第 1 回：ガイダンス，データの整理
【教科書第 2 章】

北村　友宏

2025 年 9 月 30 日

本日の内容

1. ガイダンス
2. データの種類
3. 代表値
4. ばらつきの指標
5. 母集団と標本
6. 統計解析ソフト gretl

授業の基本情報

- 科目名：
  - 【学部（大阪公立大学）】計量経済学 2
  - 【学部（大阪市立大学）】計量経済学上級講義 2
  - 【大学院】基礎研究 D：計量分析 2
- 担当教員：北村　友宏
- 開講形式：対面授業
- 開講区分：後期
- 曜日・時限：火曜日 4 時限
- 教室：杉本キャンパス 1 号館 132 教室
- 配当年次：
  【学部】3 年，【大学院】博士前期課程 1 年

担当教員の連絡先など

- メールアドレス
  - t-kitamura@haruka.otemon.ac.jp
- オフィス・アワー
  - 授業の前後の時間に質問を受け付けます.
- ホームページ
  - https://tomkitamura.static.jp/

資料の配布・授業の連絡・課題の提出

- 資料の配布
  - 担当教員の HP にて行う．
  - HP（https://tomkitamura.static.jp/）にアクセスし，「Teaching」→「計量経済学 2（計量分析 2)」と辿れば確認できる．
- 授業の連絡・課題の提出
  - Moodle にて行う．

授業概要と到達目標

ミクロ計量経済学（＝個票データを用いた実証分析の手法）を学ぶ.

- 前半：
  - 大学中級レベルの統計的推測の理論を復習する.
  - 回帰分析による因果関係の分析の意義と限界を学ぶ.
- 後半：
  - より高度な因果推論の手法を学ぶ.

$$\Downarrow$$

各自のゼミ研究で，個票データを用いた実証分析を実践できるようになることが，この授業の到達目標.

扱う予定のトピック

- 「計量経済学 1（計量分析 1）」の復習
  - 記述統計
  - 確率分布
  - 統計的推測
  - 単回帰分析
  - 重回帰分析
- 不均一分散
- 操作変数法
- パネル・データ
- 2 値応答モデル
- マッチング法

教科書

- 田中隆一 (2015)『計量経済学の第一歩—実証分析のススメ』有斐閣.

授業中に教科書を参照してもらうことがあるので，必ず購入し，授業に持参すること.

履修済みであることが望ましい科目

- 学部科目
  - 基礎数学 A
  - 基礎数学 B
  - データサイエンス入門
  - 経済数学
  - 計量経済学入門
  - 統計解析論
  - 計量経済学 1
- 大学院科目
  - 計量分析 1

成績評価

- 提出課題 50%
  - Excel や統計解析ソフト gretl を用いたデータ分析の実習の問題を出題する．
- 期末試験 50%
  - 計量経済学の専門用語の定義を答える問題【20 点】
  - 計算・証明・分析結果の読み取り問題【80 点】

  $\Longrightarrow$ 期末試験計 100 点満点を 50 点満点に換算する．

$$\Downarrow$$

- 合格（C 以上）の最低基準：
  - すべての提出課題について，データ分析を正確に行い，期限内に提出したうえで，計量経済学の専門用語を正確にに定義できること．

期末試験

- 期末試験は，1 月 20 日（火）4 限に実施する．
- やむを得ない事情で期末試験を受験できなかった場合の追試験の申請方法は，UNIPA にて通知されるので確認すること．
- 不合格になった場合，再試験やレポートなどによる救済措置は一切行わない．

履修上の注意

- 授業では，教科書に掲載されていない内容の補足説明も行う．
  ⇒ 授業に出席せず教科書を読むだけでは，単位を取得できない可能性があるので注意すること．
- 定期的に Moodle の計量経済学 2（計量分析 2）のページにアクセスし，連絡事項や提出課題などを確認すること．
- 授業内容に関して理解できない箇所が出てきたら，担当教員に質問し，解決を試みること．

個票データ・集計データ

- 個人，家計，事業所，企業などの観測単位からなるデータを個票データ（individual data）という．
  - マイクロ・データともいう．
  - e.g., 日本に住んでいる個人それぞれの所得
- 個票データを市町村，都道府県，国などの単位で合計または平均したデータを集計データ（aggregated data）という．
  - e.g., 日本に住んでいる個人の所得の各都道府県における平均

