

R の使い方 (R 2.8.1)

--- R コマンダー を使って楽をする ---

34

インストール (とパッケージの追加) : これは最初に一度だけする。

■ Windows 版

- ・ ファイルをダウンロードする。

<http://cran.md.tsukuba.ac.jp/bin/windows/base/R-2.8.1-win32.exe>

1. ダブルクリックして普通にインストールする。
2. プログラム「R 2.8.1」を起動する。
3. GUI で使用するためのパッケージ「R Commander」を追加する。
 1. メニューの「パッケージ」から「パッケージのインストール」を選ぶ。
 - ・ 最初は「CRAN mirror」というダウンロードサイトの一覧が出るので「Japan (Tsukuba)」を選ぶとよい。
 - ・ インターネット上から直接ダウンロードしてインストールされる。
 2. パッケージの一覧表が出るので「Rcmdr」を選びインストールする。
 - ・ 「Rcmdr」及びそれに依存するパッケージがインストールされる。

R コマンダーの起動

(プログラム「R 2.8.1」は起動されているものとする)

1. メニューの「パッケージ」から「パッケージの読み込み」を選ぶ。
2. パッケージの一覧表が出るので「Rcmdr」を選ぶ。
 - ・ 場合によっては、初回、足りないパッケージのインストールを促されるので指示に従いインストールする。(90 個近くインストールされる)

データの読み込み

■ 他のファイルを読み込む場合

【サンプルデータ】 (プリテストとポストテスト : 対応のあるデータ)

- ・ 「R コマンダー」のメニュー「データ」より「データのインポート」を選ぶ。
1. テキストファイルは「テキストファイルまたはクリップボード ,URL から」を選ぶ。
 1. テキストファイルの場合
 - ・ データは変数ごとに縦に並べておく。
 - ・ 「フィールドの区切り記号」でフィールドの区切りがタブ区切りの場合はタブを選択
 - ・ 「OK」を押すと「ファイルを開く」ウインドウが開くので、目的のファイルを選ぶ。
 2. クリップボードの場合
 - ・ データの範囲をマウスで選び「コピー」しておく
 3. インターネットの URL の場合
 2. Excel のファイルは「Excel または、、、のデータセットから」を選ぶ。
 3. SPSS のファイルは「SPSS データセットから」を選ぶ。
- ・ 本当に読み込まれたかどうかは「R コマンダー」の「データセットを表示」ボタンを押し

てみる。

■ 直接、データを入力する場合

「R コマnder」のメニューの「データ」から「新しいデータセット」を選ぶ。

1. 「データセット名を入力」
2. 適当に名前を付けて「OK」を押す。（「Dataset」のままでもよい）
3. 「データエディタ」ウインドウが開くので、入力する。（変数名も入力 OK。ただし半角英数字のみ）

統計解析（R コマnder 上のメニューの「統計量」）

（データセットは読み込まれているものとする）

■ 「要約」

- ・「アクティブデータセット」（5 数要約値と平均）

```
reading (変数名)
Min.    : 0.00 (最小値)
1st Qu.: 11.00 (第1四分位): 全部を並べて、4 分の 1 の位置にある数
Median : 15.00 (中央値: メジアン)
Mean    : 15.03 (平均)
3rd Qu.: 18.50 (第3四分位)
Max.    : 31.00 (最大値)
```

- ・「数値による要約」（平均、標準偏差、分位数 (quantile)、パーセンタイル等）
 - ・変数（一つ以上選択）（Ctrl を押しながら連続選択）

mean(平均) sd(標準偏差) quantiles(分位数) n(総数)

```
> numSummary(Dataset[, "pre"], statistics=c("mean", "sd", "quantiles"),
+   quantiles=c(0, .25, .5, .75, 1))
   mean      sd 0% 25% 50% 75% 100%   n
39.36301 15.20702  3  29  39  48  78 146
```

- ・「相関行列」

1. 相関を見たい変数を選ぶ（二つ以上）（Ctrl を押しながらクリック）
2. 相関の種類を選ぶ
 1. 「ピアソンの積率相関」（Pearson product-moment）
 2. 「スピアマンの順位相関」（Spearman rank-order）（ノンパラメトリック）
 3. 「偏相関」（三つの変数の関係で、一つの影響を除いた残りの二つの相関）

```
> cor(Dataset[, c("grammar", "reading", "vocabulary")], use="complete.obs")
           grammar  reading vocabulary
grammar  1.0000000 0.1993079  0.4200132
reading   0.1993079 1.0000000  0.3321875
vocabulary 0.4200132 0.3321875  1.0000000
```

