

Rによる調査データの整理

茂木 豊

(<http://www.shakaigaku.net>)

オープンソースの統計解析システム「R」

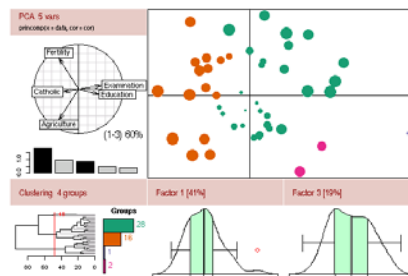


About R
[What is R?](#)
[Contributors](#)
[Screenshots](#)
[What's new?](#)

Download
[CRAN](#)

R Project
Foundation
[Members & Donors](#)
[Mailing Lists](#)
[Bug Tracking](#)
[Developer Page](#)
[Conferences](#)
[Search](#)

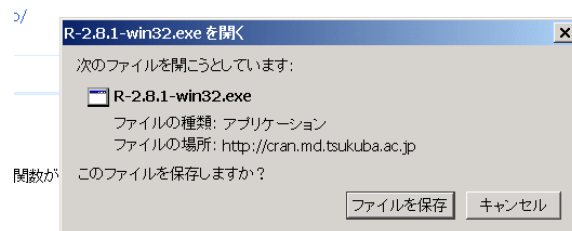
The R Project for Statistical Computing



<http://www.r-project.org/>

Rのインストール

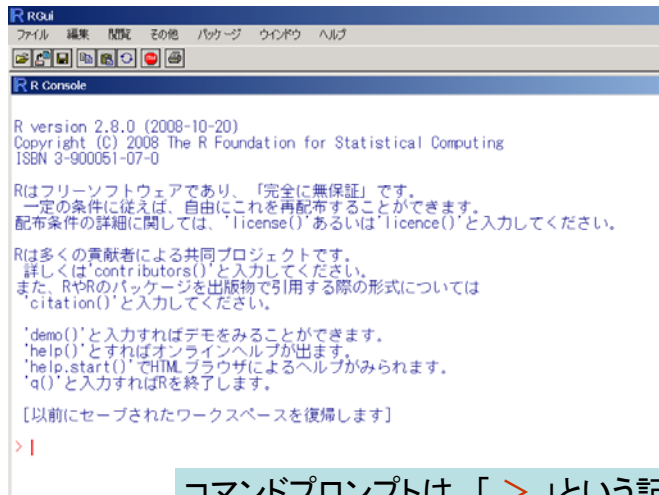
- <http://cran.md.tsukuba.ac.jp/bin/windows/base/R-2.8.1-win32.exe>
(<http://cran.md.tsukuba.ac.jp/>)
ファイルをダウンロードし実行するとRがパソコンにインストールされる。



Rの使い方(その1)

1. 代入
2. 関数
3. ベクトル成分の取り出し
4. 行列(マトリックス)
5. クロス表の検定(分割表における独立性の検定、カイ2乗検定)
6. 比率の差の検定(2つの比率の差の検定)
7. プログラムの終了

プログラム起動直後の画面



```
R GUI
ファイル 編集 閲覧 その他 パッケージ ウィンドウ ヘルプ
R Console
R version 2.8.0 (2008-10-20)
Copyright (C) 2008 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

Rはフリーソフトウェアであり、「完全に無保証」です。
一定の条件に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、「license()」あるいは「licence()」と入力してください。

Rは多くの貢献者による共同プロジェクトです。
詳しくは「contributors()」と入力してください。
また、RやRのパッケージを出版物で引用する際の形式については
「citation()」と入力してください。

「demo()」と入力すればデモをみることができます。
「help()」とすればオンラインヘルプが出ます。
「help.start()」でHTMLブラウザによるヘルプがみられます。
「q()」と入力すればRを終了します。

[以前にセーブされたワークスペースを復帰します]
> |
```

コマンドプロンプトは、「>」という記号。

代入

```
> x1 <- c(3, 4, 5, 6, 7)
> x1
[1] 3 4 5 6 7
> |
```

「<-」という記号が使われる。

関数

```
> x1 <- c(3, 4, 5, 6, 7)
> x1
[1] 3 4 5 6 7
> mean(x1)
[1] 5
> sd(x1)
[1] 1.581139
> |
```

ベクトル成分の取り出し

```
> x1[2]
[1] 4
> x1[3]
[1] 5
> |
```

行列の作成

```
> matrix(1:6, nrow=2, ncol=3)
```

```
  [,1] [,2] [,3]  
[1,]  1  3  5  
[2,]  2  4  6
```

```
> matrix(1:6, nrow=2, ncol=3, byrow=T)
```

```
  [,1] [,2] [,3]  
[1,]  1  2  3  
[2,]  4  5  6
```

```
> m1 <- matrix(1:6, nrow=2)
```

```
> m1 <- matrix(1:6, 2)
```

```
> m1
```

```
  [,1] [,2] [,3]  
[1,]  1  3  5  
[2,]  2  4  6
```

```
> |
```

