

基礎統計学

(基礎医科統計学)

第1回 ガイダンス Rを用いた記述統計

2016. 9.30

開講日

- 博士課程 共通領域 選択科目 (1単位)
(修士課程 共通教育科目「基礎医科統計学」を兼ねる。)
担当：長崎 生光, 吉井健悟 (数学教室)
- 博士課程開講日 (全7回) :
9/30, 10/14, 10/21, 10/28, 11/11, 11/25,
12/2 (吉井講師担当)
修士課程：上記に加え (吉井講師担当)
10/17, 11/7, 11/14

演習？

シラバスには演習についての記述があるが・・・

オフィスアワーでとして設定されているので、集まる必要はありません。原則は各自で勉強してください。

ただし質問・相談は随時受け付けます。(e-mailでOK)

必要なもの

- ノートパソコン（可能ならば）
- R（R Commander, EZR を含む）
- Excel

を使います。

成績について

博士課程の場合（修士課程は別途指示します）

- 1 規定回数以上の出席.
- 2 出席回数が足りない場合, レポート提出が必要.
該当者は来年1月6日（金）までに長崎宛にメールにて提出のこと
- 3 やむを得ない事情のある人は個別にご相談ください.

HP について

講義の資料や課題は下記におきます (期間限定).

URL: http://iku3.webcrow.jp/Med_Stat/

メールで連絡する場合があります.

受講連絡がまだの人は長崎宛にメールしてください.

連絡先 nagasaki@koto.kpu-m.ac.jp

統計学の必要性

- 現代の医学研究には統計学的な裏付けが必須
- Evidence の発見のために (EBM)
- 論文の読み書きに統計の知識も必要
- 最近では論文の統計分析が正しいかどうかチェックされる。

最低限の（医学 or 生物）統計学の考え方・手法を理解しておくのが望ましい。

医学・生物統計の参考書

加納, 高橋「基礎 医学統計学」南江堂

柳川, 荒木「バイオ統計の基礎」近代科学社

森實敏夫「入門医療統計学」東京図書

古川, 丹後「医学への統計学」朝倉書店

丹後「統計学のセンス」朝倉書店

など

補足 — ベイズ統計学の参考書

近年コンピューターソフトの進歩により，ベイズ流の統計学が浸透しつつある。

(ある統計学のジャーナルでは過半数の論文がベイズ統計に関する論文であるという.)

豊田「はじめての統計データ分析」朝倉書店

豊田「基礎からのベイズ統計学」朝倉書店

アルバート「Rで学ぶベイズ統計学入門」シュプリンガー

統計ソフト

実際の統計分析には、統計ソフトが必須.

SPSS

- 1 ver.16 から Win, Mac, Linux 版が用意されている.
- 2 アカデミック版あり
- 3 解説本が多数

統計専用ソフト

JMP(ジャンプ)

- 1 ver. 7 から Win, Mac, Linux 版が用意されている.
- 2 解説本もいくつかある
- 3 アカデミック版あり.

フリー統計ソフト R

これらのソフトは一般には高価.

そこで, 本講義ではフリーの統計ソフト「R」およびそのパッケージを使う.

フリー統計ソフト R

R

- 1 フリーソフト. Win, Mac, Linux 版あり
- 2 最近は解説本も多数
- 3 キーボードからのコマンド入力 (CUI) が標準だが, **EZR** (後述) をインストールすることでマウス操作 (GUI) も可能
- 4 Web 上 (CRAN, Bioconductor など) に多数の R 用のパッケージが公開されている.
- 5 rstan などを導入するとベイズ流の分析も可能になる.

Rのインストール

情報源

- 1 Rによる統計処理
<http://aoki2.si.gunma-u.ac.jp/R/>
- 2 RjpWiki
<http://www.okada.jp.org/RWiki/>

インストールファイル

CRANのミラー <http://cran.ism.ac.jp/>

からリンクをたどっていけば見つかる。

Rで使うデータファイル

- R単体ではすべてコマンドで操作する.
- データファイルはエクセルでつくるのが楽.
- コンマ区切り形式 (拡張子 csv) で作成し, 作業フォルダーに保存.

Rで使うデータファイル (例)

- 1 body.csv : 青年 (Y) と老年 (A) の男 (M), 女 (F) の身長, 体重のデータ. (産総研: 人体寸法データベース 1991-92 から抽出. ただし一部を改変)
- 2 bodyMY.csv : 上のデータのうち, 青年男子の身長, 体重のデータ

Rで使うデータファイル (例)

ここでは bodyMY.csv を読み込む.

- Rから直接読み込むには

```
data <- read.csv("bodyMY.csv")
```

と入力

Rcommander (EZR)

- インストールすることでマウスで操作 (GUI) が可能になる.
- R Commander の機能を大幅に拡張したものが EZR.
- 医療統計学でよく使う検定や分析法は, ほぼ網羅している.

EZR (イージーアール)

EZR の本家は

EZR

<http://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.html>

参考書：神田善伸著，EZR でやさしく学ぶ統計学
改訂2版，中外医学社 (CDつき)

上記の HP にインストール方法 (Windows, Mac, Linux) が説明されている。

R Commander (EZR) の起動

- 1 R を起動する.
- 2 R に `library(Rcmdr)` と入力し return.
- 3 R Commander (EZR) が立ち上がる.

R Commander から読み込むには . . .

EZRのメニュー

一例をあげると

- パラメトリック検定
 - t検定
 - 一元配置分散分析
 - 多重比較 (Tukey-Kramer 法、Bonferoni 法)

- カイ二乗検定、Fisher の正確検定
ロジスティック回帰
- Mann-Whitney U 検定
Wilcoxon 符号付順位和検定
- 生存期間の解析
Logrank 検定、一般化 Wilcoxon 検定
Cox 比例ハザード回帰
- ROC 曲線
陽性適中率、陰性適中率の計算
- Mantel-Haenzel 検定
条件付ロジスティック回帰
- 必要サンプルサイズの計算

記述統計学

記述統計学とは、観察されたデータを整理し、その性質を記述することを目的とする。

一方、検定や推定は、推測統計学といわれる。

考え方（思想）の違いで、頻度流とベイズ流の2つの流れがある。

記述統計

具体的には

統計的数値

- 1 平均, メディアン (中央値)
- 2 標準偏差, 分散
- 3 相関係数など

表とグラフ

- 1 度数分布表
- 2 ヒストグラム
- 3 回帰直線

EZR を用いた記述統計

- 1 R Commander (with EZR) を立ち上げる
- 2 bodyMY.csv の読み込み
- 3 基本統計量の計算
- 4 ヒストグラム
- 5 回帰直線

R の操作 — 例

① bodyMY.csv の保存 (Excel でカンマ区切り形式
でドキュメントフォルダーに保存)

② bodyMY.csv の読み込み

```
data <- read.csv("bodyMY.csv")
```

③ 身長データ data\$Ht

④ 基本統計量の計算

```
mean(data$Ht), sd(data$Ht),
```

```
summary(data$Ht)
```

⑤ ヒストグラム hist(data\$Ht)