# Introduction to R

宇佐美慧
（東京大学）

Email: usami_s@p.u-tokyo.ac.jp
HP: http://www.satoshiusami.com/

# 資料のアウトライン

- なぜRか？

- Rを使ってみよう

(1)基本操作1　基本統計量，変数作成，関数 etc

(2)基本操作2　パッケージ，データの読み取り etc

(3)統計分析・図の作成

(4)シミュレーションの基礎

# なぜ、R?

- **Excel, SPSS, SASなど様々なソフトウェアがあるのに？**

**理由を簡単に言えば,**

- Rはフリー(無料、かつ自由な配布）。
- Windows, Unix, Macなど様々なプラットフォームで実行可能。
- 既存の豊富な統計解析が実行できるばかりでなく、最新の方法についても利用可能性が高い。
- グラフィックも充実。
- コードを走らせて大量の分析を自動処理化（クリックの繰り返しは面倒！）
- 参考資料(本・オンラインのユーザーサポート）も充実。
- プログラミング言語として，特定の計算処理を実行する関数の作成やシミュレーションの実行が可能。

# もう少し詳しく:

http://kohske.github.io/ESTRELA/201503/index.htmlより引用. 一部改変.

## R言語の最新の動向

- 2010年代に入って以降、「ビッグデータ」「データサイエンティスト」がバズワードとなり、官民問わずデータ分析の有用性が意識され始めています。R言語は当初は学術分野を中心に普及が進みましたが、最近では民間企業でも多くの専門家がR言語を駆使してビジネス上の意思決定を支えています。事実、米国での2014年の報告によると、R言語はSQLなどのデータベース技術、Hadoopなどの大規模データ処理環境構築スキルなどを抑え、IT関連技術の中で最も高いサラリーが期待できるスキルとされています。またプログラミング言語について成長度、ユーザ数、求職数などを総合的に評価したIEEE Spectrumの2014年のランキングではJava、C、Javascriptなどについで R言語が9位にランクインしています。R言語は今後しばらくはデータ分析ツールのデファクト・スタンダードであり続けるに違いありません。

## 豊富な統計解析手法の利用

- Rの強みは何と言っても、多数の統計解析手法が容易に利用できる形ですでに提供されていて、さらに最先端の統計解析手法も利用できるという点にあるでしょう。統計解析手法の多くはRのパッケージとして公開されてます。Rの公式パッケージシステムであるCRANには2015年1月現在、6000以上のパッケージが登録されています。この中には従来の統計解析手法を使いやすくするパッケージ、効果的な可視化を行うためのパッケージなども多数含まれています。

## 他のツールとの比較

- 2015年現在、他のツールと比較すると、速度面での極端に高いパフォーマンスや大規模データの処理が必要な状況以外では、利用できる統計解析手法の豊富さや信頼性、ユーザ数や情報の多さという点から、R言語に一日の長があるといえるでしょう。

## データ解析環境としてのRstudio （ただし，本資料では触れません）

- 2011年にβ版が公開されて以来、RStudioチーム（ http://www.rstudio.com/ ）はRコミュニティの中で急速に存在感を高めています。RStudioはR用IDE(統合開発環境)で、変数ビューアやデータビューア、グラフィック出力パネルの統合、Rの利用に役立つ様々な機能を実装しているだけでなく、プロジェクト管理、バージョン管理、レポート・スライド作成支援など、日々のデータ分析業務に役立つ機能も備えています。是非一度試してみることをお勧めします。

## 高速化・並列化・大規模データへの対応

- 「ビッグデータ」と呼ばれる超巨大データの登場で、現在ではデータ分析を取り巻く環境が大きく変化しています。R言語自体は大規模データを扱う能力に乏しいため、大規模なデータに対してR外部での高速・並列なデータ抽出、変換、読み込み（Extract/Transform/Load、ETL）環境の構築が不可欠ですが、最近ではR上でETL環境にアクセスできるパッケージが提供され始めています。

# Rの利用とサポート

**ダウンロード/インストール**：http://www.r-project.org/(2016年9月現在R-3.3.1)

**Rのユーザーサポート.**

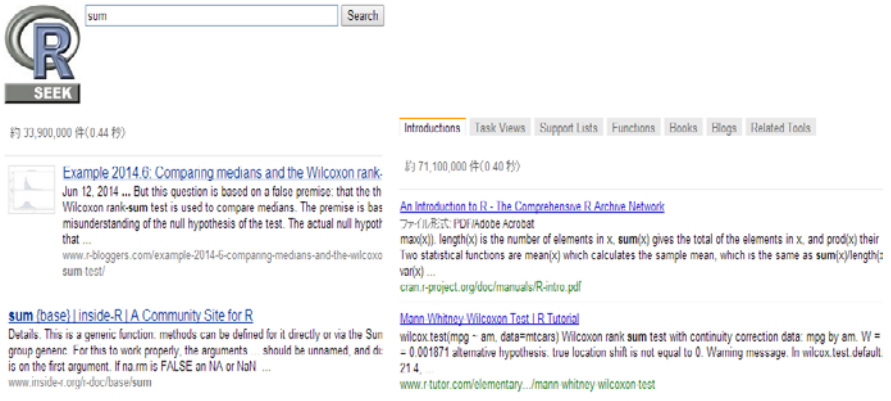
(a)Rseek:  http://www.rseek.org/

(b)オンラインマニュアル：https://cran.r-project.org/doc/manuals/R-intro.pdf
Venables, W.N., Smith, D.M., & R Core Team (2014). An introduction to R.

(c)RFAQ：http://cran.r-project.org/doc/FAQ/R-FAQ.html

(d)help function/example function;
e.g., help(sum)

(e)他，書籍・オンライン上の資料
（次ページおよび引用文献参照）。



# 役立つオンライン資料（一部）

(1)The R Project for Statistical Computing (英語)
http://www.r-project.org

(2) R Tips (日本語)
http://cse.naro.affrc.go.jp/takezawa/r-tips/r.html

(3)小杉考司・押江隆(2013). Rチュートリアルセミナーテキスト
http://psycho.edu.yamaguchi-u.ac.jp/wordpress/wp-content/uploads/2014/01/R_tutorial20131.pdf

(4)統計数理研究所　林真広先生　はじめての「R」(スライド)
http://www.slideshare.net/m884/japan-r-15432969

(5)群馬大学　青木繁伸先生のサイト
http://aoki2.si.gunma-u.ac.jp/R/

# 資料のアウトライン

- なぜRか？

- Rを使ってみよう

(1)基本操作1　基本統計量，変数作成，関数 etc

(2)基本操作2　パッケージ，データの読み取り etc

(3)統計分析・図の作成

(4)シミュレーションの基礎

# 使ってみましょう

**（私が思う）R言語学習の基本原則：**

- 関数・コマンドは，いきなり沢山覚えようとしない。少しずつ覚えていけますし，忘れても，既述の通りサポートツールが沢山あります。

- どんどん手を動かして，沢山覚えて（忘れる）の繰り返しでOK。

- 本資料は，Windows版のRをもとにした解説をしますが，今回の内容程度のことは基本的に他のOS版でも相違なく実行できます。

- また、統計分析は単に「使えれば（動けば）それで良し」というのは極力避けるべきで、どのような統計的処理がされて眼前のアウトプットが得られているのかも平行して学ぶ態度が常に重要。

→「いくら(ソフトで)分析ができても，統計的知識がなければいつかは詰みます」。

# 基本演算

四則演算+α

> 4+2

[1] 6

- 4-2　4*2　4/2　4^2　4**2　4/(2*2)　sqrt(4)

対数・指数

- log(10)（自然対数),
- log(10,10)（常用対数),
- exp(3) (指数)
- 関数は,「関数名(a)」や「関数名(a,b)」などの形をとる。a,bは**引数(argument/parameter)**と呼ぶ。
- 英数字は基本的に半角でタイプする。
- 大文字・小文字の区別に注意。

# 変数の定義

- X<-3　　# コメントの挿入にシャープ”#”が利用できる。

**注:変数名は半角数字で始まってはいけない。一部の記号も利用不可。**

- X
- [1] 3

- A<-2 ; B<-4 # 一行内に複数個の命令を書くときに”;”が利用できる。

文字変数による計算

- A+B
- [1] 6

定義の変更と計算結果

- X<-2 ;(A+B)^X

# 基本演算その2

## ベクトルの作成

- X<-c(1,2,3,4,5) # c(数字列)によりデータベクトルを作る。

  Y<-c(1,5,3,2,4)

## 基本統計量の計算

sum(X)　　mean(X)　　var(X)　　sd(X)　　cov(X,Y)
cor(X,Y)　　sum(1:5)　　X+Y　　X*Y　　X^2　　X+1
#1:5は1,2,3,4,5を指す。

## 注意点

var(X) は不偏分散なので，標本分散を計算するときは，例えば，var(X)*((length(X)-1)/length(X))とする必要がある（後述）.