■ 「分割表」（Contingency tables）

- ・ Two-way table
- ・ multi-way table
- ・ 二元表の入力と分析（Enter and analyze two-way table）（カイ 2 乗検定）

■ 「平均」（Means）

- Sigle-sample t-test
- Independent-sample t-test

• 対応のある t 検定 (Paired t-test)

1. 「 First variable 」 を選ぶ。
2. 「 Second variable 」 を選び 「 OK 」 を押す。

```
> t.test(Dataset$PRE, Dataset$POST, alternative='two.sided', conf.level=.95, paired=TRUE)
Paired t-test
data: Dataset$PRE and Dataset$POST
t = -4.1009, df = 19, p-value = 0.0006085
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-5.210833 -1.689167
sample estimates:
mean of the differences -3.45
```

• One-way ANOVA (一元配置分散分析 : 三つ以上のグループの平均に差があるか)

- データを以下のように整形しておく必要がある。
 - それぞれのデータがどのグループに属するかを明記するカラム 「 factors 」 を設ける。
 - 変数名は 「 factors 」 (文字) にしておかないと分析できないので注意。
 - 変数名の編集は、 「 Edit data set 」 ボタンを押して、変数名をクリックする。変数名を入力し、 「 numeric 」 ではなく 「 character 」 を選ぶ。
 - たとえば三つのグループの得点 (score) のデータだとして以下ようになる :

factors	score
1	5
1	4
1	6
1	6
1	4
2	6
2	6
2	7
2	5
2	6
3	7
3	7
3	7
3	9
3	7
3	8

1. データセットを読み込む : 「データ」 > 「インポートデータ」 > 「テキストファイルから」 (もしくは、 「 Data Editor 」 で直接入力)
2. 「統計」 > 「代表値」 > 「One-way ANOVA」
3. 左の 「 Groups 」 と右の 「 Response Variable 」 それぞれ表示されている項目をクリック。
4. 下の 「 Pairwise comparisons of means 」 は多重比較。これにチェックをいれる。で、 「 OK 」

```
> anova(lm(score ~ factors, data=ano))
Analysis of Variance Table

Response: score
      Df Sum Sq Mean Sq F value    Pr(>F)
factors  2 17.5000   8.7500  11.974 0.001125 ** (0.01 の危険率で有意)
Residuals 13  9.5000   0.7308
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> tapply(ano$score, ano$factors, mean, na.rm=TRUE) # means ( 平均 )
      1      2      3
5.0 6.0 7.5

> tapply(ano$score, ano$factors, sd, na.rm=TRUE) # std. deviations ( 標準偏差 )
```

```

      1      2      3
1.0000000 0.7071068 0.8366600

> tapply(ano$score, ano$factor, function(x) sum(!is.na(x))) # counts (数)
1 2 3
5 5 6

> summary(simtest(score ~ factor, type="Tukey", data=ano))

Simultaneous tests: Tukey contrasts (テューキーの方法による多重比較)

Call:
simtest.formula(formula = score ~ factor, data = ano, type = "Tukey")

Tukey contrasts for factor factor

Contrast matrix:
      factor1 factor2 factor3
factors2-factors1 0      -1      1      0
factors3-factors1 0      -1      0      1
factors3-factors2 0       0     -1      1

Absolute Error Tolerance: 0.001

Coefficients:
      Estimate t value Std.Err. p raw p Bonf p adj
factors3-factors1      2.5 -4.830   0.541 0.000 0.001 0.001 (有意)
factors3-factors2      1.5 -2.898   0.518 0.012 0.025 0.023 (有意)
factors2-factors1      1.0 -1.850   0.518 0.087 0.087 0.087 (有意差なし)

```

• Multi-way ANOVA

■ 「Proportions」(割合)

- Single-sample proportion test
- Two-sample proportion test

■ 「Variances」(分散)

- Two-variances F-test
- Bartlett's test
- Levene's test

■ 「Nonparametric tests」(ノンパラメトリック)

- Two-sample Wilcoxon test
- Paired-sample Wilcoxon test
- Kruskal-Wallis test

■ 「Dimensional analysis」(次元解析)

- Scale reliability
- Principal-components analysis
- Factor analysis

■ 「Fit models」

- Linear regression
- Linear model
- Generalized linear model

グラフ

■ ヒストグラム

■ 箱ひげ図

■ QQ プロット

- ・「分布」を選ぶことで、その分布が想定している分布にあっているかどうか
- ・「正規」を選べば、正規分布になっているかがわかる。

■ 散布図

R for Beginners

The R Project for Statistical Computing

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

R の本家。

The Comprehensive R Archive Network (通称 CRAN)