横断面・時系列・パネルデータ

- ある 1 時点において複数の個体を観測したデータを横断面データ（cross section data）という．
  - e.g., 47 都道府県，2009 年のみ
- ある特定の個体を複数の時点にわたり，一定の時間間隔で観測したデータを時系列データ（time series data）という．
  - e.g., 大阪府のみ，1999 年～2014 年，5 年間隔
- 複数の個体を複数の時点にわたり，一定の時間間隔で観測したデータをパネル・データ（panel data）という．
  - e.g., 47 都道府県，1999 年～2014 年，5 年間隔

参考：総和記号

- e.g.,

$$\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5.$$

  - 「$\sum$」は総和記号.
  - $\sum_{i=1}^{5} x_i$ と書いた場合，$i = 1$ から出発し，i を 1 ずつ増やしながら「$i = 1$ のときの x_i」，「$i = 2$ のときの x_i」，$\cdots$ の順に足していき，$i = 5$ になった時点で足すのをやめる，という意味.

- $\sum_i x_i$ と書くこともある．この場合，「i が何番目から何番目までなのか」は文脈で判断.
- $\sum x_i$ と書くこともある．この場合，「何について何番目から何番目まで合計するのか」は文脈で判断.

総和記号の性質１：

a を定数とすると，

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i .$$

（証明）

$$\begin{aligned}\sum_{i=1}^{n} a x_i &= a x_1 + a x_2 + \cdots + a x_n \\ &= a(x_1 + x_2 + \cdots + x_n) \\ &= a \sum_{i=1}^{n} x_i . \qquad \text{（証明終）}\end{aligned}$$

※ 証明するときは，始まりと終わりを必ず明記すること．「（証明）～（証明終）」や"*Proof.*～□"など．

総和記号の性質２：

$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i.$$

$n = 2$ なら，

$$\sum_{i=1}^{2}(x_i + y_i) = \sum_{i=1}^{2} x_i + \sum_{i=1}^{2} y_i.$$

（証明）

$$\begin{aligned}\sum_{i=1}^{2}(x_i + y_i) &= x_1 + y_1 + x_2 + y_2 \\ &= x_1 + x_2 + y_1 + y_2 \\ &= \sum_{i=1}^{2} x_i + \sum_{i=1}^{2} y_i.\end{aligned}$$

（証明終）

総和記号の性質３：

a を定数とすると，

$$\sum_{i=1}^{n} a = na.$$

（証明）

$$\sum_{i=1}^{n} a = \underbrace{a + a + \cdots + a}_{n \text{ 個}}$$

$$= na. \qquad \text{（証明終）}$$

算術平均

- データがどの程度の数値になりやすいかを表す数値を代表値（representative value）という．
  - e.g., 算術平均，中央値，最頻値
  - 中央値と最頻値については，この授業では説明を省略する．

- 観測値数 n のデータ $(x_1, x_2, \cdots, x_n)$ の算術平均（arithmetic mean）は，

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \cdots + x_n).$$

分散

- 観測値数 n のデータ $(x_1, x_2, \ldots, x_n)$ の分散（variance）は,

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$
$$= \frac{1}{n}\left\{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\right\}.$$

  - 観測値のばらつきの程度を表す.
  - 分散が大きいほど観測値のばらつきが大きい.
  - 分散が小さいほど観測値のばらつきが小さい.
  - 単位は元の単位の 2 乗.

- 分散を，

$$
\begin{aligned}
s^2 &= \frac{1}{n - 1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \\
&= \frac{1}{n - 1} \left\{ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots \right. \\
&\qquad\qquad\qquad\qquad \left. + (x_n - \bar{x})^2 \right\},
\end{aligned}
$$

と定義する場合もある．
  - 観測値数 n が十分大きい場合，算術平均との差の 2 乗の合計を n で割ったバージョンの分散 σ^2 と $n - 1$ で割ったバージョンの分散 s^2 がほとんど同じ値になる．
  - s^2 は不偏性を持つ（詳細は第 3 回授業で説明）．

標準偏差

- 分散は単位が 2 乗されているため，数値の意味が分かりにくい．
  ⇒ 正の平方根をとれば単位が元に戻る．
- 分散の正の平方根を標準偏差（standard deviation）という．

- 標準偏差は,

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

  または

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

  のように計算される.
  - 観測値のばらつきの程度を表す.
  - 標準偏差が大きいほど観測値のばらつきが大きい.
  - 標準偏差が小さいほど観測値のばらつきが小さい.
  - 単位は元の単位と同じ.