コロン(:)を使い、1:6とすると、c(1,2,3,4,5,6)と同じになる。

rowは行、colは列(column)の意味。

行列成分の取り出し

```
> m1
```

```
  [,1] [,2] [,3]  
[1,]  1  3  5  
[2,]  2  4  6
```

```
> m1[1,2]
```

```
[1] 3
```

```
> m1[2,3]
```

```
[1] 6
```

```
> m1[1,]
```

```
[1] 1 3 5
```

```
> m1[2,]
```

```
[1] 2 4 6
```

```
> m1[,1]
```

```
[1] 1 2
```

周辺度数の追加

```
> m1
  [,1] [,2] [,3]
[1,]  1   3   5
[2,]  2   4   6
> m2 <- addmargins(m1, 1)
> m2
  [,1] [,2] [,3]
[1,]  1   3   5
[2,]  2   4   6
[3,]  3   7  11
> m2 <- addmargins(m1, c(1, 2))
> m2
  [,1] [,2] [,3] [,4]
[1,]  1   3   5   9
[2,]  2   4   6  12
[3,]  3   7  11  21
> |
```

addmargins(A)は、以下と同じ。
addmargins(A, margin = 1:length(dim(A)), FUN = sum)

クロス表の検定

```
> m3 <- matrix(c(24, 12, 12, 18, 9, 5), nrow=2)
> m3
  [,1] [,2] [,3]
[1,] 24  12   9
[2,] 12  18   5
> chisq.test(m3)
```

Pearson's Chi-squared test

data: m3
X-squared = 5.1737, df = 2, p-value = 0.07526

```
> |
```

カイ2乗値が5.2、自由度2、p値が0.075という結果。帰無仮説を棄却することができないので、関連がないと解釈。

フィッシャーのexact test

```
> fisher.test(m3)
```

```
Fisher's Exact Test for Count Data
```

```
data: m3  
p-value = 0.084  
alternative hypothesis: two.sided
```

```
> |
```

比率の差の検定

```
> prop.test(c(278, 242), c(800, 745))
```

```
2-sample test for equality of proportions  
with continuity correction
```

```
data: c(278, 242) out of c(800, 745)  
X-squared = 0.7891, df = 1, p-value = 0.3744  
alternative hypothesis: two.sided  
95 percent confidence interval:  
-0.02574150 0.07107707  
sample estimates:  
prop 1 prop 2  
0.3475000 0.3248322
```

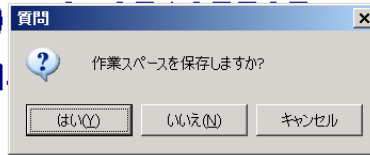
278/800と242/745と
いう2つの比率の差について
検定。

プログラムの終了

```
95 percent confidence interval:  
-0.02574150  
sample estimates:  
prop 1  
0.3475000 0.3248322
```

```
> q()
```

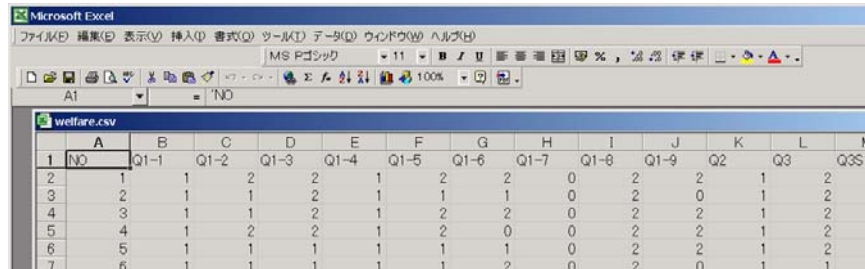
R version 2.8.0 (2008-10-20)



Rの使い方(その2)

1. Excelデータの読み込み
2. 単純集計
3. クロス集計
4. 周辺度数の計算
5. 列パーセント、行パーセントの計算
6. グラフの作成
7. 検定(ピアソン、フィッシャー)

Excelデータの読み込み



| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|----|------|------|------|------|------|------|------|------|------|----|----|------|
| 1 | NO | Q1-1 | Q1-2 | Q1-3 | Q1-4 | Q1-5 | Q1-6 | Q1-7 | Q1-8 | Q1-9 | Q2 | Q3 | Q3S1 |
| 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 0 | 2 | 2 | 1 | 2 | |
| 3 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 0 | 2 | 0 | 1 | 2 | |
| 4 | 3 | 1 | 1 | 2 | 1 | 2 | 2 | 0 | 2 | 2 | 1 | 2 | |
| 5 | 4 | 1 | 2 | 2 | 1 | 2 | 0 | 0 | 2 | 2 | 1 | 2 | |
| 6 | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 2 | 2 | 1 | 2 | |
| 7 | 6 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 2 | 0 | 1 | 1 | |

```
> data01 <- read.delim("clipboard")
```