# 乱数の発生

- rnorm(1)
- rnorm(10)
- rnorm(10,mean=0,sd=1)　#3つ目の引数は分散ではないので注意。
- rnorm(10,0,1)
- runif (10,0,1)

library(MASS) #library,matrixについては後述

- mvrnorm(100,c(0,0),matrix(c(1,0.5,0.5,1),2,2))

#各変数の平均が0で分散が1,相関が0.5の2変数乱数を100個発生。

# その他役立つ関数

- seq(1,10,by=1)　#数列の生成。byは間隔。
- seq(from=1,to=10,by=3)
- rep(1,5)  #1つ目の引数に示されている系列の複数発生。
- X<-c(1,2,3); rep(X,5)
- floor(3.24) #整数打ち切り。
- round(3.24,1) #四捨五入。二つ目の引数は桁数。

# 補足:表示の注意

**(1)" e-8"などの表示。**

```
x<-rnorm(100)  #乱数の発生
z<-(x-mean(x))/sd(x)  #標準(z)得点の算出
mean(z)  #平均(0になるはず)の計算
[1] -2.751705e-18
```

→e-18は10の-18乗の意味→限りなく0に近い。

**(2)"Inf"などの表示。**

```
> 10^500
[1] Inf
```

→" Inf" (or "-Inf") は (負の)無限大を表す。

# 閑話休題：Rの特徴とRStudio

小杉・押江(2013). Rチュートリアルセミナーより引用。一部改変。

http://psycho.edu.yamaguchi-u.ac.jp/wordpress/wp-content/uploads/2014/01/R_tutorial20131.pdf

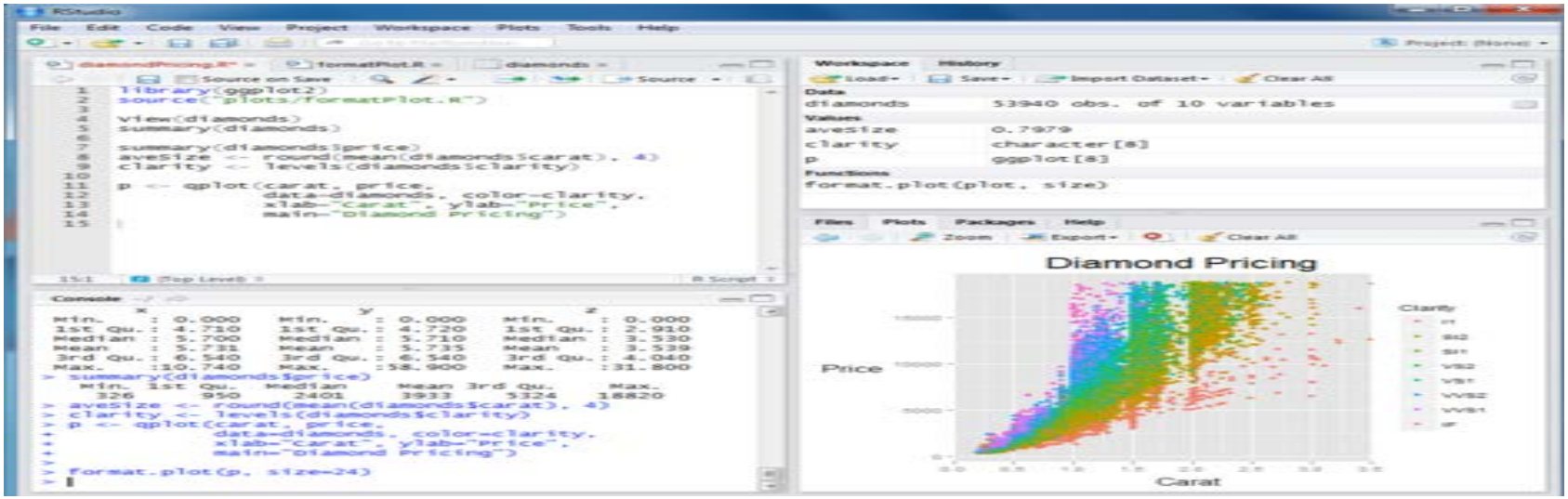
- RはCUI(キャラクタ・ユーザ・インターフェイス，あるいはコマンド・ユーザ・インターフェイス) であり，GUI(グラフィカル・ユーザ・インター フェイス）ではない。GUIは，ポインタがボタンを押すと分析などが実行される，という画面を持っていることを指します。

*宇佐美注：RにもGUI機能をもつパッケージとしてR コマンダー(Rcmdr)がありますが，大量の反復計算には向いていないためオススメしません。

- コードを書いて実行するときに感じる不便のひとつが，どこにコードを記録しておこうか迷う，というものです。R にはエディタもありますが，メモ帳などでコードを書いて，コピー・ペーストで実行し，うまくいったものだけ残して 保存する，という方法を取っている方もいるでしょう。コードを保存しておいても，複数のデータを扱ったり，同じデータでも異なる分析をする（異なるコードを書く）こ とがあると，混乱してしまうことがあるかもしれません。

# 続き (RStudio)

- そこでご紹介するのが RStudio です。RStudio は統合開発環境とよばれ，R と R にまつわるファイル，関数，変数，パッケージ，図版などをこのソフトだけで管理することができるのです。例えて言うなら，R は飯ごう炊爨のようにー つーつ手作りで進んでいく楽しさがありますが，RStudio はキッチンスタジアム，とまではいかないまでも，システム キッチンでお料理する程度には整えられた環境なのだ，とイメージしていただければと思います。




宇佐美注：本資料はRstudioは直接触れませんが，上達度や好みに応じてRstudioに切り替えていくのも手でしょう。

# 資料のアウトライン

- なぜRか？

- Rを使ってみよう

(1)基本操作1　基本統計量，変数作成，関数 etc

(2)基本操作2　パッケージ，データの読み取り etc

(3)統計分析・図の作成

(4)シミュレーションの基礎

# （本題に戻って、）欠測値がある場合

X<-c(1,2,3,4,NA); Y<-c(3,4,5,1,2)

**一変数の平均**

```
> mean(X)
[1] NA
> mean(X,na.rm=T)
[1] 2.5
```

**二変数間の相関**

```
> cor(X,Y)
[1] NA
> cor(X,Y,na.method<- "complete.obs")　#欠測のないケースを使う対処。
[1] -0.3779645
“”　" "
```

- データの欠測に関する基礎的事項については宇佐美・荘島(2015)を参照。

# ベクトルの要素の抽出・処理

[要素番号]を使う。

```
X<-c(5,4,2,3,1)
> X[3]
[1] 2

> X[2:4]
[1] 4 2 3

> X[c(3,5)]
[1] 2 1
```

# ビルト・インデータセットの利用

data()

data(package = .packages(all.available = TRUE))

```
Data sets in package 'datasets':

AirPassengers                           Monthly Airline Passenger Numbers 1949-1960
BJsales                                 Sales Data with Leading Indicator
BJsales.lead (BJsales)                  Sales Data with Leading Indicator
BOD                                     Biochemical Oxygen Demand
CO2                                     Carbon Dioxide Uptake in Grass Plants
ChickWeight                             Weight versus age of chicks on different diets
DNase                                   Elisa assay of DNase
EuStockMarkets                          Daily Closing Prices of Major European Stock Indices, 1991-1998
Formaldehyde                            Determination of Formaldehyde
HairEyeColor                            Hair and Eye Color of Statistics Students
Harman23.cor                            Harman Example 2.3
Harman74.cor                            Harman Example 7.4
```

iris　#(フィッシャーも判別分析で使った，有名な）あやめのデータ

| | Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species |
|---|---|---|---|---|---|
| 1 | 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| 2 | 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| 3 | 4.7 | 3.2 | 1.3 | 0.2 | setosa |
| 4 | 4.6 | 3.1 | 1.5 | 0.2 | setosa |
| 5 | 5.0 | 3.6 | 1.4 | 0.2 | setosa |
| 6 | 5.4 | 3.9 | 1.7 | 0.4 | setosa |
| 7 | 4.6 | 3.4 | 1.4 | 0.3 | setosa |
| 8 | 5.0 | 3.4 | 1.5 | 0.2 | setosa |

・Sepal「がく片」の長さと幅・Petal「花びら」の長さと幅。
・Species「あやめ３品種　[setosa・versicolor・virginica]」

# (データ)行列の要素の抽出・処理

iris[1,2]　iris[,2]　iris[2,]　iris[,-2]

iris$Sepal.Length　#「$+変数名」による読み込み

colMeans(iris[,1:4])

rowMeans(iris[,1:4])

rowSums(iris[,1:4])

cor(iris[,1:4])

round(cor(iris[,1:4]),3)

# パッケージの活用

- メモリの軽量化や同一名義による関数のcrashを避けるため，一部の関数については，それが格納されているパッケージを読み込まないといけない。

例えば…

> describe

エラー：オブジェクト ‘describe’ がありません

install.packages("psych")

```
> install.packages("psych")
 パッケージを 'C:/Users/satoshi/Documents/R/win-library/2.15' 中にインストールします
 ('lib' が指定されていないので)
--- このセッションで使うために、CRANのミラーサイトを選んでください ---
 URL 'http://cran.ism.ac.jp/bin/windows/contrib/2.15/psych_1.4.3.zip' を試しています
Content type 'application/zip' length 2907097 bytes (2.8 Mb)
 開かれた URL
downloaded 2.8 Mb
```

一度インストールすれば以後不要。

Rを一度閉じたら再度読み込む必要。

library(psych) #ライブラリにある，インストールしたpsychの読み込み

エラーが出てインストールできない場合の対処

*ユーザー名が日本語だと×な場合も（その場合、管理者としてログインして修正する方法があります）.R自体を管理者として実行(右クリック）する方法も。
*回線の種類によっては×な場合があります。学内で×な場合，自宅の回線から試みれば解決する場合が多いです。また，日本ではなく他の国(Singerporeなど)をしているとセキュリティの問題が回避できるかもしれません。
*PCとの相性でRのversionが3だとどうしてもうまくいかないことがあるため，version2を使ってみてください。
*どうしても×な場合直接（送付した）psychのファイルをディレクトリ（windowsの場合：document/R/win-library）内に保存する方法があります）。

# Psychパッケージは役立ちます！

describe(iris) #基本統計量の計算

```
> describe(iris)
             vars   n mean   sd median trimmed  mad min max range  skew kurtosis   se
Sepal.Length    1 150 5.84 0.83   5.80    5.81 1.04 4.3 7.9   3.6  0.31    -0.61 0.07
Sepal.Width     2 150 3.06 0.44   3.00    3.04 0.44 2.0 4.4   2.4  0.31     0.14 0.04
Petal.Length    3 150 3.76 1.77   4.35    3.76 1.85 1.0 6.9   5.9 -0.27    -1.42 0.14
Petal.Width     4 150 1.20 0.76   1.30    1.18 1.04 0.1 2.5   2.4 -0.10    -1.36 0.06
Species*        5 150 2.00 0.82   2.00    2.00 1.48 1.0 3.0   2.0  0.00    -1.52 0.07
```

alpha(iris)　#α係数の計算 ( help(alpha)も役立つ。)

```
Reliability analysis
Call: alpha(x = iris)

  raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd
       0.81      0.85    0.94      0.59 5.9 0.055  3.9 0.86

 lower alpha upper     95% confidence boundaries
0.71 0.81 0.92

 Reliability if an item is dropped:
             raw_alpha std.alpha G6(smc) average_r  S/N alpha se
Sepal.Length      0.71      0.81    0.86      0.59  4.2     0.077
Sepal.Width-      0.88      0.96    0.96      0.88 22.9     0.064
Petal.Length      0.72      0.70    0.77      0.43  2.3     0.075
Petal.Width       0.68      0.73    0.86      0.47  2.7     0.079
```

