

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

2017. 9. 29

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) :
9/29, 10/13, 10/20, 10/27 : 17:00~18:00
11/10, 11/24, 12/1 : 17:30~18:30
修士課程 : 上記に加え (吉井講師担当)
10/2, 10/23, 11/6

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
- R (R Commander, EZR を含む)
- Excel (データファイル用)

成績について

29度入学生から 優, 良, 可, 不可 で判定
(可以上で合格)

それ以前入学は 合格, 不合格 の判定

- 1 出席回数
- 2 レポート問題 (数回) の提出
レポートは来年1月9日 (火) までに随時
長崎宛に提出のこと (e-mailでOK)

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

講義の資料やレポート問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
- 一流のジャーナルでは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 . . .

医学でよく使われる統計手法

これらのほとんどは R (EZR) で実行可能

2群の差の検定

- t検定 (対応なし, あり)
- Wilcoxon の順位検定
- Wilcoxon の符号付き順位検定
- 符号検定

相関係数と回帰分析

- 相関係数と回帰直線
- Spearman の順位相関係数
- 重回帰分析

分割表の検定

- χ^2 検定
- Fisher の exact test

分散分析と共分散分析

- 分散分析
- 共分散分析
- 多重比較

リスクファクターの分析

- リスク比とオッズ比
- Mantel-Haenszel 法
- ロジスティック回帰分析

生存曲線の分析

- Kaplan-Meier 法
- log-rank 検定
- Cox の比例ハザードモデル

記述統計学と推測統計学

- **記述統計学**とは、観察されたデータを整理し、その分布や特徴を記述することを目的とする。
- 一方、検定や推定、予測などは、**推測統計学**に属する。
- また、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。
- 統計学の理論は、確率論がベースになる。統計的な正しさは確率的なものである。

記述統計

数値的な記述

- 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

視覚的な記述

- 1 ヒストグラム
- 2 箱ひげ図
- 3 相関図, 回帰直線

Rのみで操作すると・・・

- 1 bodyMY.csv の読み込み

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

- 2 データの表示 data

- 3 身長データの表示 data\$Ht

- 4 基本統計量（平均，標準偏差など）の計算

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

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

Rのみで操作すると・・・

- 1 身長と体重の相関係数

```
cor(data$Ht, data$Wt)
```

- 2 身長, 体重のヒストグラム, 箱ひげ図

```
hist(data$Ht) , boxplot(data$Wt)
```

- 3 身長と体重の相関図 + 回帰直線の表示

```
plot(data$Ht, data$Wt)
```

```
abline(lm(data$Wt~data$Ht), col="red")
```

EZR を用いると . . .

- 1 R Commnder (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 身長, 体重のヒストグラム
- 5 相関図と回帰直線