

確率と統計

担当教員： 中山 謙二

電子メール： nakayama@t.kanazawa-u.ac.jp

オフィスアワー：メールで連絡されたい

確率と統計の必要性

仕事、研究のあらゆる場面で「データを処理し、そこから知識を得る」ことが行われる。

例1：創造実験でデータを取る

例2：プロジェクトデザインでアンケート調査をする

生のデータ＝数字が並んでいるだけ

→ 眺めても何も分からない。

データ&統計処理 → 数字の意味、傾向がわかる
(平均いくつ、何人中何人がそれを欲しいと思うか)

→ 意味を持つ知識(方針が立てられる)

確率と統計の例(1)

「大学生のお小遣いはいくらぐらい？」と聞かれたら？

自分は2万円だけど、これだけでは何とも言えない。
友達にも聞いてみる→1万円、2万円、3万円、4万円いろいろ
いったいいくらなんだ？→平均的には2万5千円かな。

もし、2.2, 2.4, 2.5, 2.6, 2.8万円なら、自信を持って2万5千円
と言える。

もし、0.5, 1, 1.5, 4, 5.5万円なら、平均値は2万5千円だが自信
は持てない。→「2万5千円ぐらいですが、人それぞれです」

データの数が(少ない／多い)→信頼性が(低い／高い)
データのばらつきが(小さい／大きい)→信頼性が(高い／低い)

データの平均値やばらつきを調べる→統計解析

授業の概要と目標

学習教育目標：統計学的数学手法の基本概念を学び、
統計的資料の処理および解釈の方法を習得する。

概要：

- 統計解析ソフトウェア
- 一変数の記述統計
- 二変数の記述統計
- 母集団と標本
- 統計的仮説検定

確率と統計の例(2)

ある箱の中に赤、青、黄色のボールが同じ数だけ入っている。
無作為に1個取り出すとしたら、「何色になるか期待しない」。

赤が20個、青が5個、黄色が1個だとすると、「ほとんど赤が
出る、またに青が出る、黄色は減多に出ない」と期待する。

15個のボールを取り出し、赤が12個、青が3個だったとすると
「そろそろ黄色が出るかも」と期待する。確率を計算している

ボールを取り出した後、箱に戻すとすると、黄色が出るまで何
回取り出せばいいのか？

→断定できないが、「確率的には」26回でよい。

授業の進め方

教科書：「Rによるやさしい統計学」
講義では第1章～第5章を学ぶ予定である。

コンピュータ演習を組み合わせる。
毎回ノートパソコンを持参すること。

“技術者のための統計”「新統計入門」(数学的側面)
この教科書も持参すること。

学習の達成目標

標準的な達成レベル:

1. 統計解析ソフト「R」で基礎的なデータ処理ができる.
2. 一変数と二変数の記述統計の基本統計量を説明できる.
3. 確率分布や条件からその確率を計算できる.
4. 正規分布を理解し, 応用できる.
5. 標本から母集団の統計分布を推定できる.

★理想的な達成レベル→学習支援計画書参照

講義のWeb Page

http://leo.ec.t.kanazawa-u.ac.jp/~nakayama/edu/kit_prob_sta.html

をクリック(準備中)

- 授業の予定表
- 講義スライド
- 各種ダウンロード
- 課題情報

コンピュータ演習は, 教科書を見ながら進める. スライドで適宜補足する.

成績評価方法

試験:	40%
小テスト:	30%
演習, レポート:	30%
(合計)	100点
合格	60点以上

出欠

1/3以上の欠席はF判定となる(自動的に不合格).

★不合格者には再試験, 又はレポートを課す場合があるので, 教員からの連絡には十分に注意すること.(条件があります!)

私語について

学生は高い授業料を払って静かに授業を履修する権利を持っているが, その権利を妨害する権利はさらさら持っていないことを自覚してほしい.

私語等, 授業の妨害になる行為は注意する.

1回の注意で(100点満点)→20点のペナルティを課す.

学生の病気欠席等

- インフルエンザ関係→公欠の手続きをすること。(教務課から連絡があった場合のみ公欠である.)
- 授業内容は自習で補うこと.
講義ウェブページからダウンロードできる資料, 及び教科書などを自習する.
- 公欠の場合は追試験を行う.

講義内容について

今日の内容:教科書第1章

1. Rとはなにか
2. Rのインストール
3. Rを使ってみる

使用パソコン:Windowsを前提とするが, Mac, linuxでもOK. (ただし, OS依存部分は自分で補うこと.)