Versionによって計算結果が変わる可能性？

fa()による因子分析の実行(後述)

# データの保存と読み取り

ここでは，一般的なCSV(Comma separated values) Fileに則って説明。

## データの保存(write.csv)

write.csv(iris, "C:/Users/satoshi/Desktop/PUBLIC2/irisdata.csv")

**作業ディレクトリを指定する必要有！また「irisdata」はファイル名であり，名前のつけ方は任意。**" "を忘れない(“” ではない！)。スラッシュは/の向き。.csvも忘れない。　Macの場合はC:（ドライブ名）を消せばよい。

日本語のファイル名（例えば,協調性発達.csvという名前）を読み込んでエラーが出る場合、
read.csv(file(“C:/Users/Satoshi Usami/Desktop/PUBLIC2/協調性発達.csv”,encoding=‘Shift_JIS’))とするとよい。

## データの読み取り(read.csv)

＃私のPCの場合

read.csv("C:/Users/satoshi/Desktop/PUBLIC2/irisdata.csv")

なお，作業ディレクトリ(wd)の指定が面倒ならば，

setwd("C:/Users/satoshi/Desktop/PUBLIC2")

read.csv("irisdata.csv")

# データの部分抽出

**subset関数**

subset(iris, iris$Species=="virginica")

注：2つの目の引数が抽出条件。等号(==)や不等号を使う。""は文字変数のときに利用する。

複数条件の場合：

subset(iris, iris$Species== "virginica" & Sepal.Width>3.0)　# &=かつ

subset(iris, iris$Species== "virginica" | Sepal.Width>3.0)　# |=または

subset(iris ,select=c("Species","Sepal.Length"))　#条件指定しない方法

**補足：attach関数**

> Species

エラー：　オブジェクト ‘Species’ がありません

attach(iris)

Species

detach(iris)

# その他役立つ関数（その2）

- cbind(iris[,c(1,3)],iris[,c(2,4)])　#列方向の結合。行方向はrbind.
- nrow(iris); ncol(iris)　#行数と列数
- head(iris) #最初の数行のデータを表示
- DATA<-matrix(1:15,3,5) #行列の作成
- t(DATA) #データ行列の転置(入れ替え)
- scale(iris[,1:4])　#データの標準化

# Exercises

1, Calculate basic statistics of variable X=(1,3,5,7,9,11,13) including sample mean, variance, standard deviation.

2, Let X=(1,3,5,7,9) and Y=(3,5,3,5,7) be two variables. Calculate correlation between X and Y with rounding to two decimal place (e.g., 0.72763… as 0.73).

3. Express data of X=(1,2,3,1,2,3,1,2,3,1,2,3,1,2,3,1,2,3,1,2,3) by using rep function.

4, Let X=(1,2,3,4,NA,5,8,9,11) be data including missing. Make new variable that does not include missing in X.

5. Confirm data named “lsat7” exists that includes 5 variables with size 1000 from data(), then calculate correlations of variables and alpha coefficient (reliability) of this data. Here results should be shown within three decimal points.

# 資料のアウトライン

- なぜRか？

- Rを使ってみよう

(1)基本操作1　基本統計量，変数作成，関数 etc

(2)基本操作2　パッケージ，データの読み取り etc

(3)統計分析・図の作成

(4)シミュレーションの基礎

# 統計分析（*t*検定）

```
DATAsetosa<-subset(iris,iris$Species=="setosa")
DATAversicolor<-subset(iris,iris$Species=="versicolor")
```

```
# t test　　がくの長さの差の検定
t.test(DATAsetosa$Sepal.Length, DATAversicolor$Sepal.Length,var.equal=T)
#Welch's test　がくの長さの差の検定
t.test(DATAsetosa$Sepal.Length, DATAversicolor$Sepal.Length,var.equal=F)
```

```
> t.test(DATAsetosa$Sepal.Length, DATAversicolor$Sepal.Length,var.equal=T)

        Two Sample t-test

data:  DATAsetosa$Sepal.Length and DATAversicolor$Sepal.Length
t = -10.521, df = 98, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.1054165 -0.7545835
sample estimates:
mean of x mean of y
    5.006     5.936
```

# 統計分析（相関係数の検定）

```
cor.test(DATAsetosa$Sepal.Length, DATAsetosa$Petal.Length)
```

#Setosaにおける，がく片と花弁の長さの相関

```
        Pearson's product-moment correlation

data:  DATAsetosa$Sepal.Length and DATAsetosa$Petal.Length
t = 1.9209, df = 48, p-value = 0.0607
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.01206954  0.50776233
sample estimates:
      cor
0.2671758
```

# 統計分析(回帰分析）

単回帰分析(lm(y～x))

```
lm(iris$Sepal.Length~iris$Petal.Length)
```

summary()などによる，より詳細な情報の抽出。

```
summary(lm(iris$Sepal.Length~iris$Petal.Length))
summary(lm(iris$Sepal.Length~iris$Petal.Length))[[4]]
AIC(lm(iris$Sepal.Length~iris$Petal.Length))#情報量規準
```

重回帰分析 (lm(y～x1+x2))

```
summary(lm(iris$Sepal.Length~iris$Petal.Length+iris$Petal.Width))
```

# 統計分析(一般化線形モデル)

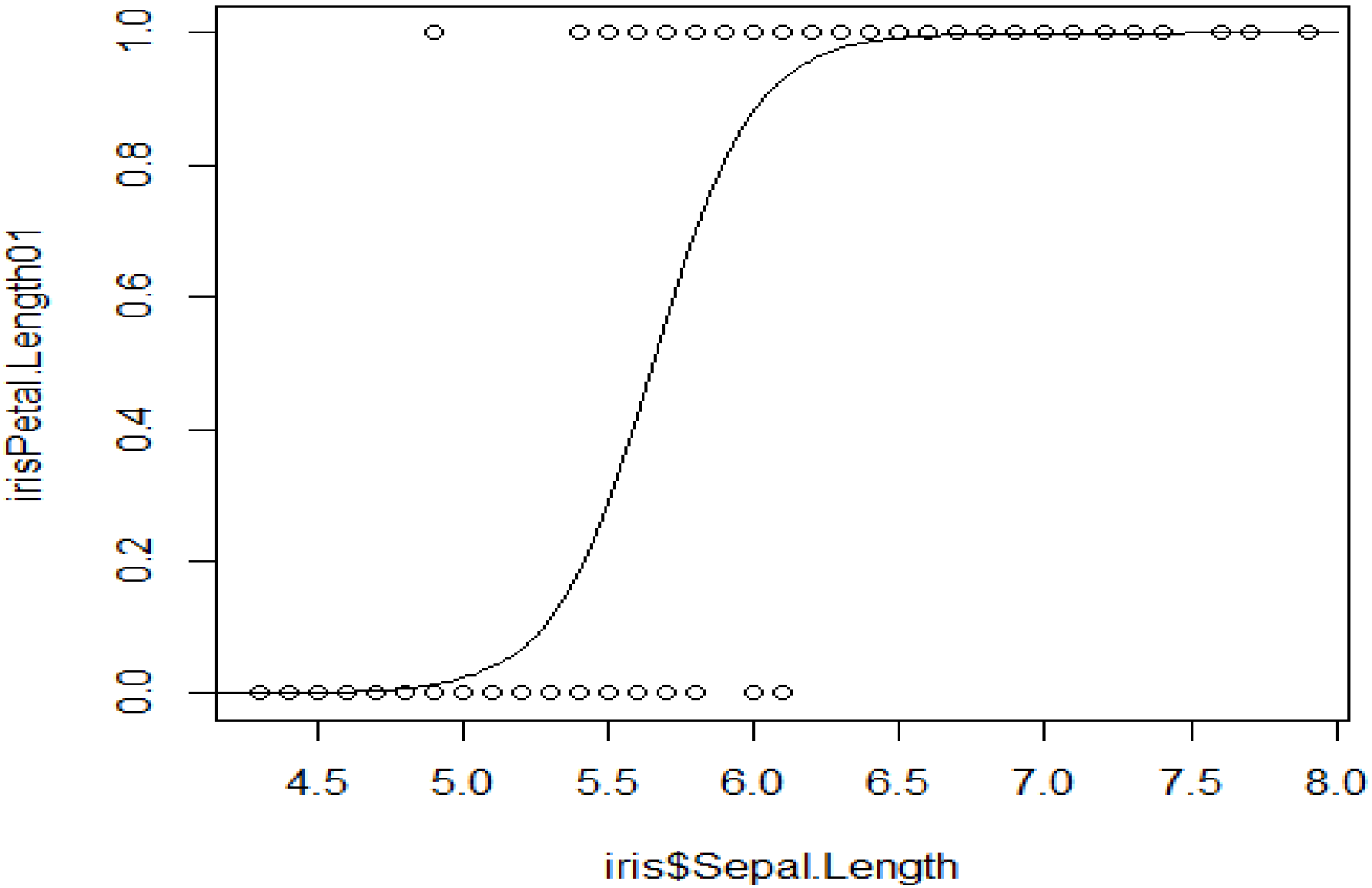
#0,1データの作成　(ifelse関数はデータ変換に役立つ関数)
irisPetal.Length01<-ifelse(iris$Petal.Length>4,1,0)

## ロジスティック回帰分析

```
glm(irisPetal.Length01~iris$Sepal.Length, family=binomial (link="logit"))
summary(glm(irisPetal.Length01~iris$Sepal.Length, family=binomial(link="logit")))
```

## データと予測式のプロット（予習）

```
result<- summary(glm(irisPetal.Length01~iris$Sepal.Length,
    family=binomial(link="logit")))
Esintercept<-result[[12]][1,1]; Esreg<-result[[12]][2,1]#回帰式の切片と回帰係数
xaxis<-seq(from=4, to=8, length=200)　#x軸
yaxis<-exp(Esintercept+Esreg*xaxis)/(1+exp(Esintercept+ Esreg* xaxis))　#y軸
plot(iris$Sepal.Length,irisPetal.Length01) #データプロット
lines(xaxis, yaxis) #生成データで曲線を描画する
```


irisPetal.Length01
0.0
0.2
0.4
0.6
0.8
1.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
iris$Sepal.Length