読み込まれたデータの確認

```
> data01 <- read.delim("clipboard")
> head(data01)
```

| | NO | Q1.1 | Q1.2 | Q1.3 | Q1.4 | Q1.5 | Q1.6 | Q1.7 | Q1.8 | Q1.9 | Q2 | Q3 | Q3S1 |
|---|----|------|------|------|------|------|------|------|------|------|----|----|------|
| 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 0 | 2 | 2 | 1 | 2 | |
| 2 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 0 | 2 | 0 | 1 | 2 | |
| 3 | 3 | 1 | 1 | 2 | 1 | 2 | 2 | 0 | 2 | 2 | 1 | 2 | |
| 4 | 4 | 1 | 2 | 2 | 1 | 2 | 0 | 0 | 2 | 2 | 1 | 2 | |
| 5 | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 2 | 2 | 1 | 2 | |
| 6 | 6 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 2 | 0 | 1 | 1 | |

```
      Q6.2 Q7 Q8 Q8S1 Q8S2 Q8S3 Q9 Q10 Q10S1 Q10S2 Q10S3 Q11 Q11
1      2  1  1    1    0    0  2    1    1    0    0    2
2      2  3  2    0    0    0  2    3    0    0    0    1
```

単純集計、クロス集計

```
> attach(data01)
> table(Q4)
Q4
 1  2  3  4
33 86  6 11
> table(Q4, 学科)
  学科
Q4   0  1  2  3
 1   1 10 10 12
 2   6 28 42 10
 3   0  1  5  0
 4   0  8  2  1
> |
```

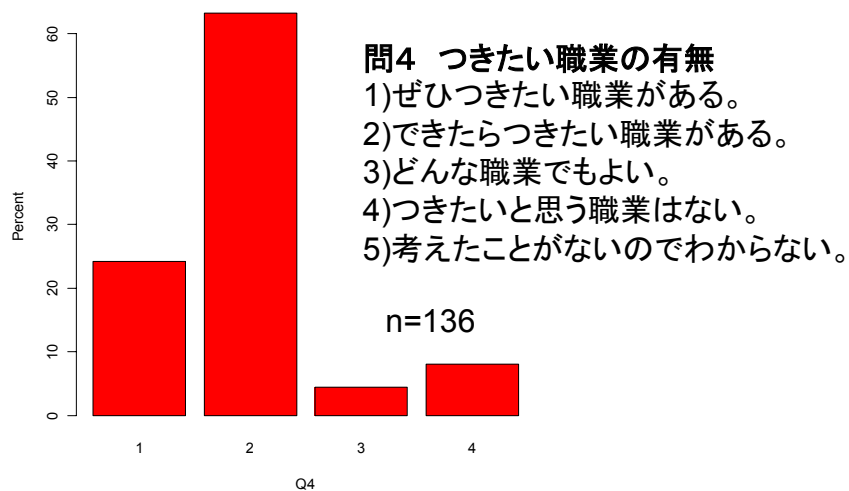
周辺度数の表示

```
> table(Q4, 学科)
  学科
Q4   0  1  2  3
 1   1 10 10 12
 2   6 28 42 10
 3   0  1  5  0
 4   0  8  2  1
> t1 <- table(Q4, 学科)
> t2 <- addmargins(t1, c(1,2))
> t2
  学科
Q4   0  1  2  3 Sum
 1   1 10 10 12 33
 2   6 28 42 10 86
 3   0  1  5  0  6
 4   0  8  2  1 11
Sum   7 47 59 23 136
```

列パーセントの計算

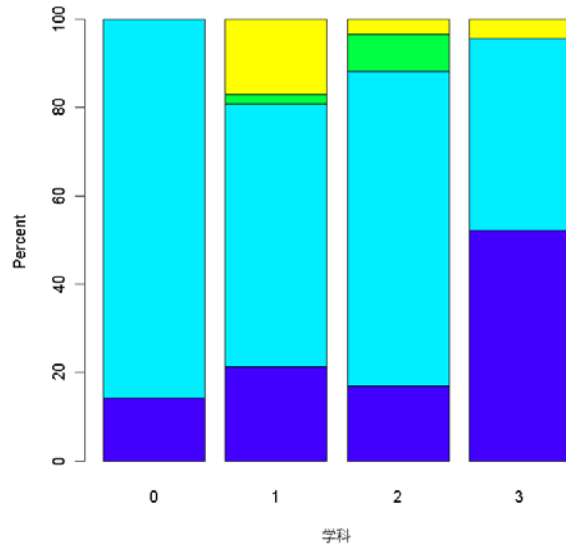
```
> t3 <- 100*prop.table(t1, 2)
> t3
  学科
Q4   0         1         2         3
  1 14.285714 21.276596 16.949153 52.173913
  2 85.714286 59.574468 71.186441 43.478261
  3  0.000000  2.127660  8.474576  0.000000
  4  0.000000 17.021277  3.389831  4.347826
```

