

多変量解析 ～ RとRStudioを用いて～

林 篤裕 (おもひ領域&社会工学専攻)
@11号館 5階 516室
hayashi.atsuhiro@nitech.ac.jp
052-735-5119 (内線 5119)



1

第2回(10/08): 目次

1. 前回のアンケートに対して
2. Rを使ってみる
3. データをファイルから読み込むには
4. 初歩的な統計関数
 - u 基礎統計量
 - u 視覚的に把握する
5. 来週に向けて

2

1. 前回のアンケートに対して

- u 16名が提出。
 - u 全員が経営システム分野 3年生
- u 受領者の学籍番号を掲載済み。
 - u 10/1 22:05, 10/2 22:00
 - u 気になる方は確認ください。
- u メールは受領されたかが送信者には確認できないため。

3

b. 「データサイエンス」に抱くイメージや印象

- u 「データサイエンス」と聞くと、過去の結果から得られるデータを活用してよりよい未来を実現するために必要な技術という印象を抱きます。
- u 「データを研究・分析する」ことに特化した分野
- u これからビッグデータやIoT化によって多くのデータが手に入るようになって考えている。その中でデータから何が読み取れるか、データから次にどう生かしていけるかが重要になってくる。そのためデータサイエンスは将来的に必ず必要になってくるのではないかと印象がある。
- u データサイエンスの印象は、集められたデータを分析して規則性などを見つけ出す学問だと思っていた。データ通信の知識がない人がデータを見ていても、見出だせないようなものを発見するといった感じである。
- u データサイエンティストという言葉が思い浮かび、データを収集して統計を行うというイメージ
- u データで物事や人の行動をモデル化し、予測するためのもの
- u データに基づいて、問題発見と解決を行うことというイメージがある。情報社会の現代で重宝される分野だと思っている。
- u データの分析、考察から社会に役立つ知見を得る学問
- u データを何のために使うかや、達成したい目標を設定し、的確な方法でデータを集めそれを分析する。
- u データを適切に分析して、元のデータを何倍にも活用するための技術であるとイメージしております。
- u データを様々な手法で分析することで、データ利用の手助けをする。データの分析を行うところまでがデータサイエンスの領域だと思っていた。
- u データ解析のためにソフトを用いる能力だけでなく、結果から自ら分析する能力も必要だと思う。
- u 技術的な効果の測定や、高度なマーケティング・社会の動向の調査などで用いられている。
- u 重回帰分析などを用いて問題解決や経営戦略に役立てたりする
- u 情報を多く保存、活用されている今の時代では様々な分野で使われている学問
- u 対象に新たな付加価値を見出すときにデータからアプローチする取り組みだと考えています。マーケティング的な知見ではなく、モデルやプログラムといったデータ構造を読み取ることに特化しているイメージがある。
- u 難しい。実際の経営方針の決定との結びつきがよくわからない

データ
分析
規則性
モデル
知見
活用
難しい

4

c. 講義についての要望等

- u Rを使うのは初めてなので、使い方の説明については詳細にしてくださいと助かります。
- u プログラミングは苦手なので丁寧(に説明して)いただけると嬉しいです
- u 演習課題がある場合、解説もして欲しいです。
- u 課題が少ないとうれしいです
- ★ 課題提出やアンケート等はmoodle上で行えるようにするとスムーズにできるので可能であればmoodleでしてほしい。
- ★ 咳払いをした際、音量が一時的にとでも大きくなってしまっていたため、マイクから離れてしていただきたいと感じました。
- u 興味をもって取り組めるような講義であること
- u 現時点で講義形式や進め方は問題なく感じています。
- ★ 講義がちゃんと受けれていない時連絡が欲しい。心配であるため
- ★ 講義スライドを共有していただけるのとことでしたので、“PCでのメモの取りやすさ”の観点から(PowerPointでスライドを制作していた場合)PowerPointファイルのままで共有していただけるとうれしいです。
- ★ 講義の予習等に活用するために、講義資料をなるべく早めに公開されると非常に助かります。
- u 自分のスピーカーが悪いのかもかもしれませんが、少し声が小さいように感じました。
- ★ 分析方法や考察方法などの実例があるとより深い分析になると前期で感じました。
- ★ 翌週以降も第1回講義の内容を公開した状態にしてほしいです。同様に、第2回講義以降の講義もすべて公開した状態にしてほしいです。前期に「その講義日/その週」で非公開にする講義もいくつかありましたが、メモを取っていても新しく気になってきたとき、「教授に連絡する」か「未解決のまま進む」かの2択になってしまい、重要度がわからないためとても窮屈な思いをしていました。
- u 今のところありません。
- u 今のところ特になし
- u 特になし: 2件