# 統計分析(分散分析）

```
aov(iris$Sepal.Length ~ factor(iris$Species))
```

```
summary(aov(iris$Sepal.Length ~ factor(iris$Species)))
```

```
                      Df Sum Sq Mean Sq F value Pr(>F)
factor(iris$Species)   2  63.21  31.606   119.3 <2e-16 ***
Residuals            147  38.96   0.265
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ただし，下位検定や効果量の算出などもう少し細かな計算上のニーズがある場合，他のソフトを併用することもある。たとえば，SPSS, ANOVA4(http://www.hju.ac.jp/~kiriki/anova4/),

**ANOVA君(http://riseki.php.xdomain.jp/index.php)→Rで実行可！**

# anovakunによる分散分析

*回線の種類によっては×な場合があります。学内で×な場合，自宅の回線から試みれば解決する場合が多いです。

```
source("http://riseki.php.xdomain.jp/index.php?plugin=attach&refer=ANOVA%E5%90%9B&openfile=anovakun_480.txt")
subdat <- subset(iris ,select=c("Species","Sepal.Length"))
anyNA = function(x) {any(is.na(x))} #ver3.1以前だとうまく動かないため必要
anovakun(subdat , "As",3,peta=T)    #データ、要因計画，水準数を書く。
```

Ａｓ(１要因被験者間)
ｓＡ(１要因被験者内)
ＡＢｓ(２要因被験者間)
ＡｓＢ(２要因混合)
ｓＡＢ(２要因被験者内)
ＡＢＣｓ(３要因被験者間)
ＡＢｓＣ(３要因混合)
ＡｓＢＣ(３要因混合)
ｓＡＢＣ(３要因被験者内)

```
[ As-Type Design ]

This output was generated by anovakun 4.8.0 under R version 2.15.0.
It was executed on Thu Sep 08 13:16:41 2016.


<< DESCRIPTIVE STATISTICS >>

--------------------------
  A   n    Mean    S.D.
--------------------------
 a1  50  5.0060  0.3525
 a2  50  5.9360  0.5162
 a3  50  6.5880  0.6359
--------------------------


<< ANOVA TABLE >>

-----------------------------------------------------------------
 Source       SS  df      MS  F-ratio  p-value      p.eta^2
-----------------------------------------------------------------
      A  63.2121   2 31.6061 119.2645   0.0000 ***    0.6187
  Error  38.9562 147  0.2650
-----------------------------------------------------------------
  Total 102.1683 149  0.6857
                +p < .10, *p < .05, **p < .01, ***p < .001
```

# 統計分析(因子分析）

#人工データを発生(復習)

DATA<-mvrnorm(500,rep(0,5),0.5+0.5*diag(5)) #diag()は対角行列。

#相関行列の確認

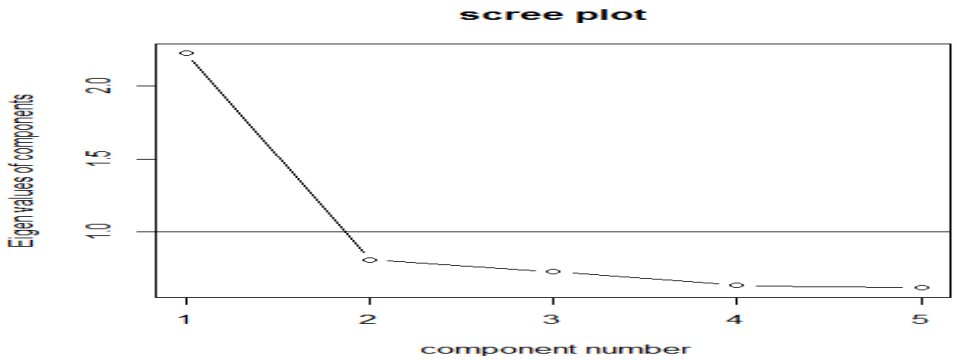
cor(DATA)

#因子分析の実行　（GPArotationパッケージをインストールする必要)

```
FA1<-fa(DATA,nfactors=1,fm="ml",rotate="promax",scores=TRUE)
FA2<-fa(DATA,nfactors=2,fm="ml",rotate="promax",scores=TRUE)
```

#スクリーテスト

VSS.scree(DATA)



# print()による出力

- print(FA2 ,sort=TRUE ,digit =3,cut=.6)

```
Standardized loadings (pattern matrix) based upon correlation matrix
  V   ML1    h2    u2 com
3 3 0.628 0.394 0.606   1
2 2 0.598 0.358 0.642   1
4 4 0.525 0.276 0.724   1
1 1 0.524 0.275 0.725   1
5 5 0.490 0.240 0.760   1

                 ML1
SS loadings    1.542
Proportion Var 0.308

Mean item complexity =  1
Test of the hypothesis that 1 factor is sufficient.

The degrees of freedom for the null model are  10  and the objective function was  0.688 with Chi Square of  341.4
The degrees of freedom for the model are 5  and the objective function was  0.014

The root mean square of the residuals (RMSR) is  0.027
The df corrected root mean square of the residuals is  0.038

The harmonic number of observations is  500 with the empirical chi square  7.085  with prob <  0.214
The total number of observations was  500  with MLE Chi Square =  7.004  with prob <  0.22

Tucker Lewis Index of factoring reliability =  0.9879
RMSEA index =  0.0287  and the 90 % confidence intervals are  NA 0.0727
BIC =  -24.069
Fit based upon off diagonal values = 0.993
Measures of factor score adequacy
                                                ML1
Correlation of scores with factors             0.834
Multiple R square of scores with factors       0.695
Minimum correlation of possible factor scores  0.391
```

# 統計分析(構造方程式モデリング)

*例示のため，伊藤・宇佐美(2016,教育心理学研究)から一部のデータを利用。

学級風土質問紙(5件法)

A…学級活動への関与(7項目)

(e.g.,行事へ一生懸命取り組む)

B…学級への満足感(5項目)

(e.g.,クラスを心から楽しむ)

C…学級内の不和(6項目)

(e.g.,クラスでもめ事が少ない)

E…生徒間の親しさ(6項目)

(e.g.,友達同士，助け合う)

| 1 | year | A1 | A2 | A3 | A4 | A5 | A6 | A7 | B1 | B3 | B5 | B6 | B7 | C1 | C2 | C3 | C4 | C6 | C7 | E2 | E3 | E5 | E6 | E7 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 97 | 2 | 3 | 3 | 4 | 2 | 3 | 3 | 3 | 3 | 1 | 3 | 4 | 2 | 3 | 1 | 5 | 3 | 5 | 2 | 3 | 4 | 3 | 3 |
| 3 | 97 | 4 | 3 | 2 | 3 | 3 | 4 | 5 | 4 | 3 | 3 | 3 | 4 | 2 | 3 | 1 | 3 | 3 | 3 | 3 | 4 | 4 | 3 | 4 |
| 4 | 97 | 4 | 3 | 1 | 4 | 3 | 3 | 5 | 3 | 3 | 3 | 3 | 5 | 1 | 4 | 1 | 5 | 3 | 5 | 4 | 4 | 3 | 3 | 3 |
| 5 | 97 | 4 | 1 | 2 | 2 | 4 | 4 | 4 | 5 | 2 | 4 | 2 | 5 | 2 | 2 | 1 | 4 | 1 | 2 | 4 | 4 | 5 | 2 | 4 |
| 6 | 97 | 3 | 5 | 1 | 3 | 2 | 3 | 2 | 4 | 2 | 1 | 3 | 4 | 4 | 2 | 1 | 3 | 1 | 4 | 1 | 2 | 4 | 2 | 2 |
| 7 | 97 | 2 | 1 | 1 | 5 | 3 | 1 | 4 | 1 | 1 | 1 | 1 | 3 | 1 | 5 | 1 | 3 | 5 | 1 | 1 | 3 | 3 | 1 | 3 |
| 8 | 97 | 3 | 3 | 1 | 3 | 2 | 3 | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 4 | 2 | 4 | 4 | 5 | 3 | 3 | 3 | 1 | 2 |
| 9 | 97 | 2 | 1 | 1 | 3 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 3 | 2 | 4 | 1 | 5 | 5 | 3 | 4 | 3 | 5 | 1 | 1 |
| 10 | 97 | 4 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 5 | 4 | 3 | 3 | 3 | 1 | 3 | 4 | 4 | 3 | 1 | 4 |
| 11 | 97 | 4 | 4 | 4 | 3 | 3 | 3 | 4 | 5 | 4 | 3 | 3 | 4 | 3 | 3 | 3 | 3 | 2 | 3 | 4 | 4 | 4 | 3 | 4 |
| 12 | 97 | 4 | 3 | 3 | 4 | 3 | 4 | 3 | 4 | 4 | 4 | 5 | 5 | 2 | 3 | 2 | 3 | 4 | 3 | 2 | 4 | 4 | 3 | 3 |
| 13 | 97 | 3 | 2 | 1 | 1 | 1 | 2 | 2 | 3 | 1 | 2 | 2 | 4 | 1 | 2 | 1 | 4 | 2 | 4 | 1 | 1 | 2 | 2 | 2 |
| 14 | 97 | 4 | 2 | 4 | 2 | 2 | 3 | 2 | 2 | 4 | 4 | 4 | 4 | 3 | 4 | 1 | 3 | 4 | 4 | 2 | 2 | 3 | 4 | 4 |
| 15 | 97 | 3 | 5 | 3 | 5 | 1 | 3 | 5 | 3 | 3 | 1 | 1 | 5 | 2 | 3 | 2 | 5 | 1 | 5 | 3 | 2 | 5 | 1 | 3 |
| 16 | 97 | 5 | 3 | 3 | 4 | 4 | 4 | 3 | 4 | 5 | 4 | 4 | 4 | 3 | 1 | 3 | 5 | 3 | 4 | 5 | 5 | 5 | 3 | 3 |
| 17 | 97 | 2 | 2 | 1 | 3 | 3 | 1 | 1 | 2 | 3 | 3 | 1 | 1 | 2 | 3 | 1 | 3 | 1 | 5 | 1 | 1 | 3 | 2 | 1 |
| 18 | 97 | 2 | 1 | 1 | 1 | 3 | 1 | 2 | 1 | 2 | 3 | 1 | 3 | 1 | 5 | 1 | 5 | 5 | 5 | 1 | 3 | 1 | 1 | 1 |
| 19 | 97 | 4 | 2 | 1 | 2 | 2 | 2 | 3 | 3 | 2 | 1 | 2 | 2 | 2 | 3 | 1 | 5 | 4 | 5 | 2 | 3 | 2 | 2 | 3 |
| 20 | 97 | 3 | 5 | 4 | 5 | 3 | 3 | 5 | 3 | 3 | 1 | 2 | 5 | 1 | 1 | 1 | 5 | 5 | 5 | 1 | 4 | 2 | 1 | 4 |
| 21 | 97 | 5 | 4 | 3 | 5 | 4 | 4 | 4 | 1 | 2 | 1 | 1 | 3 | 1 | 5 | 1 | 5 | 4 | 5 | 3 | 1 | 2 | 2 | 4 |
| 22 | 97 | 5 | 3 | 4 | 3 | 4 | 4 | 3 | 4 | 3 | 3 | 3 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 2 | 4 | 5 | 3 | 4 |
| 23 | 97 | 4 | 3 | 5 | 5 | 5 | 5 | 5 | 4 | 1 | 5 | 5 | 5 | 3 | 5 | 1 | 5 | 2 | 5 | 5 | 5 | 5 | 4 | 5 |

# 潜在成長モデル（線形）

## 参考：http://lavaan.ugent.be/tutorial/growth.html