グラフの作成(単純集計)



```
barplot(table(Q4)/sum(table(Q4))*100, col="red", xlab="Q4", ylab="Percent")
```

グラフの作成(クロス集計)



```
barplot(100*prop.table(table(Q4,学科),2), col=topo.colors(4),xlab="学科",ylab="Percent")
```

1列目の削除

```
> t1
  学科
Q4  0  1  2  3
  1  1 10 10 12
  2  6 28 42 10
  3  0  1  5  0
  4  0  8  2  1
> t1b <- t1[,-1]
> t1b
  学科
Q4  1  2  3
  1 10 10 12
  2 28 42 10
  3  1  5  0
  4  8  2  1
```

「ピアソンのカイ2乗検定」

```
> t1b
  学科
Q4  1  2  3
  1 10 10 12
  2 28 42 10
  3  1  5  0
  4  8  2  1
> chisq.test(t1b)

Pearson's Chi-squared test

data:  t1b
X-squared = 20.6103, df = 6, p-value = 0.002155

Warning message:
In chisq.test(t1b) : カイ自乗近似は不正確かもしれません
```

期待度数が5未満のセルは？

```
> t1b
  学科
Q4  1  2  3
  1 10 10 12
  2 28 42 10
  3  1  5  0
  4  8  2  1
> chisq.test(t1b)$expected
  学科
Q4  1      2      3
  1 11.658915 14.635659  5.705426
  2 29.147287 36.589147 14.263566
  3  2.186047  2.744186  1.069767
  4  4.007752  5.031008  1.961240
Warning message:
In chisq.test(t1b) : カイ自乗近似は不正確かもしれません
> |
```

各セルの寄与度

```
> (O-E)^2/E
  学科
```

```
Q4      1      2      3
1 0.23604239 1.4682860 6.9445568
2 0.04515916 0.8001642 1.2744355
3 0.64349332 1.8543555 1.0697674
4 3.97680416 1.8260771 0.4711217
```

「フィッシャーの正確検定」

```
> fisher.test(t1b)
```

```
Fisher's Exact Test for Count Data
```

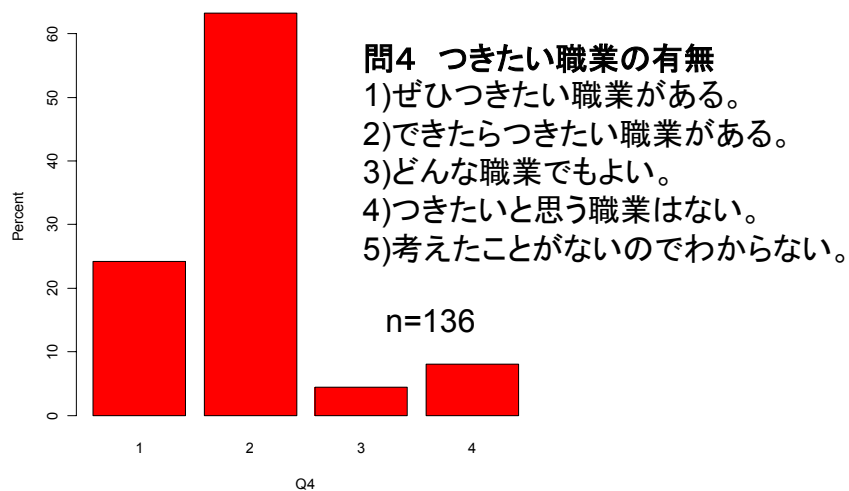
```
data: t1b
p-value = 0.003923
alternative hypothesis: two.sided
```

問4 あなたは、将来つきたいと思っている職業がありますか。

1. ぜひつきたい職業がある。
2. できたらつきたい職業がある。
3. どんな職業でもよい。
4. つきたいと思う職業はない。
5. 考えたことがないのでわからない。

「カテゴリの合併」は可能か？

どこで分けることができるか？



```
barplot(table(Q4)/sum(table(Q4))*100, col="red", xlab="Q4", ylab="Percent")
```

クラマー係数

```
cr <- function(x) {  
  if (dim(x)[1] > dim(x)[2] )  
  return(sqrt(chisq.test(x)$statistic/sum(x)/(dim(x)[2]-1)))  
  else return(sqrt(chisq.test(x)$statistic/sum(x)/(dim(x)[1]-1)))  
}
```