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

R やその追加パッケージのダウンロード元。

R の使い方 - 杉浦研究室

<http://oscar.gsid.nagoya-u.ac.jp/tech/wiki/wiki.cgi?page=R%A4%CE%BB%C8%A4%A4%CA%FD>

R と R コマンダーのインストール、データの読み込み、データの要約、出力の見方についての説明など。

CRAN(筑波大学) のミラーサイト

<http://cran.md.tsukuba.ac.jp/bin/windows/base/>

R の最新版 (Windows・日本語対応版) のインストールはこちらから。～.exe をクリック。インストール方法は、下の「R 言語マニュアル 1」を参照。

R 言語マニュアル 1

<http://web.sfc.keio.ac.jp/~maunz/wiki/index.php?R%B8%C0%B8%EC%A5%DE%A5%CB%A5%E5%A5%A2%A5%EB1>

R と R コマンダーのインストールを図解。

R 言語マニュアル 2

<http://web.sfc.keio.ac.jp/~maunz/wiki/index.php?R%B8%C0%B8%EC%A5%DE%A5%CB%A5%E5%A5%A2%A5%EB2>

R コマンダーによる、データの読み込み、データの要約、グラフの作成・保存等を図解。

統計解析フリーソフト R 入門

<http://cwoweb2.bai.ne.jp/~jgb11101/files/R-commander.pdf>

R コマンダーの使い方を全般的に図解。『R Commander ハンドブック』(九天社) は、このページが本になったものらしい。

統計解析・R-Tips

<http://cse.naro.affrc.go.jp/takezawa/r-tips/r.html>

R の基本操作、ベクトル・行列の作成から、グラフィックス、統計解析まで。目次が充実しているので、必要な情報を得やすい。

統計処理ソフトウェア R についての Tips

<http://phi.ypu.jp/swtips/R.html>

R の基本操作から、R 関連のニュースまで。

R 人口増加のきっかけとなったと言われる中澤先生の著書「R による統計解析の基礎」PDF 版がフリーでダウンロードできる。

RjpWiki

<http://www.okada.jp.org/RWiki/index.php?RjpWiki>

R に関する巨大コミュニティ。初心者のための R に関する質問コーナーあり。

R による統計処理

<http://aoki2.si.gunma-u.ac.jp/R/>

群馬大学青木先生のページ。膨大な数の統計解析手法の説明の他、日本における R 関連のリンク集もあり。

R 入門 (R 2.5.0)

最近、すっかり変わってしまったので、あらためて再入門。

入手とインストール

データはどうする？

■ Excel などから直接コピーする

1. 見出しも含めてマウスで該当セルをドラッグして選択し、メニューの「編集」から「コピー」する。
2. クリップボードからデータを読み込む

```
dat <- read.delim("clipboard")
```

■ テキストファイルに保存しておく

1. タブ区切りでテキストファイルに保存する。
 - ・保存する場所を覚えておくように。C ドライブの出来るだけ上に保存するとディレクトリを長く書かなくてよいので便利。
2. ファイルを指定し、データを読み込む。(下の例は、C ドライブ直下に置いた例)

```
dat <- read.delim("c:/file.txt")
```

■ Web サーバーに乗せておく

1. Web サーバー上で、外部からアクセス制限なしでアクセスできるところにおく。
2. 置いたファイルの URL を指定して読み込む。

```
dat <- read.delim("http://www.sample.ac.jp/~data/file.txt")
```

使用例：単語数に差があるかどうか調べる

■ どちらの検定を使うか（パラメトリックかノンパラメトリックか）

- ・ 正規性の検定：Shapiro-Wilk normality test
 - ・ データを dat に読み込み、見出しが Pretest となっている場合を例に。

```
shapiro.test(dat$Pretest)
```

- ・ これで出てくる p-value を見て、0.05 以上あれば、正規分布しているといえる。
- ・ ヒストグラムを描いて分布を見てみる。

```
hist(dat$Pretest, breaks=seq(0,100,10))
```

- ・ 上の例は、0 点から、100 点まで、10 点刻みで、という指定。

■ パラメトリック = t 検定

1. 同じ人が書いているものの場合（対応があるデータ）
2. 別の人が書いているものの場合（対応のないデータ）
 1. 等分散性の検定

■ ノンパラメトリック = Wilcoxon 符号順位検定