```
#下準備（lavaanパッケージの読み込み)
install.packages("lavaan")
library("lavaan")
#データの読み込み
DATA<- read.csv("C:/Users/satoshi/Desktop/PUBLIC2/Transcend/研究/74,岩波本
    /iwanamimath/第7章データ東大附属2009-2013.csv")
attach(DATA)
#モデルの記述・潜在成長モデル（線形） #” クォテーション(’) “に注意！
growthmodel<-
    'fI=~1*sleeptime1+1*sleeptime2+1*sleeptime3+1*sleeptime4+1*sleeptime5+1*sleeptime6
fS=~0*sleeptime1+1*sleeptime2+2*sleeptime3+3*sleeptime4+4*sleeptime5+5*sleeptime6
'
#分析と出力
fit <- growth (growthmodel, DATA, missing = 'fiml' , fixed.x = FALSE)
summary(fit, standardized = T,rsq=T,fit.measures = TRUE)
```

# 潜在成長モデル（二次）

## 参考：http://lavaan.ugent.be/tutorial/growth.html

```
#データの読み込み
DATA <- read.csv("C:/Users/satoshi/Desktop/PUBLIC2/Transcend/研究/74,岩波本/iwanamimath/第7章データJAHEAD.csv")
attach(DATA)

#モデルの記述
growthmodel<-‘
fI=~1*W60+1*W63+1*W66+1*W69+1*W72+1*W75+1*W78+1*W81+1*W84+1*W87+1*W90
fS=~0*W60+(1-exp(-para*1))*W63+(1-exp(-para*2))*W66+(1-exp(-para*3))*W69+(1-exp(-para*4))*W72+(1-exp(-para*5))*W75+(1-exp(-para*6))*W78+(1-exp(-para*7))*W81+(1-exp(-para*8))*W84+(1-exp(-para*9))*W87+(1-exp(-para*10))*W90
para>0.05 #公開時除外
‘

#分析と出力
fit <- growth (growthmodel, DATA, missing = 'fiml' , fixed.x = FALSE)
summary(fit, standardized = T,rsq=T,fit.measures = TRUE)
```

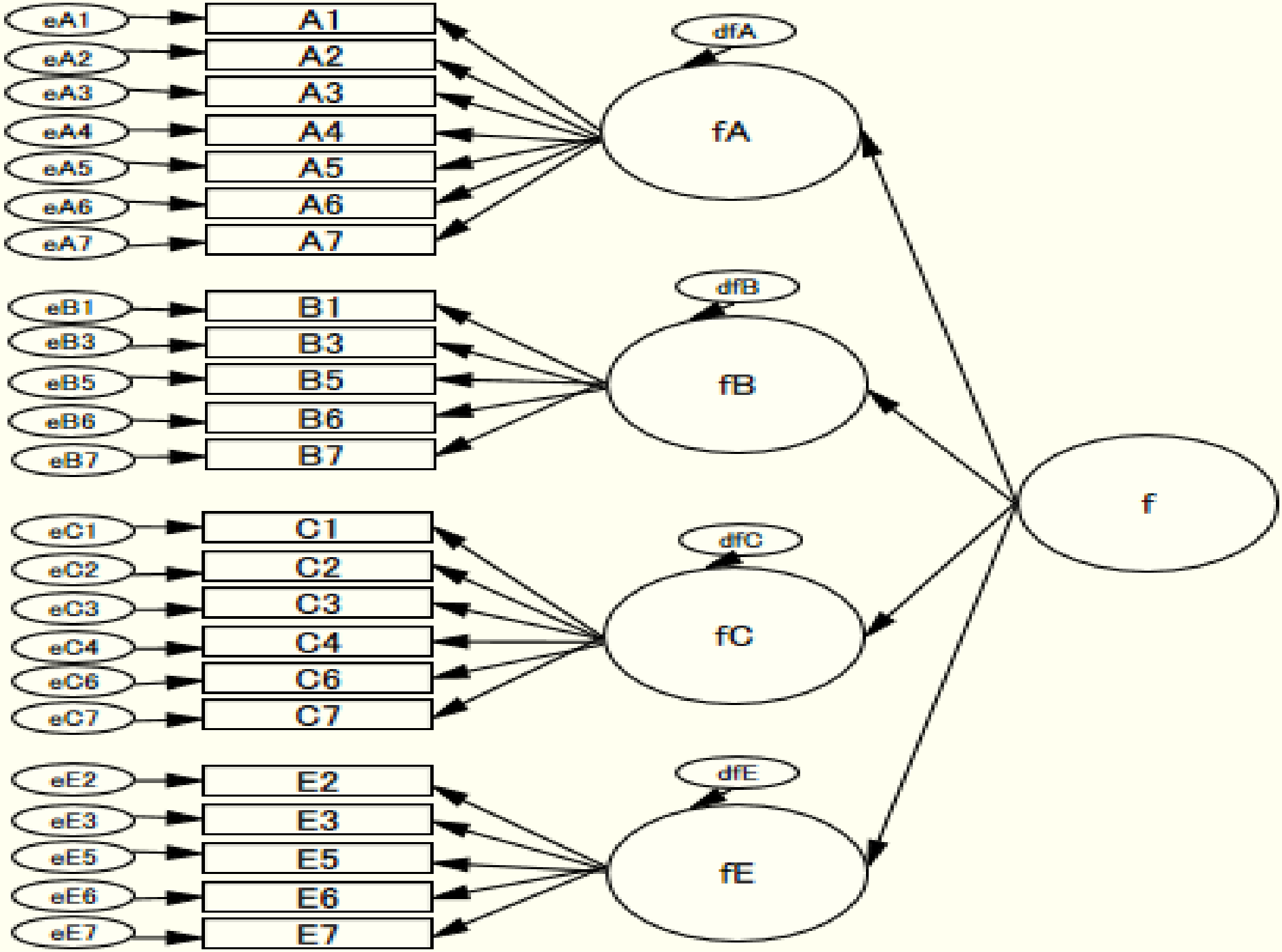
# 二次(高次)因子分析モデル

