

オープンソースの特徴

誰もがフリーで自由に同じ環境で統計解析を行うことができるため、同じデータを用いて同じ解析を行っているにも関わらず、(商用の)異なるソフトウェアの違いにより別々の出力結果を得るという事態を回避することができます。現在では、R Development Core Teamによりメンテナンスと拡張が行われています。RCRAN(<http://cran.r-project.org/>)から最新のリリース版(2015年1月15日現在では、2014年10月31日にリリースされたR-3.1.2が最新バージョンです)を入手することができます。さらに、商用のパッケージとは異なり、最近の論文で提唱された分析に関するパッケージも次々にRCRANにおいてアップロードされています。これは、商用のソフトウェアには無い特徴と言えます。

1 フリーソフトウェアRについて

2 Rを利用した実際の統計解析

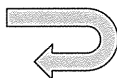
3 本日のまとめについて

データの読み込みについて

直接格納する方法

例えば、変数「data1」に1、2、3、4、5の5つのデータを格納するためには、以下のようなコマンドを書きます。

```
> data1<-c(1,2,3,4,5)
```

> |  Enterキーを押してください。

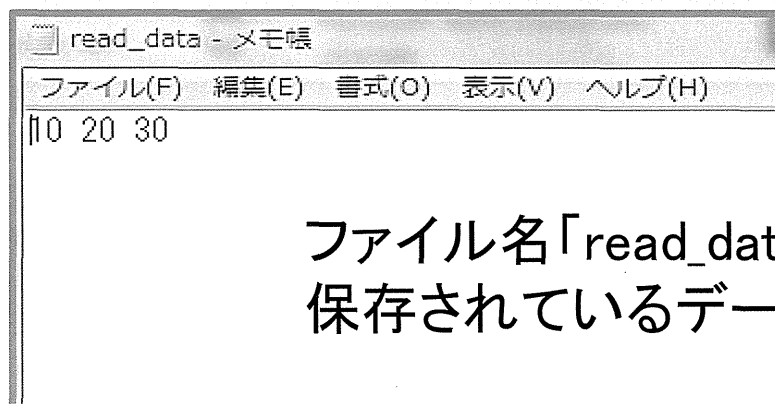
「data1」の中身を確認するために、「data1」と入力し、Enterキーを押してみてください。次のような出力を得ます。

```
> data1  
[1] 1 2 3 4 5
```

データの読み込みについて

ファイルからデータを格納する方法

デスクトップ上のファイルからもデータを読み込むことができます。



ファイル名「read_data.txt」に
保存されているデータ

```
> data2<-scan("C:/Users/ ユーザー名 /Desktop/read_data.txt")
Read 3 items
```



Enterキーを押してください。(ユーザー名の部分はご自身のPCのユーザー名になります。)

C:/Users/ ユーザー名 /Desktop は、データが保存されている場所を示しています。その他の場所に保存されているデータもこの部分を変更することでRに読み込むことができます。

実際に、Rに正しくデータが読み込まれているか確認してみましょう。

```
> data2
```



Enterキーを押してください。

```
[1] 10 20 30
```

Rのサンプルデータについて

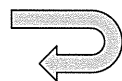
Rには、サンプル用のデータセットが多数用意されています。「MASS」と「stats」を読み込んでみましょう。

```
> library(MASS)
```



Enterキーを押してください。

```
> library(stats)
```



Enterキーを押してください。

```
> data()
```



Enterキーを押してください。

次のようにデータセットの一覧をご確認いただけます。

Data sets in package 'MASS':	
AirPassengers	Monthly Airline Passenger Numbers 1966-1967
BJsales	Sales Data with Leading Indicator
BJsales.lead (BJsales)	Sales Data with Leading Indicator
SCD	Bioclimatic Oxygen Demand
CO2	Carbon Dioxide Uptake in Brass Plants
ChickWeight	Weight versus age of chicks on different diets
UScars	Price versus age of cars
EUStockMarkets	Daily Closing Prices of Major European Stock Indices

データ「birthwt」について

例えば、「birthwt」というデータフレームがあります。これは、Springfield の Baystate 医療センターの 189 の出生について低体重出生とそのリスク因子の関連を調べるためのデータセットです。

```
> birthwt
```

	low	age	lwt	race	smoke	ptl	ht	ui	ftv	bwt
85	↑0	↑19	132	↑2	↑0	↑0	↑0	↑1	↑0	2523
↑ 低体重出生の有無を示す 2 値変数 (児の出生時体重 2.5kg 未満が 1)										
87	0	20	105	1	1	0	0	0	1	2557
↑ 年齢										
88	0	21	108	1	非熟練労働経験数			1	2	2594
89	0	1	最終月経時体重 (ポンド)			1	0	0	0	2600
↑ 高血圧の既往 (1=あり)										
91	0	2	人種 (1=白人, 2=黒人, 3=その他の有色人種)			0	0	0	0	2622
92	0	22	118	1	0	0	0	0	1	2637
↑ 子宮神経過敏の有無 (1=あり)										
93	0	17	103	喫煙の有無 (1=あり)		0	0	0	1	2637
94	0	29	123	1	1	0	妊娠の最初の 3 ヶ月の受診回数			-
95	0	26	113	1	1	0	0	0	児の出生時体重 (g)	

Rの関数を使ってみよう

最も基本的な関数について(その①)

