	wna	mes	(D3)	<-	c(1:	:nrow((D3))	#	番号	を整	列
ro D3	wna I D	Imes (H	(D3) W	<-	c(1:	:nrow((D3))	#	番号	を整	到
ro D3 1	wna ID A	mes H 158	(<mark>D3)</mark> W 51	<-	c(1:	:nrow((D3))	#	番号	を整	到
ro D3 1 2	wna ID A E	mes H 158 166	(D3) W 51 55	<-	c(1:	:nrow((D3))	#	番号	を整	逐列
ro D3 1 2 3	wna ID A E B	H 158 166 NA	(D3) W 51 55 55	<-	c(1:	:nrow((D3))	#	番号	を整	逐列
ro D3 1 2 3 4	ID A E B D	H 158 166 NA NA	(D3) W 51 55 55 57	<-	c(1 :	:nrow((D3))	#	番号	を整	逐列

	D					
ID	Н	W				
Α	158	51				
С	177	NA				
E	166	55				
В	NA	55				
D	NA	57				





D

Η

158

177

166

NA

NA

W

51

NA

55

55

57

ID

А

С

Ε

В

D

sortlist <- order D[sortlist,] ID H W 1 A 158 51 3 <u>E 166</u> 55 4 <u>B NA</u> 55 5 D NA 57	(<mark>D\$W</mark> , pmax(<u>D\$W</u> , D\$H)) # W を昇順に並べる # W で同じ値がある場合は # H の小さい方を上にする	
2 C 177 NA		
sortlist <- order	(D\$W. pmax(D\$W. D\$ID))	ſ
D[sortlist,]	# W を昇順に並べる	
ID H W	# ₩ で同じ値がある場合は	
1 A 158 51	# ID の小さい方を上にする	F
4 <u>B NA</u> 55		╞
3 <u>E 166</u> 55		
5 D NA 57		
2 C 177 NA		

データの加工・抽出



コマンド	機能
head(x, n=a)	先頭から a 行だけ抽出する
tail(x, n=b)	末尾から b 行だけ抽出する
na.omit(x)	NA を含む行を削除する
transform(x, y=値)	データフレーム x に新たな列 y を追加 する
df[sapply(x, 論理ベクトル)]	論理ベクトルが TRUE となっている行 にのみアクセスする
subset(x, 条件式)	条件式に合う行のみを抽出する
subset(x, 条件式, ベクトル)	ベクトルで指定した列に対し,条件式に 合う行のみを抽出する
split(x,列名や条件式)	(列がカテゴリーデータならば)列名で データフレームを分割する

データの加工・抽出(1)



sum(DF\$W) DF\$W <- DF\$W * 1000 # kg から g に変換する DF ID SEX H W 1 1 F 158 51000 2 2 F 162 55000 3 3 M 177 72000 4 4 M 173 57000 5 5 M 166 64000 # 以下は DF\$G <- DF\$W * 1000 と同じ

.

transform(DF, G=DF\$W * 1000) ID SEX H W 1 1 F 158 51000 2 2 F 162 55000 3 3 M 177 72000

体重の和を求める

DF

١D	SEX	Н	W
1	F	158	51
2	F	162	55
3	М	177	72
4	М	173	57
5	М	166	64

データの加工・抽出(2)



<pre>DF\$W <- ifelse(DF\$SEX== DF</pre>	:"F", NA, DF\$W) # 女	、性の体	本重を	隠す	
ID SEX H W					
1 1 F 158 NA					
2 2 F 162 NA					
cond $ (DE^{U}) = 170)$	# 4 170 の人た地中		П	F	
DE[cond]					
		D	SEX	Н	W
3 3 M 177 72		1	F	158	51
4 4 M 173 57		2	F	162	55
subset(DF ID>3)	# ID>3の人を抽出	3	М	177	72
ID SEX H W	# 10>00>7(С)ЩЦ	4	М	173	57
4 4 M 173 57		5	М	166	64
5 5 M 166 64					