```
#下準備（lavaanパッケージの読み込み)
install.packages("lavaan")
library("lavaan")
#下準備（データの読み込み)
DATA<-read.csv("C:/Users/Satoshi Usami/Desktop/PUBLIC2/example.data.csv")
#モデルの記述
bifamodel<-'fA=~A1+A2+A3+A4+A5+A6+A7
fB=~B1+B3+B5+B6+B7
fC =~C1+C2+C3+C4+C6+C7
fE=~E2+E3+E5+E6+E7
f=~fA+fB+fC+fE'
#分析と出力
fit <- cfa(bifamodel, data = DATA)
summary(fit, fit.measures = TRUE)
```

#私のPCの場合

=~はmanifested byの意味

前後に'　を忘れない。


eA1
eA2
eA3
eA4
eA5
eA6
eA7
A1
A2
A3
A4
A5
A6
A7
dfA
fA
eB1
eB3
eB5
eB6
eB7
B1
B3
B5
B6
B7
dfB
fB
eC1
eC2
eC3
eC4
eC6
eC7
C1
C2
C3
C4
C6
C7
dfC
fC
eE2
eE3
eE5
eE6
eE7
E2
E3
E5
E6
E7
dfE
fE
f

# 出力

- lavaan (0.5-16) converged normally after  34 iterations

- Number of observations 8961
- Estimator ML
- Minimum Function Test Statistic 10506.631
- Degrees of freedom 226
- P-value (Chi-square) 0.000

- Model test baseline model:
- Minimum Function Test Statistic 105729.828
- Degrees of freedom 253
- P-value 0.000

- User model versus baseline model:

適合度指標

Comparative Fit Index (CFI) 0.903
Tucker-Lewis Index (TLI) 0.891

- Loglikelihood and Information Criteria:

対数尤度

Loglikelihood user model (H0) -268732.279
Loglikelihood unrestricted model (H1) -263478.964

- Number of free parameters 50

情報量規準

Akaike (AIC) 537564.559
Bayesian (BIC) 537919.591
Sample-size adjusted Bayesian (BIC) 537760.699

- Root Mean Square Error of Approximation:

適合度指標

RMSEA 0.071
90 Percent Confidence Interval 0.070 0.072
P-value RMSEA <= 0.05 0.000

- Standardized Root Mean Square Residual:

SRMR 0.050

| | Estimate | Std.err | Z-value | P(>\|z\|) |
|---|---|---|---|---|
| Latent variables: | | | | |
| fA =~ | | | | |
| A1 | 1.000 | | | |
| A2 | 0.899 | 0.018 | 51.175 | 0.000 |
| A3 | 1.131 | 0.018 | 61.350 | 0.000 |
| A4 | 1.204 | 0.019 | 62.916 | 0.000 |
| A5 | 1.315 | 0.019 | 69.687 | 0.000 |
| A6 | 1.255 | 0.019 | 67.619 | 0.000 |
| A7 | 1.058 | 0.017 | 62.571 | 0.000 |
| | | | *途中略* | |
| fE =~ | | | | |
| E2 | 1.000 | | | |
| E3 | 1.109 | 0.014 | 78.161 | 0.000 |
| E5 | 0.779 | 0.012 | 64.319 | 0.000 |
| E6 | 0.858 | 0.014 | 63.020 | 0.000 |
| E7 | 0.884 | 0.012 | 71.519 | 0.000 |
| f =~ | | | | |
| fA | 1.000 | | | |
| fB | 1.505 | 0.026 | 58.972 | 0.000 |
| fC | 1.028 | 0.024 | 43.688 | 0.000 |
| fE | 1.486 | 0.025 | 58.706 | 0.000 |

分散成分
推定値

パスの推定値
標準誤差
検定統計量
p値

Variances:

| | | |
|---|---|---|
| A1 | 0.471 | 0.008 |
| A2 | 0.741 | 0.012 |
| A3 | 0.622 | 0.010 |
| A4 | 0.634 | 0.011 |
| A5 | 0.455 | 0.008 |
| A6 | 0.491 | 0.009 |
| A7 | 0.502 | 0.008 |
| B1 | 0.500 | 0.009 |
| B3 | 0.591 | 0.010 |
| B5 | 0.401 | 0.008 |
| B6 | 0.349 | 0.007 |
| B7 | 0.528 | 0.008 |
| C1 | 0.800 | 0.015 |
| C2 | 1.207 | 0.019 |
| C3 | 1.012 | 0.018 |
| C4 | 1.107 | 0.019 |
| C6 | 0.973 | 0.017 |
| C7 | 1.040 | 0.018 |
| E2 | 0.516 | 0.009 |
| E3 | 0.550 | 0.010 |
| E5 | 0.561 | 0.009 |
| E6 | 0.727 | 0.012 |
| E7 | 0.502 | 0.008 |
| fA | 0.134 | 0.004 |
| fB | 0.069 | 0.005 |
| fC | 0.358 | 0.012 |
| fE | 0.033 | 0.005 |
| f | 0.322 | 0.010 |

# テトリコリック・ポリシリアル相関

- テトラコリック相関…2値データどうしの相関。ポリシリアル相関…2値と連続量データの間の相関。

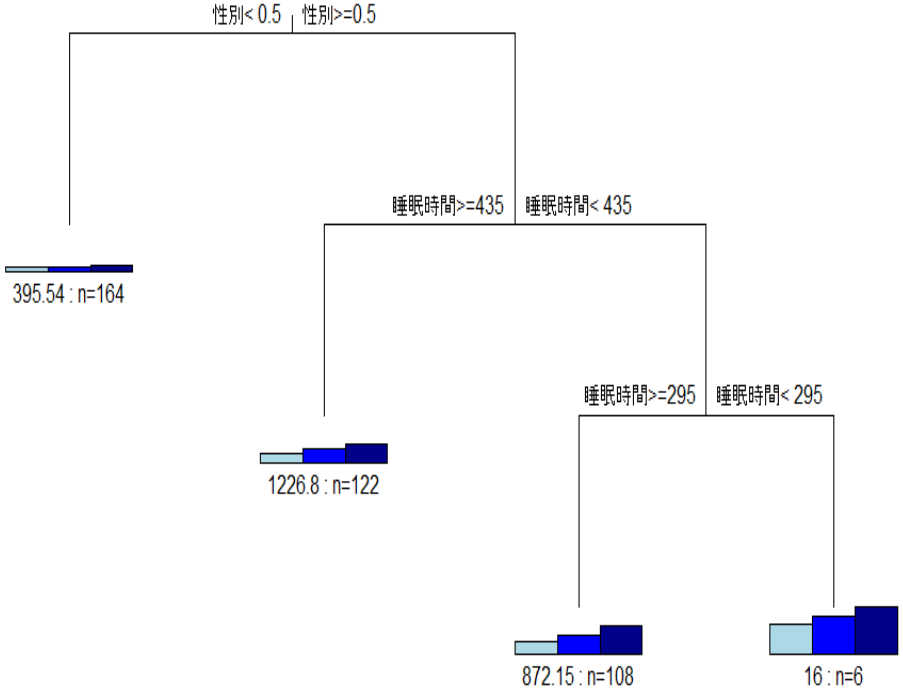
```
library(psych)
library(MASS)#多変量正規分布からの乱数の発生の 為利用
Data<-mvrnorm(10000,c(0,0,0,0) ,0.5+0.5*diag(4)) #N=10000の疑似データ（真の相関が全て0.5)の発生
Data01<-ifelse(Data>0,1,0) #01データの発生
sumscore<-rowSums(Data01) #合計点の計算