キャプチャボード: 発注済み
マイク: 発注中
キーボード: 考慮中

回答を希望する質問であった場合は、その旨お知らせください。来週取り上げます。5

e. その他、質問やお気づきの点があれば何なりと。

- ★ 「出席」の基準を教えてください。「講義当日(公開~24:00)」に「動画を視聴」して「その日の課題」を「締め切りまでに提出」することが要件なのか、それ以外に「講義時間内に動画を視聴すること」や「講義動画を「1倍速」で見ること」なども要件なのか、それとも成績の採点には使用しない(欠席回数のカウントのみ)のため、あまり重視していないのかを教えてください。
- u 厳密な規定は難しい。
- u 大学は「自学」。
- u 従来はいわゆる「出席点」はなかった。レポートで評価。
- u ただ、今回の講義形態(リモート講義)では学生の動静が全く観察ができない。少なくとも受講確認ぐらいはほしいところ。==== 何か示唆や希望がございましたか?
- u 翌朝ぐらいまでにはフィードバックがほしい。==== 時刻に希望があればお聞かせください。
- u 一度決めたら8回の中では固定しておきたい = 「ルール」だから。
- u 説明が丁寧で聞きやすいです。
- u 音声も問題なく、pdfの提示をしていただき、大変受けやすい環境を整えていただきありがとうございます。
- u [欄外に記入] 履修するかどうかは前期の成績及び研究室配属の結果によって決めたいと考えています。よろしくお祈りします。
- u よろしくお祈りします。
- u ありません
- u 特にありません: 2件
- u 特にないです: 1件
- u 特になし: 6件

f. 教育用端末の利用予定についての『有/無』
u 無: 9件
u 有 問題があった場合: 1件

6

d. データ収集

u	1●3	●5	●8.5	自宅生	0	9000
u	1●5	●1	84	自宅生	0	8500
u	1●7	●5	●5	自宅生	0	5000
u	1●7	●5	NA	自宅生	0	10000
u	1●7.5	●8	不明	自宅生	0	3000
u	1●9	●0	●0.5	下宿生	●●000	●●00
u	1●9.3	●3	●0	自宅生	0	5000
u	1●0	●5	92	自宅生	0	5500
u	1●0	●4	●9	自宅生	0	●●●●
u	1●0	●4	不明	自宅生	0	5300
u	1●3	●1.5	76	自宅生	0	5500
u	1●5	●8	●3	NA	NA	NA
u	1●6.6	●2.3	88	自宅生	NA	5000
u	1●9	●7	●2	自宅生	NA	NA
u	1●0	●0	88	自宅生	NA	7700
u	1●1	●7	●2	自宅生	0	不明

- u 上記の「NA」(No Answer)は回答欄はあるものの、値が記入されていなかったもの。
- u 一方、「不明」は回答欄自身がなかったもの。
- u 「NA」、「不明」には少なくとも3つの意味があるだろう。
 - u 「答えたくない」/「測定していない」/「忘れた」: 識別の必要の有無
- u 今回は「不明」は「NA」に修正(集約)する。