■データをセル形式で見る場合は関数 edit(データフレーム名)を用いる

DF <- edit(DF)</pre>

Rテ-	·9エディタ			_ 🗆	×
	ID	SEX	Н	Ψ	
1	1	F	158	51	
2	2	F	162	55	
3	3	M	177	72	
4	4	М	173	57	
5	5	M	166	64	

Windows 版

Tana Tana Tana	25			reand rearded
ID	SEX	H	W	sgan 1
1	F	158	51	
2	F	162	55	
3	М	177	72	
4	М	173	57	
5	М	166	64	

Mac OS X 版





- 手入力でデータフレームを作成する場合で欠損が含まれているデータを読み込む場合は、ベクトル中の 欠損部分を NA としておけば、該当部分に欠損値 (NA)が入る。
- sex <- c("F",NA,"M"); height <- c(158,162,NA);</pre>
- weight <- c(51,55,72)
- (x <- data.frame(SEX=sex, HEIGHT=height, WEIGHT=weight))
 SEX HEIGHT WEIGHT</pre>
- 1 F 158 51
- 2 <NA> 162 55
- 3 M NA 72





ファイルからデータを読み込む場合で欠損が含まれているデータを読み込む場合は、データ間がコンマで区切られている方が処理しやすい。 この場合、単に欠損部分を空白にしておけば、該当部分に欠損値(NA)が入る。

, ")

X	<-	read.tab	ole("data06.tx	ct",
			header=T,	sep="
	sex	height	weight	
1	F	158	51	
2	F	162	55	
3	М	<u>NA</u>	72	
4	М	173	57	
5	М	166	64	
4 5	M M	173 166	57 64	

🧇 C:¥	
sex, height, weight	
F, 162, 55	
M, ,72 M,173,57	
M, 166, 64	
data06 tyt	
ualau0.lXl	





データフレームについて(復習) データ加工・データハンドリング データフレームを使った統計解析の例





モデル式の立て方



■モデルについて統計的処理を行う場合の書式 関数名(モデル式) ■モデル式の例(は誤差項) \Box Y ~ X : Y = a + bX + $\Box Y \sim X_1 + X_2$: Y = a + b₁X₁ + b₂X₂ + □ Y ~ . : Y = (Y以外の変数を説明変数に) + $\Box Y \sim X_1 * X_2$: Y = a + b₁X₁ + b₂X₂ + $b_3X_1X_2$ + (交互作用モデル) □ Y ~ X₁ + X₂ + X₁*X₂ : 上と同じ交互作用モデル □ Y ~(X₁ + X₂)²: 上と同じ交互作用モデル





height <- c(177, 165, 175, 168, 171, 168, 190, 167, 173, 172, 171, 177)weight <- c(62, 65, 75, 58, 59, 66, 74, 61, 70, 80, 71, 68)bmi <- weight/(height/100)^2 MYDATA <- data.frame(BMI=bmi, HGT=height, WGT=weight) result <- Im(BMI ~ WGT, data=MYDATA)</pre> summary(result) Call: $Im(formula = BMI \sim WGT, data = MYDATA)$ Residuals: Min 10 Median 3Q Max -3.6088 -0.6052 0.1780 0.9304 1.8451 Coefficients: Estimate Std. Error t value Pr(>|t|)(Intercept) 7.02588 4.72554 1.487 0.1679 WGT 0.23083 0.06976 3.309 0.0079 ** codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.586 on 10 degrees of freedom Multiple R-Squared: 0.5226, Adjusted R-squared: 0.4749 F-statistic: 10.95 on 1 and 10 DF, p-value: 0.007896

BMI	HGT	WGT
19.78	177	62
23.87	165	65
24.48	175	75
20.54	168	58
20.17	171	59
23.38	168	66
20.49	190	74
21.87	167	61
23.38	173	70
27.04	172	80
24.28	171	71
21.70	177	68