#example
biserial(sumscore,Data01[,1]) #bis相関(IT相関)
#example2
tetrachoric(data.frame(cbind(D ata01[,1],Data01[,2])))
#example3
Cormatrix<-matrix(rep(0,16),4, 4) #結果を格納する行列の作成
for(i in 1:4){;for(j in i:4){
```

- Cormatrix[i,j]<-tetrachoric(da ta.frame(cbind(Data01[,i],Data 01[,j])))[[1]][1,2]#相関係数のみ抽出
- Cormatrix[j,i]<-Cormatrix[i,j]
- };}
- Cormatrix　#テトラコリック相関
- cor(Data01)　#Φ係数（ピアソンの積率相関)

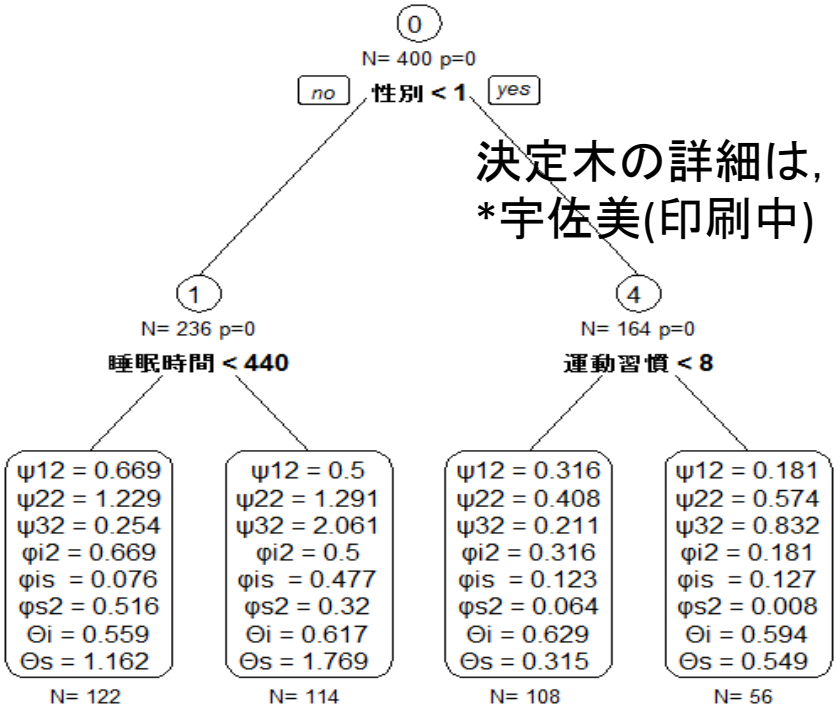
# 統計分析(決定木)

決定木は，（通常のクラスター分析のように）分類の対象となる基準変数のデータのみ利用して分類を行うのではなく，基準変数の量的な違いを説明するための変数である説明変数を積極的に利用して，それによる基準変数の分類を段階的に行う方法。データマイニング手法の一つ。

rpart, mvpartパッケージ




OpenMx,semtreeパッケージ




決定木の詳細は，
*宇佐美(印刷中)

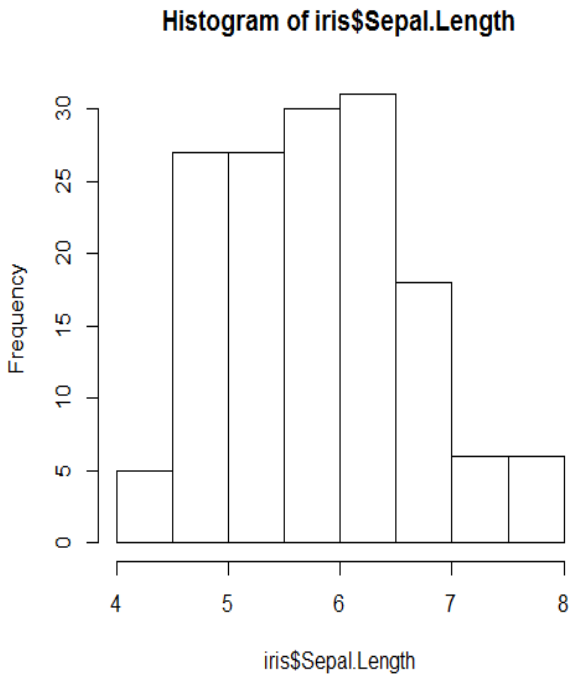
# 図の作成

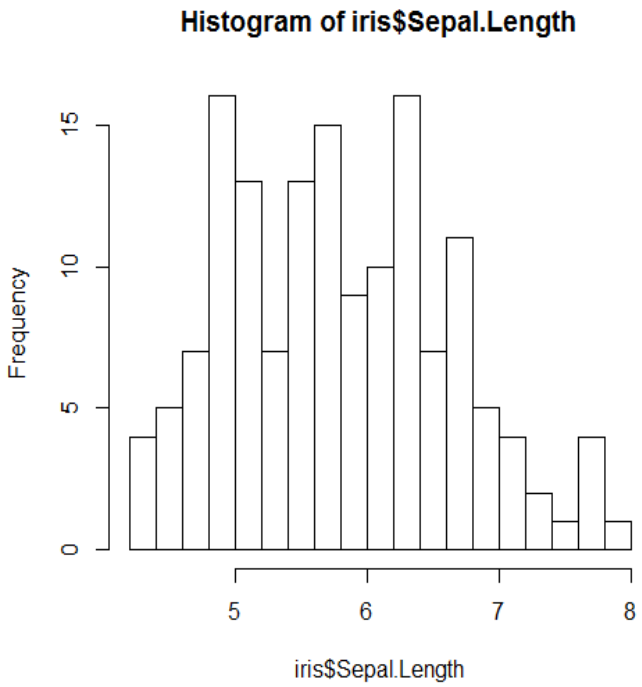
*小杉・押江(2013). Rチュートリアルセミナーより引用。一部改変。

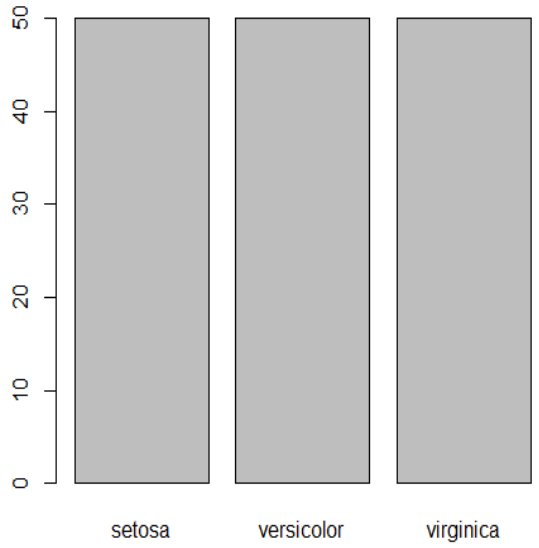
hist(iris$Sepal.Length) #ヒストグラム

hist(iris$Sepal.Length ,breaks =20)  #階級数の設定。

barplot(table(iris$Species)) #棒グラフ



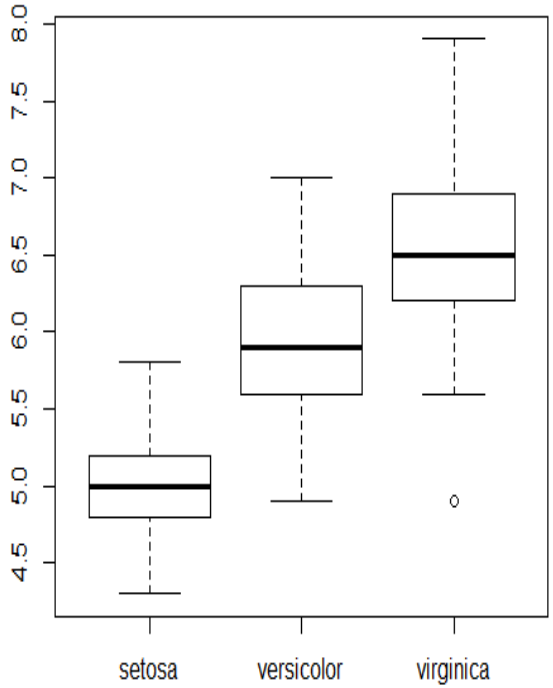
boxplot(Sepal.Length~Species ,data=iris) #箱ひげ図

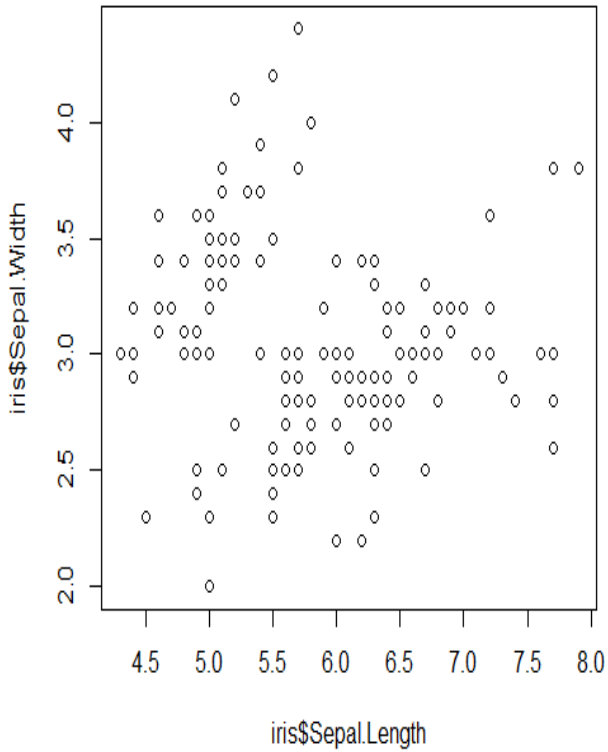
plot(iris$Sepal.Length ,iris$Sepal.Width) #散布図

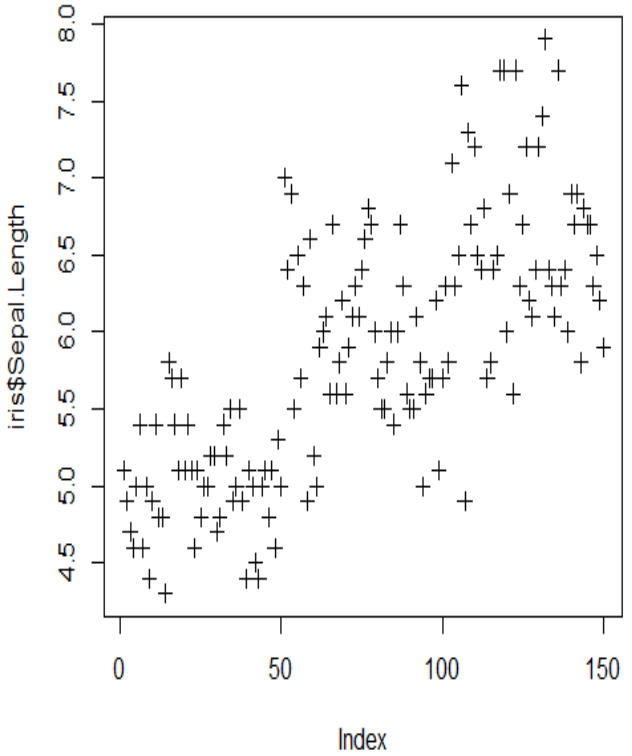
plot(iris$Sepal.Length ,pch =3)　#マークの変更

plot(iris$Sepal.Length ,col="blue")　#色の変更

plot(iris$Sepal.Length ,type="l")　#線の変更



# 資料のアウトライン

- なぜRか？

- Rを使ってみよう

(1)基本操作1　基本統計量，変数作成，関数 etc

(2)基本操作2　パッケージ，データの読み取り etc

(3)統計分析・図の作成

(4)シミュレーションの基礎

# シミュレーションの基礎 (if関数)

**一般表現**

```
if (expr_1) {
expr_2
}else{
expr_3
}
```

**Example:**

```
A<-10
if (A<5) {
C<-1
}else{
C<-0
}
C
```

# シミュレーションの基礎
# (for関数,while 関数)

## for関数

**一般表現**

```
for (name in expr_1) {
expr_2
}
```

**Example:**(1)

```
A<-c(7,5,3,2,5)
SUM<-0
for(j in 1:5){
SUM<-SUM+A[j]
}
SUM
```

## while関数

**一般表現**

```
while (expr_1) {
expr_2
}
```

**Example:**

```
SUM<-0
while(SUM<15){
SUM<-SUM+7
}
SUM
```

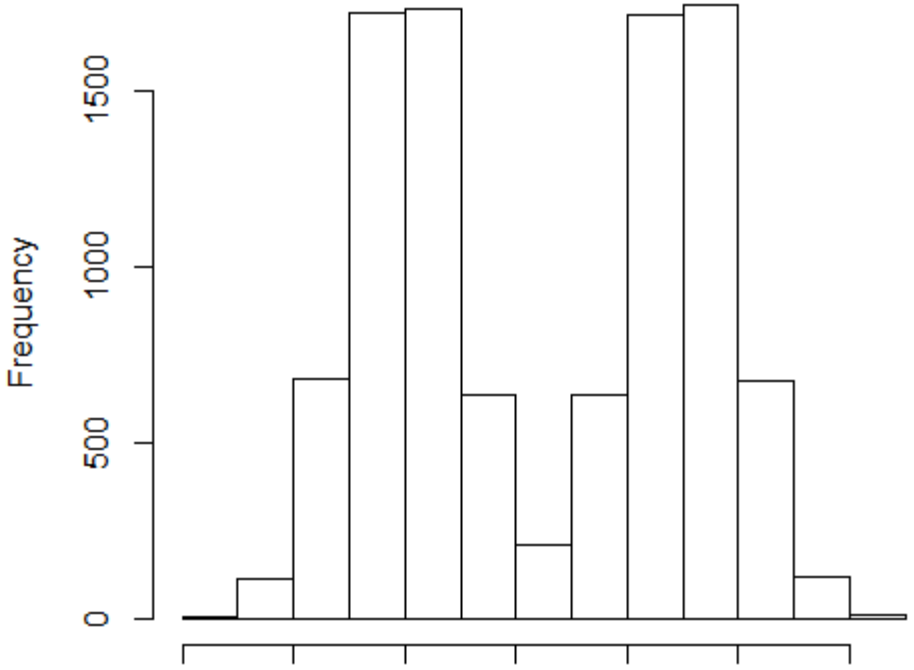
# ブートストラップの例(相関係数の推測）

```
library(MASS)
Data<-mvrnorm(100,c(0,0),0.5+0.5*diag(2))#多変量正規分布からのデータの
    発生。
Y<-Data[,1];X<-Data[,2];cor(Data)
(1-0.5^2)/sqrt(100) # 近似的な標準誤差

#ブートストラップ
CORR<-rep(0,10000)
for(j in 1:10000){
Number<-sample(1:100,100,replace=TRUE)
NewData<-Data[Number,]
CORR[j]<-cor(NewData)[2,1]
}
sd(CORR) #より正確な標準偏差;
hist(CORR)
```

# 確率計算への応用例(混合分布)

