

# Rによる計量分析：データ解析と可視化

## 第1回 イントロダクション

伊藤 岳

人間文化研究機構／富山大学  
富山大学 経済学部 2017 年度後期



Email: [gito@eco.u-toyama.ac.jp](mailto:gito@eco.u-toyama.ac.jp)

October 2, 2017

# Outline

- 1 ガイダンス
- 2 履修上の注意
- 3 アンケートと次回講義までの課題

ガイダンス



# 講義の概要と目的

## 一般学習目標

- ▶ 基礎的な計量経済・統計学の理論を学んだ (あるいは, これから本格的に学びたい) 学生を主な対象として, 計量分析の**応用的な**講義を提供する
- ▶ 具体的には, フリーの強力なソフトウェア/統計処理言語である R を用いたデータの可視化・解析の方法を学び, さらに統計的シミュレーションや現実のデータの解析を通じて統計的推論上の問題 (たとえば, 欠落変数バイアス) を理解する
- ▶ あわせて, ときとして煩雑な作業が必要になるデータ取得や統計的シミュレーション, データ解析・可視化の作業を R によって自動化・再現可能にする方法も学び, 自ら計量分析を用いる素地を身につける

# 講義の概要と目的

## 達成目標

この講義では、初歩的な計量経済・統計学の知識を前提に、履修者が次のような応用的な経験・知識・スキルを身につけることを目指す。

- ▶ データの可視化・解析のスキルを身に付ける —— 基礎的なデータの可視化および計量経済・統計学の推定を、R言語を用いて行なう方法・スキルを身につける
- ▶ 基礎的な推定法、仮定、概念を理解する —— 統計的シミュレーションと経験的な(現実の)データの解析を通して、基礎的な計量経済・統計学の推定法を理解する
- ▶ 再現可能な研究を行なう力を身につける —— R言語による実装・プログラミングを通して、再現可能な研究(データの取得・整理・可視化・解析の一連の作業を、他者が追試・確認・再現可能な研究)を行なう方法を身につける

# 授業計画

## 授業の進め方

- ▶ 毎回の授業は、講義と履修者の個人 PC を用いたハンズオン (実習) を交互に行なう
- ▶ シラバスに記載の通り、個人のノート PC 持ち込みと、学内無線 LAN への接続が前提 (各自第 2 回講義までに設定しておくこと)

## 講義の内容と計画

- ▶ ざっくりといえば、(1) R の導入とデータの可視化 (作図)、(2) (プログラミング的な) データ取得と、回帰分析の基礎、(3) 発展的な回帰分析、の 3 部構成
- ▶ 詳細はシラバスを参照
- ▶ ただし、履修者にあわせて適宜内容を変更する (今日アンケートをとります)

# 休講予定

## 次の日程は休講

- ▶ 10/30 (Mon.)
- ▶ 1/9 (Tue.) ← 月曜講義日
- ▶ 補講については講義の進捗具合をみながら、今後アナウンスする

# 教科書と副読本

## 教科書

- ▶ Gelman, A & J Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press
- ▶ 星野 匡郎・田中 久稔. 2016. 『Rによる実証分析：回帰分析から因果分析へ』 オーム社
- ▶ Imai, K. 2017. *Quantitative social science: An introduction*. Princeton, NJ: Princeton University Press (近刊の邦訳でも可)
  - ▶ 毎回の講義では、上記の教科書に加えて以下の副読本からも適宜課題文献(次回講義までに読んでおく章)を割り当てる

## 副読本

- ▶ シラバス参照



# 評価

## 中間課題 2 回と最終課題 (予定)

- ▶ 中間課題 2 回 ( $20 \times 2 = 40\%$ ) と最終課題 (60%) による。課題の詳細は講義時に指示するが、いずれも R のコードと出力 (解析・可視化) 結果の報告・解釈を求める
- ▶ 当然ながら、正当な理由 (e.g., インフルエンザ) なく締切を過ぎた提出物は一切受け付けない

- ▶ 講義で用いるスライド、R コード等の資料は、講義ウェブページにアップロードする
  - ▶ URL: [http://cfes-project.eco.u-toyama.ac.jp/education/education\\_2017/r\\_2017/](http://cfes-project.eco.u-toyama.ac.jp/education/education_2017/r_2017/)
- ▶ **原則、紙媒体では配布しない**
- ▶ 講義前日 19:00 までに、上記のウェブページに資料をアップロードするので、各自ダウンロードすること