(例)重回帰分析(1)



result <- Im(BMI ~ . , data=MYDATA)
summary(result)</pre>

Call: $Im(formula = BMI \sim ..., data = MYDATA)$ Residuals: 10 Median Min 3Q Max -0.10499 -0.03116 -0.01093 0.03824 0.12068 Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 43.579130 0.574259 75.89 6.06e-14 *** HGT -0.252870 0.003678 -68.76 1.47e-13 *** 0.336904 0.003557 94.70 8.28e-15 *** WGT codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07288 on 9 degrees of freedom Multiple R-Squared: 0.9991, Adjusted R-squared: 0.9989 F-statistic: 4956 on 2 and 9 DF, p-value: 2.039e-14

BMI	HGT	WGT
19.78	177	62
23.87	165	65
24.48	175	75
20.54	168	58
20.17	171	59
23.38	168	66
20.49	190	74
21.87	167	61
23.38	173	70
27.04	172	80
24.28	171	71
21.70	177	68

(例)重回帰分析(2)



result <- Im(BMI ~ HGT+WGT-1, data=MYDATA) # -1 は切片項を除く命令 summary(result)

Call: Im(formula = BMI ~ HGT + WGT - 1, data = MYDATA) Residuals: Min 1Q Median 3Q Max -4.2323 -0.6161 0.4206 1.1296 2.1624 Coefficients: Estimate Std. Error t value Pr(>|t|) HGT 0.005504 0.033396 0.165 0.87237 WGT 0.320069 0.085272 3.754 0.00376 ** ---codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.75 on 10 degrees of freedom Multiple R-Squared: 0.995, Adjusted R-squared: 0.994 F-statistic: 1003 on 2 and 10 DF, p-value: 3.006e-12

BMI	HGT	WGT
19.78	177	62
23.87	165	65
24.48	175	75
20.54	168	58
20.17	171	59
23.38	168	66
20.49	190	74
21.87	167	61
23.38	173	70
27.04	172	80
24.28	171	71
21.70	177	68

(例)重回帰分析(3)



result <- Im(BMI ~ HGT*WGT, data=MYDATA)
summary(result)</pre>

Call: Im(formula = BMI ~ HGT * WGT, data = MYDATA)

Residuals:

Min	1Q	Median	3Q	Max
-0.104412	-0.036019	-0.009247	0.049600	0.127699

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	39.8632534	9.3798326	4.250	0.00280	* *
HGT	-0.2311712	0.0547951	-4.219	0.00292	**
WGT	0.3895558	0.1326834	2.936	0.01883	*
HGT:WGT	-0.0003071	0.0007736	-0.397	0.70176	
codes: 0 '	***' 0.001 '	'**' 0.01 '*	' 0.05	'.' 0.1 '	' 1

Residual standard error: 0.07655 on 8 degrees of freedom Multiple R-Squared: 0.9991, Adjusted R-squared: 0.9988 F-statistic: 2995 on 3 and 8 DF, p-value: 1.540e-12

BMI	HGT	WGT
19.78	177	62
23.87	165	65
24.48	175	75
20.54	168	58
20.17	171	59
23.38	168	66
20.49	190	74
21.87	167	61
23.38	173	70
27.04	172	80
24.28	171	71
21.70	177	68

(例)重回帰分析(4)



result <- Im(BMI ~ HGT*WGT-1, data=MYDATA) # -1 は切片項を除く命令 summary(result)

Call: Im(formula = BMI ~ HGT * WGT - 1, data = MYDATA) Residuals: Min 1Q Median 3Q Max -0.14473 -0.08449 -0.02458 0.06949 0.21044 Coefficients: Estimate Std. Error t value Pr(>|t|) HGT 1.619e-03 2.487e-03 0.651 0.531 WGT 9.520e-01 1.620e-02 58.751 6.04e-13 HGT:WGT -3.588e-03 8.466e-05 -42.384 1.13e-11 *** 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 codes:

Residual	star	ldard	error	: 0).1	303	on	9	degrees	of	fre	edor	n
Multiple	R-Sc	luarec	:	1,		A	١dj	ust	ed R-squ	lare	ed:		1
F-statist	tic:	1.213	3e+05	on	3	and	9	DF,	p-value): <	< 2.	2e-7	16

BMI	HGT	WGT
19.78	177	62
23.87	165	65
24.48	175	75
20.54	168	58
20.17	171	59
23.38	168	66
20.49	190	74
21.87	167	61
23.38	173	70
27.04	172	80
24.28	171	71
21.70	177	68

(例)生存時間解析(1)



Χ

Nonmaintained

Nonmaintained

Nonmaintained

Nonmaintained

Nonmaintained

Nonmaintained

.

Maintained

Maintained

Maintained

- 急性骨隋白血病データ aml
 time:「生存時間」または「打ち切りまでの時間」
 - status:フラグ(0:打ち切り,1:イベント)
 - □ x: 群(Maintained:化学療法維持群, Nonmaintained:非維持群)

lik Mye	orary(sı DATA <-	urviva aml	al)				
MYE	DATA\$tin	ne2 <	- <u>Surv(MYDATA\$</u> 1	<u>time, My</u>	<u> (DATA\$status)</u>	time	status
MYE	DATA[ord	der(M)	YDATA\$time),]			5	1
	time st	tatus	Х	<u>time2</u> #	ἑ打ち切り		1
12	5	1	Nonmaintained	5 🛔	* には + がつく	5	1
13	5	1	Nonmaintained	5		8	1
14	8	1	Nonmaintained	8			
15	8	1	Nonmaintained	8		8	1
1	9	1	Maintained	9		9	1
16	12	1	Nonmaintained	12		12	1
2	13	1	Maintained	13		12	1
3	13	0	Maintained	13 <mark>+</mark>		13	1
17	16	0	Nonmaintained	16 <mark>+</mark>		13	0
4	18	1	Maintained	18		16	0

(例)生存時間解析(2)



■ カプラン・マイヤー法によるメディアン生存関数



(例)生存時間解析(3)



■ ログランク検定

```
survdiff(time2 ~ x, data=MYDATA)
```

Call: survdiff(formula = time2 ~ x, data = MYDATA)

	Ν	Observed	Expected	(0-E)^2/E	(0-E)^2/V
x=Maintained	11	7	10.69	1.27	3.40
x=Nonmaintained	12	11	7.31	1.86	3.40

Chisq= 3.4 on 1 degrees of freedom, p= 0.0653

有意水準5%で検定した場合は,維持群と非維持群の 間の生存時間の差は有意ではない

(例)生存時間解析(4)



■ コックス回帰

```
coxph(time2 ~ x, data=MYDATA)
```

```
Call:
coxph(formula = time2 ~ x, data = MYDATA)
```

```
        coef exp(coef) se(coef)
        z
        p

        xNonmaintained 0.916
        2.5
        0.512
        1.79
        0.074
```

Likelihood ratio test=3.38 on 1 df, p=0.0658 n= 23

```
もし,進行度データ stage のようなものを層別因子として(共変量として)分析する場合・・・
```

層(進行度)によって,異なるベースラインハザードを想定する場合など coxph(time2 ~ x + <u>strata(stage)</u>, data=MYDATA)



□ Surv(観察開始時間,観察終了時間,打ち切りフラグ)

- ―― 0:右側打ち切り
- └──── 1:イベント発生
- ----- 2:左側打ち切り
 - ---- 3:区間打ち切り





■データフレームについて(復習) □
手入力によるデータフレームの作成 □テキストファイルからのデータの読み込み □EXCELファイルからのデータの読み込み データ加T・データハンドリング □データへのアクセス,データの抽出・加工 ■データフレームを使った統計解析の例 □回帰分析 □ 生存時間解析

■次回は(あれば…)「検定関数について」