```
PROB<-0
for(j in 1:50000){ # 記号としてはj以外でもOK.
A[j]<-rnorm(1,0,1)
 if(A[j]>1){
 PROB<-PROB+1
 }else{
 PROB<-PROB
 }
}
for(j in 50001:100000){ # 記号としてはj以外
A[j]<-rnorm(1,4,1)
 if(A[j]>1){
 PROB<-PROB+1
 }else{
 PROB<-PROB
 }
}
hist(A)
PROB/100000
```

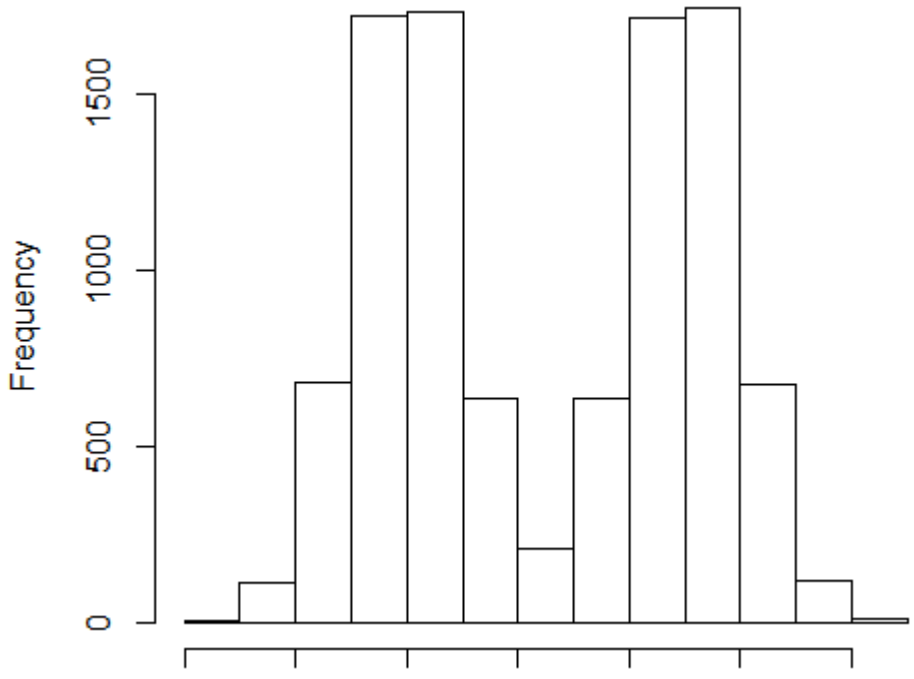
混合分布の例



# より効率的な書き方

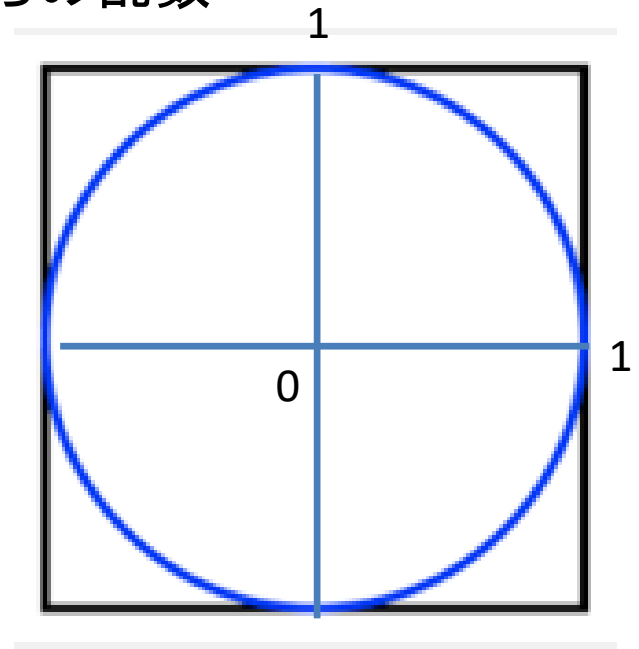
```
DATAA<-rnorm(500000,0,1)
DATAB<-rnorm(500000,4,1)
A<-ifelse(DATAA>1,1,0)
B<-ifelse(DATAB>1,1,0)
hist(c(DATAA,DATAB))
sum(c(A,B))/1000000
```

混合分布の例



# 確率計算への応用例(円周率πを求める)

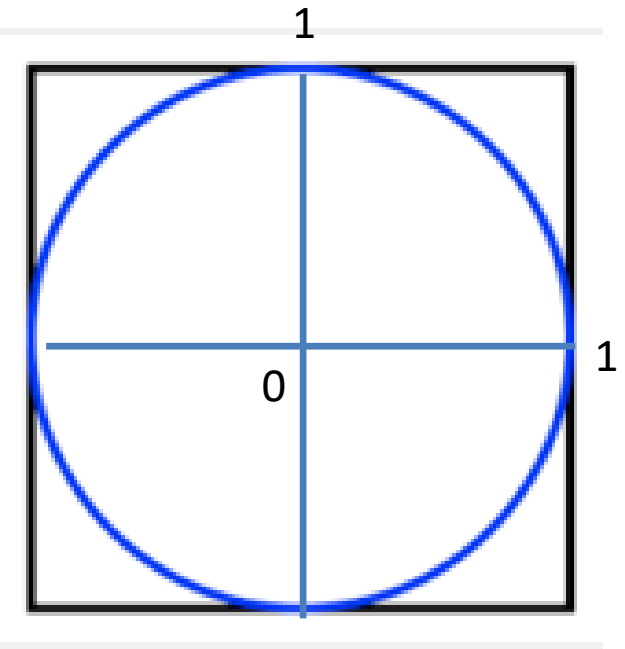
```
PROB<-0
for(j in 1:1000000){
x<-runif(1,0,1); y<-runif(1,0,1) #一様分布からの乱数
A<-x^2+y^2
if(A<1){
PROB<-PROB+1
}else{
PROB<-PROB
}
}
4*PROB/1000000
```




このような計算はモンテカルロ法による積分計算と呼ばれる。

# より効率的な書き方

```
x<-runif(10000000,0,1); y<-runif(10000000,0,1)
data01<-ifelse(x^2+y^2<1,1,0)
4*sum(data01)/10000000
```




このような計算はモンテカルロ法による積分計算と呼ばれる。

# テストの信頼性と偽陰性・偽陽性（得点が正規分布の場合）

```
P<-0.60 #(合格率)
rho<-0.80 #(信頼性)

DATA<-sort(rnorm(1000000,0,sqrt(rho)),decreasing=TRUE)
TRUEDATAPASS<-DATA[1:(1000000*P)]; TRUEDATAFAIL<-DATA[(1000000*P+1):1000000] #真の得点分布

OBSDATAPASS<-TRUEDATAPASS+rnorm(1000000*P,0,sqrt(1-rho)) #真の合格者の実際の得点分布
OBSDATAFAIL<-TRUEDATAFAIL+rnorm(1000000*(1-P),0,sqrt(1-rho)) #真の不合格者の実際の得点分布

OBSDATA<-c(OBSDATAPASS,OBSDATAFAIL)

sum((sign(rank(OBSDATA)[1:length(A)]-1000000*(1-P))+1)/2)/(1000000*P)#真陽性
1-sum((sign(rank(OBSDATA)[1:length(A)]-1000000*(1-P))+1)/2)/(1000000*P)#偽陰性

sum((sign(rank(OBSDATA)[(length(A)+1):length(c(A,B))]-1000000*(1-P))+1)/2)/(1000000*(1-P))#偽陽性
1-sum((sign(rank(OBSDATA)[(length(A)+1):length(c(A,B))]-1000000*(1-P))+1)/2)/(1000000*(1-P))#真陰性
```

# 自分の関数を作る (標本分散)

```
Samplevariance<-function(DATA){
sampleV<-var(DATA)*((length(DATA)-1)/length(DATA))
return(sampleV)
}
Samplevariance(1:5)
```

#関数名は自由につけて良いが，変数名をつける場合のように
　制約条件あり。

# 自分の関数を作る(πの計算)

```
PICALCULATION<-function(T){
  PROB<-0
    for(j in 1:T){
    x<-runif(1,0,1); y<-runif(1,0,1) #一様分布からの乱数
    A<-x^2+y^2
     if(A<1){
     PROB<-PROB+1
      }else{
      PROB<-PROB
      }
     }
   PROB<-4*PROB/T
  return(PROB)
}
PICALCULATION(10000)
```

先程と同じ

# 参考資料

**R, RStudioについての基本的な教科書**

村井潤一郎・山田剛史(2008).はじめてのR: ごく初歩の操作から統計解析の導入まで　北大路書房

舟尾暢男 (2009). The R Tips―データ解析環境Rの基本技・グラフィックス活用集　オーム社

Garrett, G(著)・大橋真也(監修), 長尾高弘(翻訳) (2015). RStudioではじめるRプログラミング入門　オライリージャパン

**Rによる(心理)データ解析・心理学研究法**

服部環 (2011). 心理・教育のためのRによるデータ解析 福村出版

山田剛史・村井潤一郎・杉澤武俊 (2015). Rによる心理データ解析　ナカニシヤ出版

山田剛史(編).  Rによる心理学研究法入門　北大路書房

Paul T(著)・大橋真也(監訳)・木下哲也(翻訳) (2011). Rクックブック　オライリージャパン

## 本資料内の引用

宇佐美慧・荘島宏二郎 (2015). 発達心理学のための統計学―縦断データの分析― 誠信書房

宇佐美慧 (2017). 縦断データの分類 ―決定木および構造方程式モデリング決定木 荘島宏二郎(編) パーソナリティ計量心理学 ナカニシヤ出版

伊藤亜矢子・宇佐美慧 (印刷中). 新版中学生用学級風土尺度（Classroom Climate Inventory；CCI）の作成. 教育心理学研究

## その他英語資料

Peter, D. (2008). *Introductory statistics with R*. 2nd edition. Springer,

Deepayan, S (2007). *Lattice: multivariate data visualization with R.* Springer, New York.

Braun, W.J., & Murdoch, D.J. (2007). *A first course in statistical programming with R*. Cambridge University Press.