# 履修上の注意

---

# 計量経済学／統計学の「応用」編科目

## 応用編科目

- ▶ 計量経済学／統計学の理論や推定法の説明・講義は (最低限の説明はするが) 大幅に省略する
- ▶ 平たくいえば, 自分で「計量分析をできるようにすること」や, 「重要な概念 (e.g., バイアス) を, 手を動かして深く理解すること」を目指す (= 「**応用**」編の講義)
- ▶ シラバスにも記載の通り, 計量経済学／統計学の関連科目を履修していない学生の履修も妨げないが, 並行して他の科目を履修することを強く勧める

## 想定している履修者

- ▶ 「計量経済学の初級科目は履修したので, 実際に手を動かして理解を深めたい, 自分で可視化・解析をしてみたい」
- ▶ 「計量経済学の科目は履修しことがないが, どんなことができるかまずやってみたい」

## この講義では、R (言語) を用いる

- ▶ オープンソースの統計解析向けプログラミング言語
- ▶ この講義では、R 言語とその統合開発環境 (IDE) である RStudio を用いて、講義・実習を進める
- ▶ エクセルのような (恐らく馴染みのある) ソフトウェアとは異なり、R の操作はコード記述・実行で行なう (クリック操作はできない)

## 利点と注意点

- ▶ プログラミング言語において日本語を扱うには追加的かつ煩雑な作業が必要なため、言語環境は原則英語 (ファイル名にも日本語を使わないこと)
- ▶ R は一見敷居が高いように思えるが、データの入手・可視化・実証分析の「標準言語」となりつつある (大学院に進みたい場合は必須)
- ▶ パッケージ (拡張機能) を追加することで、最新の推定方法も実装できる
- ▶ オープン・ソースであり、利用者コミュニティが発展 (e.g., Stack Overflow)

アンケートと次回講義までの課題

---

# 次回講義 (10/16) の注意点と課題

## 注意点と持ち物

- ▶ 10/9 (Mon.) は授業なし (祝日) なので、次回は 10/16 (Mon.)
- ▶ **ラップトップ (ノート PC) を持ち込むこと**
  - ▶ R, RStudio とパッケージをインストールします
  - ▶ 管理者権限のあるアカウントとパスワードを把握しておくこと
  - ▶ OS は Linux, macOS, Windows のいずれか

## 次回講義までの課題

- ▶ 次の単語の意味を把握しておくこと：「全角文字」「半角文字」「行」「列」「ファイルパス」「ディレクトリ」「拡張子」「文字コード/エンコーディング」
- ▶ 次のようなショートカットキーの表記が、どのようなキーボード操作を意味するのかを把握し、行えるようにしておくこと：“Command + S,” “Command + Shift + K” (macOS), “Control + C,” “Control + Shift + N” (Windows)
- ▶ 各自の OS 環境で任意のファイルのパスを取得する方法を調べ、実践しておくこと (任意のファイルのパスを取得できるように練習しておくこと)

# アンケート

手元のアンケート用紙に回答し提出してください

- ▶ 氏名・学籍番号・学部・専攻
- ▶ 履修予定 (必ず取る, これから決める, とか)
- ▶ 計量経済学関連科目の履修経験
  - ▶ 覚えている限りで科目名・担当教員名
  - ▶ 次の言葉の中で, 勉強したことがあるものを明示 (説明できなくても可, 全部分からなくても可)
  - ▶ 要約/記述統計, 最小二乗法 (OLS), 最尤法 (MLE), 一般化線形モデル (GLM), 交差項 (interaction term), 内生性 (endogeneity), 欠落変数バイアス (omitted variable bias), 同時性バイアス (simultaneity bias), 測定誤差 (measurement error), 選択バイアス (selection bias)
- ▶ 使用予定の PC の OS
  - ▶ macOS Sierra や Windows 10 のようなざっくりとした表記で可 (バージョン等は不要)
- ▶ プログラミングの経験
  - ▶ 言語はなんでも
  - ▶ 経験があれば, 言語名と使用期間を明記