Rとはなにか

Rとは: 統計計算とグラフィックスのための言語・環境

- フリーソフトウェア(無料)
- プログラムコードの公開と結果の信頼性の高さ
- プログラミング言語であることの利点
複雑な解析が簡単な操作でできる
他人が作ったプログラム(パッケージ)を使用可能
- GUIの弱さ
- 日本語対応
- 幅広いOSに対応(Windows, Mac OS, Unix)
- グラフィックス機能が充実している。

Rを使ってみる(1)

- デスクトップにあるアイコン「R」をダブルクリック
- 基本的な計算・・・1.4.2節
- 統計解析の例・・・1.5節
- ベクトル表示, 行列表示・・・1.6.1節

Rのインストール(1)

教科書の「1.3 Rを導入しよう」, P.9～に従ってインストールする。

◆Rをダウンロードする◆
IEを立ち上げる
<http://cran.md.tsukuba.ac.jp>
にアクセスする。

使用するOSをクリックする。Windowsの場合は
「Download R for Windows」を左クリック
「base」または「install R for the first time」を左クリック
「Download R 2.15.0 for Windows」を左クリック

「保存」を左クリック
保存先を選んで保存する。

Rを使ってみる(2)

- データファイル(.csv)を使用する・・・1.6.2節
Excelで表1.1を作成→csvファイルとして作業ディレクトリに保存
データファイルの読み込み

```
> ホークス2 <- read.csv("hawks.csv")
```

```
> ホークス2
```
- 作業ディレクトリ: ファイル→ディレクトリの変更→ディレクトリを選択(先に行う/どこでも良い)
- 関数の作成・・・1.6.3節

```
varp <- function(x){var(x)*(length(x)-1)/length(x)}
```

Rのインストール(2)

◆Rをインストールする◆
 > 保存した「R-2.15.0-win.exe」をダブルクリックする。
 > 言語の選択 日本語 OK
 > R for Windows 2.15.0セットアップウィザードの開始 「次へ」をクリック
 > GNU GENERAL PUBLIC LICENSE ライセンスに同意 「次へ」をクリック
 > プログラムをインストールする場所を選択する(どこでも良い) 「次へ」をクリック
 自動的に「R-2.15.0」というフォルダが作成される(変更可能)
 > コンポーネントの選択
 32-bit利用者向けインストールを選択(デフォルト)「次へ」をクリック
 > 起動時オプション 「いいえ(デフォルトのまま)」を選択し、「次へ」をクリック
 > プログラムグループの指定 デフォルトのまま「次へ」をクリック
 (追加タスクの選択 デフォルトのまま「次へ」をクリック。クイック起動アイコンを作成する場合はチェックを追加する)
 ★インストールを開始する(完了するまでPCを操作しないこと) 「完了」をクリック
 > デスクトップにあるアイコン「R」をダブルクリックする→「R」が立ち上がる。

CSVファイルとは

データをカンマ(",")で区切って並べたファイル形式。

主に表計算ソフトやデータベースソフトがデータを保存するときに使う形式だが、汎用性が高く、多くの電子手帳やワープロソフトなどでも利用できるため、異なる種類のアプリケーションソフト間のデータ交換に使われることも多い。

実体はテキストファイルであるため、テキストエディタやワープロなどで開いて直接編集することも可能。

作業ディレクトリの初期設定法

作業ディレクトリとは: プログラムやデータのファイルを置く場所. Rは作業ディレクトリからファイルを読み込んだり, 出力する.

起動時の作業ディレクトリ設定→教科書P.394

デスクトップ上のショートカットを右クリック
「プロパティ」を選択
「作業フォルダ(S)」→使用する作業ディレクトリに変更
→適用→OK

ショートカットのダブルクリック→Rが起動→作業ディレクトリは指定したディレクトリになっている.

作業スペース

「作業スペース」: それまで計算して代入した変数とその値 注意! それまでの作業そのままではない.

Rの終了時に「作業スペースを保存しますか?」と尋ねられ, 保存すると次回はその作業スペースを自動的に読み込む.

→毎回作業スペースを保存しつづけると, どんどん変数などが増えていくことに注意する.

確認: 例えば `c<-3+5` を実行し, Rを終了する. このとき作業スペースを保存しておく.

→Rを起動し, すぐに `c`(Enter) してみよう. 計算していないはずなのに, 8が表示される.

Rを使ってみる(3)

- プログラムファイルの作成と読み込み・・・1.6.4節
「ファイル」→「新しいスクリプト」→プログラム作成
→「ファイル」→「保存」→作業ディレクトリに保存される読み込み `> source("xxx.R")`
- 別の読み込み方法:
「ファイル」→Rコードのソース読み込み→ファイルを選択→「開く(O)」→読み込まれる

作業スペースの保存

「ファイル」→「作業スペースの保存」

「ファイル」→「作業スペースの読み込み」

* 読み込んだ作業スペースの状態から作業を開始できる.

* R終了時の作業スペース保存は「直前」の状態のみ.

テキストファイルとして保存

1. 「ファイル」→「ファイルを保存...」

2. ファイル名をつけてテキストファイルとして保存

3. 保存したファイルをメモ帳等で開くことができる.

4. 「ファイル」→「ファイルを表示」で表示できる

* そのファイルから有効なコマンド(命令文)をRconsoleへコピー&ペーストできる.

Rを使ってみる(4)

- プログラムパッケージをインストール・・・1.6.5節
「パッケージ」→「パッケージのインストール」
→ミラーサイト選択→パッケージ選択
→パッケージがダウンロード&インストールされる<時間がかかる>
- パッケージの読み込み
`> library(sem)` 少し時間がかかる
- Rcmdrを使ってみる・・・1.6.5節 ... 後でも良い

今日のまとめ

1. 講義全体の概要
2. Rのインストール
3. 教科書第1章+α

次回の内容:

教科書第2章, 度数分布表, ヒストグラム, 平均値(予定)

試験日程などは講義ウェブページからリンクしている講義スケジュール(予定)を参考にすること(準備中).