記号	意味
sum()	総和
mean()	平均
median()	中央値
max()	最大値
min()	最小値
range()	範囲

「birthwt」の「年齢」にこれらの関数を適用してみましょう。

```
> sum(birthwt[,2])
```

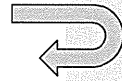
```
[1] 4392
```



Enterキーを押してください。

```
> mean(birthwt[,2])
```

```
[1] 23.2381
```



Enterキーを押してください。

```
> median(birthwt[,2])
```

```
[1] 23
```



Enterキーを押してください。

```
> max(birthwt[,2])
```

```
[1] 45
```



Enterキーを押してください。

```
> min(birthwt[,2])
```

```
[1] 14
```



Enterキーを押してください。

```
> range(birthwt[,2])
```

```
[1] 14 45
```



Enterキーを押してください。

Rの関数を使ってみよう

最も基本的な関数について(その②)

記号	意味
var()	不偏分散
sd()	不偏標準偏差
quantile()	クォンタイル点
summary()	要約統計量
scale()	標準化

「birthwt」の「最終月経時体重」にこれらの関数を適用してみましょう。

```
> var(birthwt[,3])
```

```
[1] 935.0985
```

```
> sd(birthwt[,3])
```

```
[1] 30.57938
```

```
> quantile(birthwt[,3])
```

```
0% 25% 50% 75% 100%
```

```
80 110 121 140 250
```

```
> summary(birthwt[,3])
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
80.0 110.0 121.0 129.8 140.0 250.0
```

```
> scale(birthwt[,3])
```

```
[,1]
```

```
[1,] 1.706548157
```

```
[2,] 0.823600246
```

```
[3,] -0.811488478
```

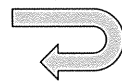
```
[4,] -0.713383154
```

```
[5,] -0.746084929
```

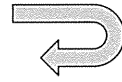
```
[6,] -0.190154763
```



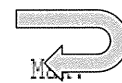
Enterキーを押してください。



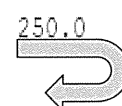
Enterキーを押してください。



Enterキーを押してください。



Enterキーを押してください。



Enterキーを押してください。

母平均の差の検定・推定 (対応のない場合)

Rを用いた2群の母平均の差の検定

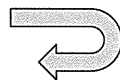
対応のない場合： 2群が互いに独立である場合

① どちらの群も正規分布に従うときは2標本のt検定を行います。

② 正規分布が仮定できない場合はWilcoxonの順位和検定を行います。

(データ「birthwt」において)2.5kg以上の児を出産した母親の「最終月経時体重」を「group1」に格納し、2.5kg未満の児を出産した母親の「最終月経時体重」を「group2」に格納しましょう。

```
> group1<-birthwt[,3][birthwt[,1]==0]
```



Enterキーを押してください。

```
> group2<-birthwt[,3][birthwt[,1]==1]
```



Enterキーを押してください。

① どちらの群も正規分布に従うときの2標本のt検定

＜両群の分散が未知であるが、等しい場合＞

```
t.test(x, y = NULL, alternative = c("two.sided",  
"less", "greater"), mu = 0, paired = FALSE,  
var.equal = TRUE, conf.level = 0.95, ...)
```

片側検定を行う場合は、こちらを指定します。

両側検定を行う場合は、こちらを指定します。

関数「t.test」を用いることで実行することができます。
ただし、引数を「var.equal = TRUE」と指定します。

```
> t.test(group1, group2, alternative=c("two.sided"), mu=0, paired=FALSE, var.equal=TRUE, conf.level=0.95)
```



Enterキーを押してください。

```
Two Sample t-test
      t値      自由度      p値
data:  group1 and group2
t = 2.3537, df = 187, p-value = 0.01962
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.807157 20.521656 ← ----- 上側信頼限界
sample estimates:
mean of x mean of y
133.3000 122.1356
                                ----- 下側信頼限界
```

group1の平均値

group2の平均値

① どちらの群も正規分布に従うときの2標本のt検定

＜両群の分散が未知であり、等しいかどうか分からない場合＞

```
t.test(x, y = NULL, alternative = c("two.sided",  
"less", "greater"), mu = 0, paired = FALSE,  
var.equal = FALSE, conf.level = 0.95, ...)
```

関数「t.test」を用いることで実行することができます。
ただし、引数を「var.equal = FALSE」と指定します。

```
> t.test(group1, group2, alternative=c("two.sided"), mu=0, paired=FALSE, var.equal=FALSE, conf.level=0.95)
```



Enterキーを押してください。

Welch Two Sample t-test

data: t値 group1 and group2 自由度 p値

t = 2.5155, df = 132.46 p-value = 0.01308

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

2.38552 19.94329 ← ----- 上側信頼限界

sample estimates:

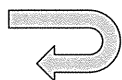
mean of x mean of y
133.3000 122.1356 ← ----- 下側信頼限界

group1の平均値

group2の平均値

(データ「birthwt」において)2.5kg以上の児を出産した母親の「年齢」を「group1.wil」に格納し、2.5kg未満の児を出産した母親の「年齢」を「group2.wil」に格納しましょう。

```
> group1.wil<-birthwt[,2][birthwt[,1]==0]
```



Enterキーを押してください。

```
> group2.wil<-birthwt[,2][birthwt[,1]==1]
```



Enterキーを押してください。

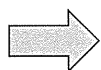
② 正規分布が仮定できない場合のWilcoxonの順位和検定

関数「wilcox.test」を用いることで実行することができます。

```
wilcox.test(x, y = NULL, alternative =  
c("two.sided", "less", "greater"), mu = 0,  
paired = FALSE, exact = NULL, correct = TRUE,  
conf.int = FALSE, conf.level = 0.95, ...)
```

信頼区間を計算するか否かを
指定します。

連続修正を行うか否かを指定
します。



引数についてさらに詳しく学びたい方は、`help(wilcox.test)`を実行すると参照いただけます。