
データの分析と知識発見
Introduction to Data Analysis

今回の構成

平均や分散などの代表値を知る

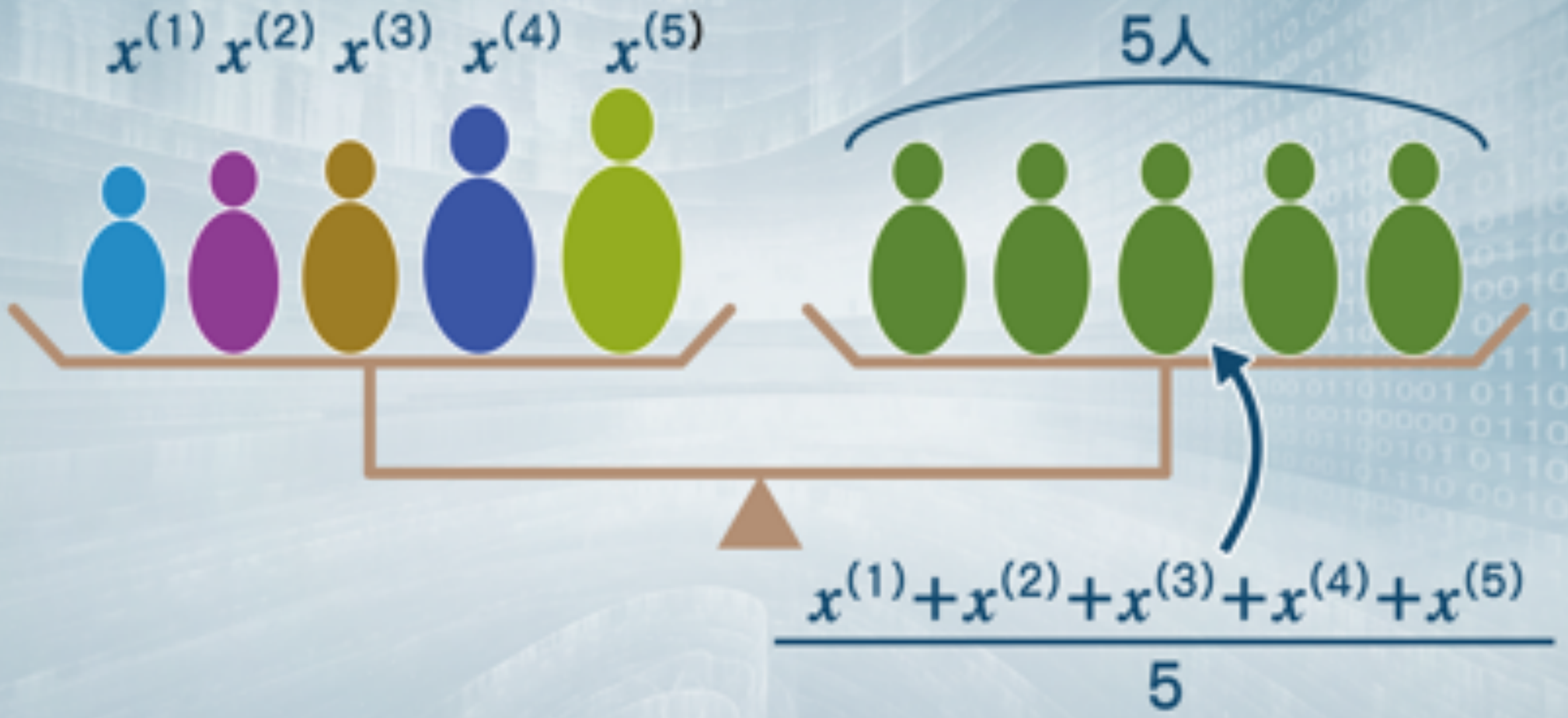
Rスクリプトを作成する

パッケージknitrを用いて
レポートを作成する

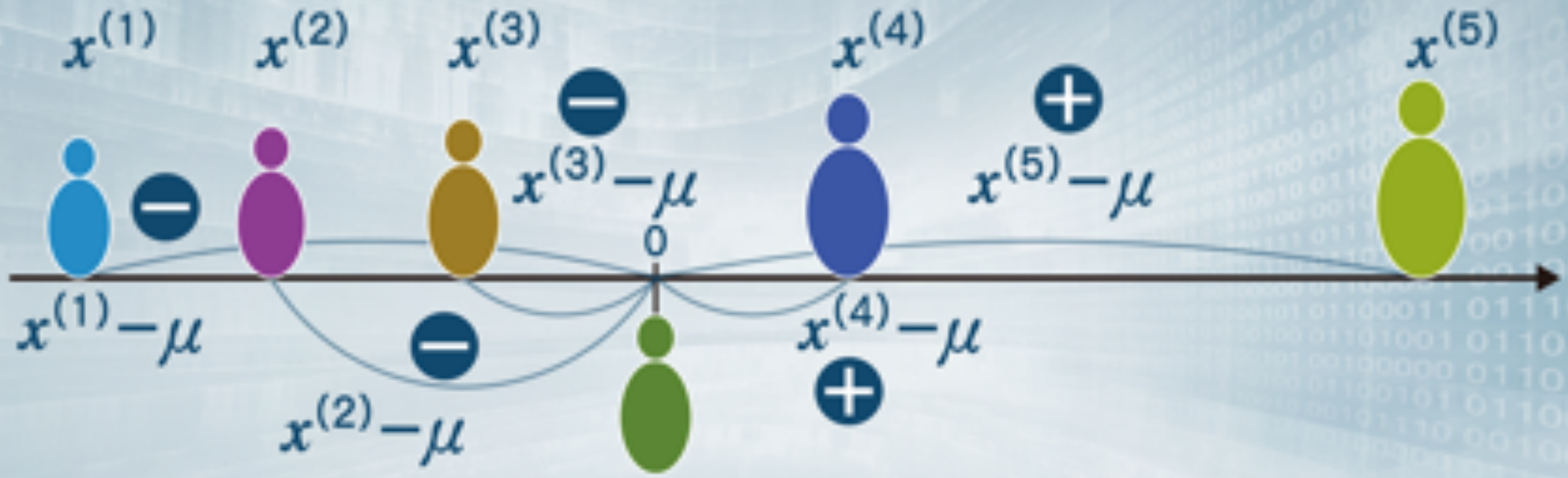
代表値

最大値	要素の中で最大の値
最小値	要素の中で最小の値
平均値	値の総和を要素の個数で割った値
最頻値	頻度の一番大きな値(階級値)
中央値	値を順に並べたときの中央の値

平均



平均



平均と分散

N個のデータの値：$x^{(1)}, x^{(2)}, \cdots, x^{(N)}$

平均

$$\mu = \frac{1}{N}\left(x^{(1)} + x^{(2)} + \cdots + x^{(N)}\right) = \frac{1}{N}\sum_{p=1}^{N} x^{(p)}$$

分散

$$\sigma^2 = \frac{1}{N}\sum_{p=1}^{N}(x^{(p)} - \mu)^2$$

不偏分散

$$\sigma^2 = \frac{1}{N-1}\sum_{p=1}^{N}(x^{(p)} - \mu)^2$$

平均と分散

5人の身長の値：

name	A	ave	dev	dev^2
(1)	118	130	−12	144
(2)	119	130	−11	121
(3)	121	130	−9	81
(4)	122	130	−8	64
(5)	170	130	40	1600
sum	650	650	0	2010
ave	130	130	0	502.5