7

【再掲】

5. インストールできたかの確認

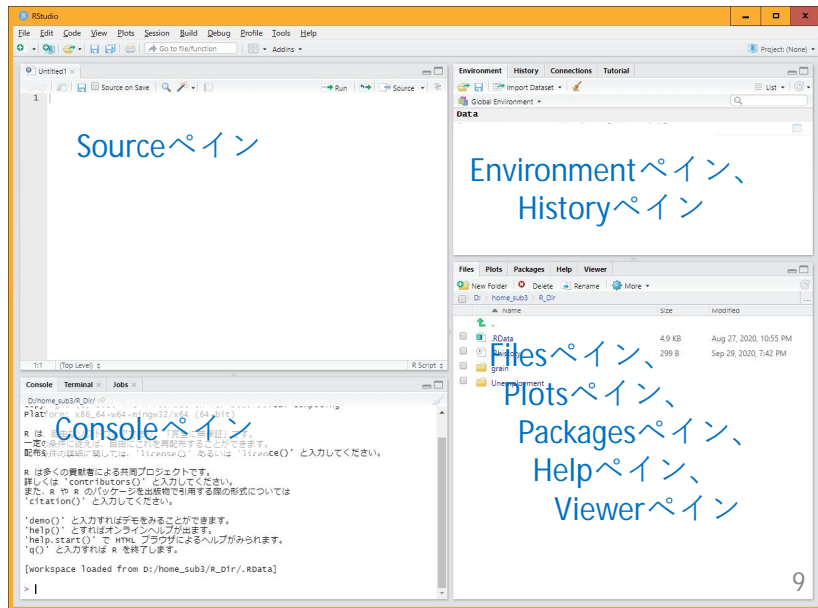
- RStudioの起動
 - デスクトップに表示されているrstudioのアイコンをクリックしてRStudioを起動する
 - Rはrstudio内から操作されるので意識する必要なし
- 画面の説明: 4つの画面がありそれぞれに役割あり
 - 左上: Sourceペイン <=== プログラムを入力する
 - 左下: Consoleペイン <=== 計算結果が表示される
 - 右上: Environmentペイン、Historyペイン
 - 右下: Filesペイン、Plotsペイン、Packagesペイン、Helpペイン、Viewerペイン
- もし、画面が3つしか表示されておらずSourceペインが1行しか表示されていなかったら、Consoleペインの上部(バー部分)をダブルクリックする。



8

【再掲】

5. インストールできたかの確認



【再掲】

5. インストールできたかの確認

- 手始めに簡単な操作(電卓のような使い方)
- Sourceペイン(左上の領域)で
 - 「3+5」と入力後、「Ctrl+Enter」(Ctrlキーを押しながら、Enterキーを押す)を行うと、Consoleペインに「8」と計算結果が出力される。
 - 「5^3」と入力後、「Ctrl+Enter」を行うと、Consoleペインに「125」と計算結果が出力される。
- 演算記号: 四則演算: +、-、*、/ べき乗: ^ 剰余: %% 平方根: sqrt()
- 「式を入力」後、「Ctrl+Enter」で実行する。
- これでは、まるで単なる電卓
- 来週からより高度な計算について紹介します。

10

5. 来週に向けて



- ☐ Rは期待通りの結果を表示してくれたであろうか。
 - ☐ 最初はなかなか慣れなくて、イライラする局面もあるとは思いますが、実用に耐えうるツールなので習得してほしい。
 - ☐ 順に新しい関数(コマンド)を体得して行けば良い。
- ☐ データの電子化にはExcelが重宝する。
 - ☐ 各自の持参データも電子化しておくように。
- ☐ 導入で躓くとこの後は付いて来られないので、早めに疑問は解決すること。
 - ☐ 例えばファイルの読み込みができないと、レポートも...
- ☐ 来週はより複雑な関数について。
 - ☐ 皆さんのアンケートデータを使って

11

∞. アンケート(ショート課題)

- ☐ 顔が見えない講義形式なので、フィードバックがほしい。
- ☐ 以下項目について「10月8日中(講義当日中)」に回答下さい。本日(第2回)の出欠調査を兼ねます。
 - a. 今回の講義の感想: 講義内容のみならず説明方法等も。
 - b. 不明な点、疑問点
 - c. その他、質問や希望、お気づきの点があれば何なりと。
- ☐ 回答にはMoodleに「課題」という形で準備してみた。初体験。うまく行けば良いが。
 - ☐ 直接入力とWord等のファイルでの提出の両方を用意した予定(意図通りか?)。
- ☐ こちらの連絡の必要が生じた場合は、先週同様新たな資料を掲載します(掲載日時を付して)。

12