平均

偏差を計算

不偏分散

5−1＝4 で割る

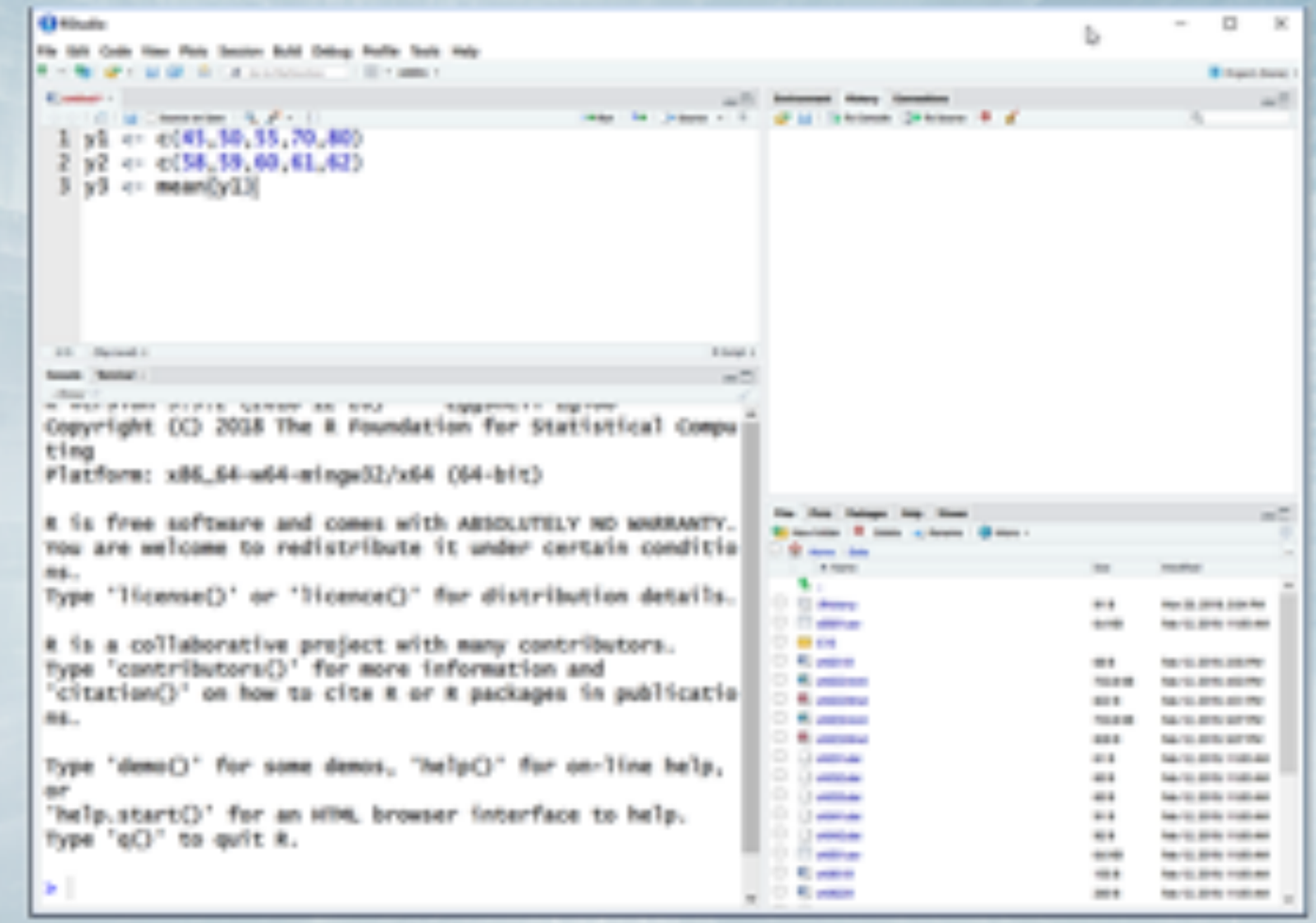
Rによる計算

```
> y1 <- c(45, 50, 55, 70, 80)
> y2 <- c(58, 59, 60, 61, 62)
```

結合（combine）

```
> mean(y1)
> var(y1)
> mean(y2)
> var(y2)
```

Rスクリプト


y1 <- c(41,50,55,70,80)
y2 <- c(58,59,60,61,62)
y3 <- mean(y1)
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Rスクリプト

```
y1 <- c(45,50,55,70,80)
y2 <- c(58,59,60,61,62)
y3 <- mean(y1)
```

```
#print(y3)
```

```
> source("ch02.R")
```

```
> y3
```

Rスクリプト

```
y1 <- c(45,50,55,70,80)
y2 <- c(58,59,60,61,62)
y3 <- mean(y1)
```

```
#print(y3)
```

#から始まる行はコメント

```
> source("ch02.R")
```

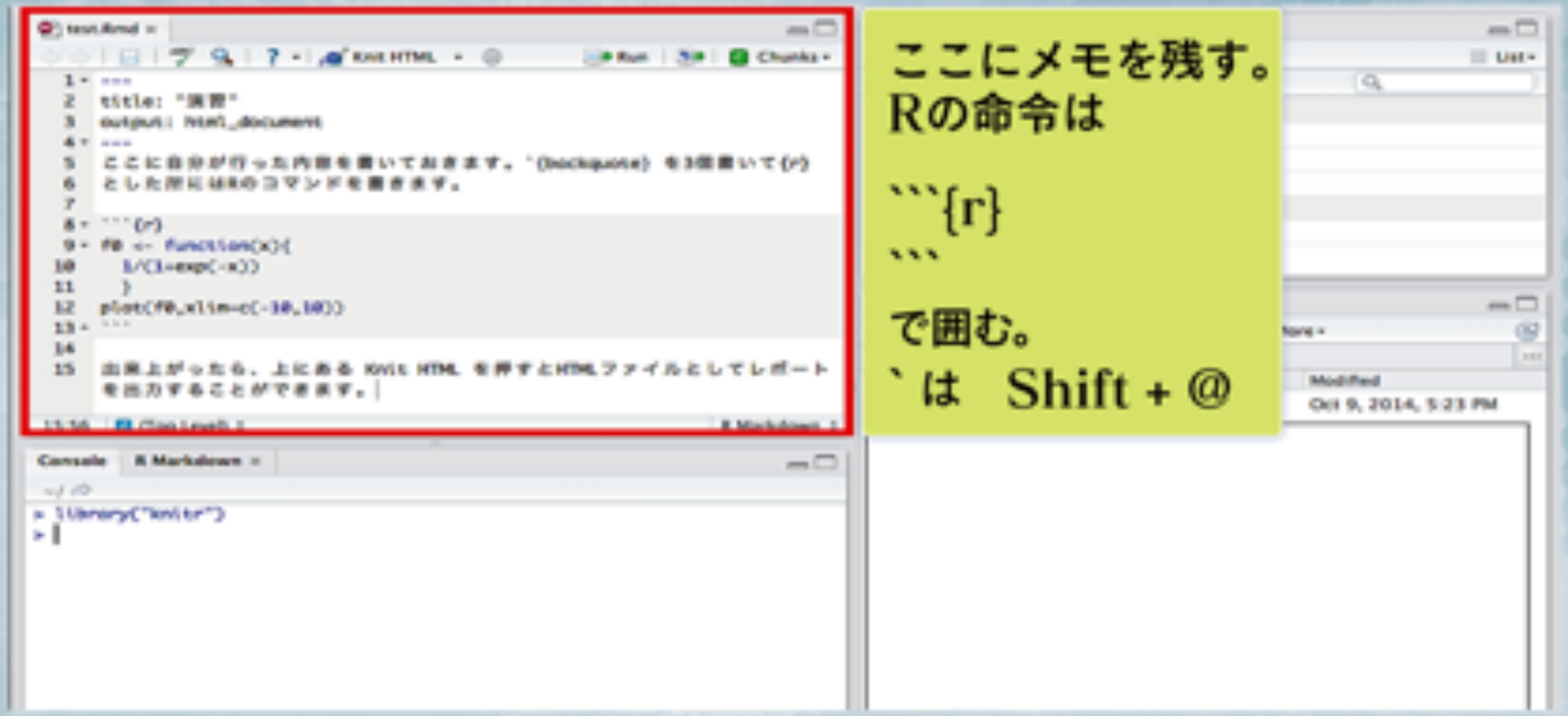
```
> y3
```

マークアップ言語←→WYSIWYG

例 HTML

```
<html>
<head>
<title>Test Page</title>
<meta http-equiv="Content-Type"
 content="text/html;charset=EUC-JP">
</head>
<body>
<h2>これはテストページです.</h2>
<br><br>
<img src="calc.jpg" width="320" height="240">
<br><br>
<a href="http://www.is.ouj.ac.jp/">他のページへ</a>
</body></html>
```

マークダウン



ここにメモを残す。
Rの命令は

```{r}
```

で囲む。
` は　Shift + @

Rにおける関数の例

統計量	説 明
sum	合計，　例 sum(a, b, c)
mean	平均値，中央値は median
max	最大値，最小値は min
var	分散, 行列を与えると分散共分散行列を計算する
sd	標準偏差, それぞれの列の標準偏差を求める
cor	相関, 行列を与えると相関行列を求める

Rにおける関数の例

関数名	説 明	例と意味
abs	絶対値	abs(a)
cos	三角関数	cos(a)　例 cos(0)＝1
sin	三角関数	sin(a)　例 sin(pi／2)＝1
tan	三角関数	tan(a)　例 tan(pi／4)＝1
round	四捨五入	round(a,n) nは小数点以下の桁数
log	自然対数	log(a)常用対数はlog10()
sqrt	平方根	sqrt(a)で$\sqrt{a}$を求める

関数の例(1)

```
f1 <- function(a){
b <- 4
a*b
}
```

```
> f1()
> f1(4)
```

関数の例(2)

```
f2 <- function(a=3) {
b <- 4
a*b
}
```

```
> f2()
> f2(4)
```