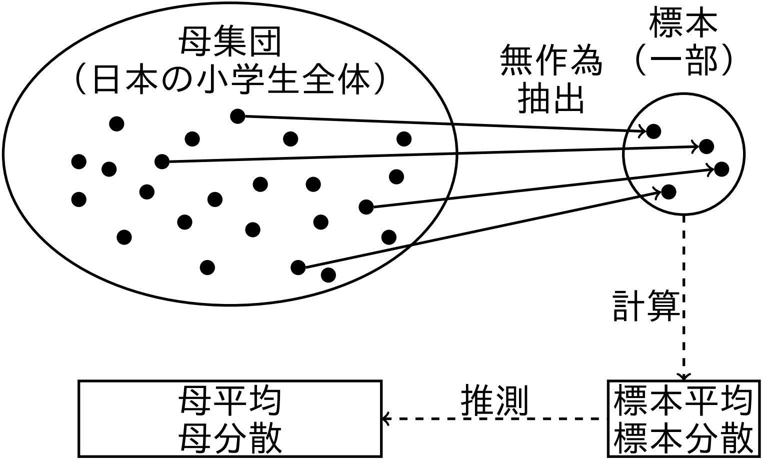
母集団と標本

- 調査対象の集団全体を母集団（population）という.
  - e.g., 日本の小学生全体
- 母集団の中から一部分を抽出して集めたものを標本（sample）という.
  - e.g., 日本の小学生全体の中から 500 人を抽出したデータ
- 母集団から標本をランダムに抽出することを無作為抽出（random sampling）という.

- 母集団全体を調査することを全数調査（census）という.
  - 悉皆調査ともいう.
  - e.g., 国勢調査
  - 費用と時間がかかり困難な場合が多い.
- 標本のみを調査することを標本調査（sample survey）という.
  - 実証研究では，これを行う場合が多い.

- 集めたデータの性質を調べることを記述統計（descriptive statistics）という．
  - e.g., ある小学校の特定のクラスの小学生 40 人の算数のテストの点数の平均や分散などを計算する．
  - 平均や分散などを計算するだけ．
- 標本から母集団の特徴を推測することを統計的推測（statistical inference）という．
  - 推測統計ともいう．
  - e.g., 日本の小学生 500 人の標本を用い，「日本の小学生全体の学力」を推測する．
  - 平均や分散などを計算し，確率論を利用して全体の性質を推測する．

統計的推測のイメージ



統計解析ソフト gretl

- 統計解析ソフト gretl は，無料でダウンロード・インストール・利用できる．
- Excel ファイルや csv ファイルのデータセットを取り込むことができる．
  - Excel ファイルについては，現行バージョンであれば xls，xlsx 両方に対応．
- 現行バージョンは日本語に対応．
- マウス操作で分析を実行する．

統計解析ソフト gretl でのデータの取り込み

最新バージョンの統計解析ソフト gretl を入手し，自分の PC にインストールする.

1. gretl の公式 HP（`https://gretl.sourceforge.net/`）にアクセス.
2. Windows の場合は「gretl for Windows」を，Mac の場合は「gretl on macOS」をクリック.

3. latest release にあるリンクをクリックしてインストールファイルを保存.
    - Windows の場合：
      最近の PC はほとんど 64bit 版なので,
      gretl-2025*-64.exe を選んでも問題ない場合が多い．自分の PC が 32bit 版であれば,
      gretl-2025*-32.exe を選ぶ．解凍ソフト（7-Zip や Lhaplus など）を持っていれば,
      gretl-2025*-win32.zip を選んでもよい.
    - Mac の場合：
      gretl-2025*-macos-arm64.pkg または
      gretl-2025*-macos-intel.pkg を選ぶ.

    ※ *には，その年に更新された順に a, b, c, d, · · · の小文字アルファベットが付される.
4. 保存したインストールファイルを実行してインストールまたは解凍.

今日のキーワード

個票データ，集計データ，横断面データ，時系列データ，パネル・データ，代表値，算術平均，分散，標準偏差，母集団，標本，無作為抽出，全数調査，標本調査，記述統計，統計的推測

次回までの準備

- 今回の講義スライドを読み直す.
- 「提出課題 1」に取り組む.
- 教科書第 3 章を読む.

付録：算術平均の性質

算術平均の性質１：

a, b を定数とすると，

$$\frac{1}{n}\sum_{i=1}^{n}(ax_i + b) = a\bar{x} + b.$$

（証明）

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(ax_i + b) &= \frac{1}{n}\sum_{i=1}^{n} ax_i + \frac{1}{n}\sum_{i=1}^{n} b \\ &= \frac{1}{n}\cdot a\sum_{i=1}^{n} x_i + \frac{1}{n}\cdot nb \\ &= a\cdot\frac{1}{n}\sum_{i=1}^{n} x_i + b \\ &= a\bar{x} + b. \qquad \text{（証明終）}\end{aligned}$$

$$\Downarrow$$

すべての観測値を a 倍して b を加えたものの算術平均は，もとの算術平均を a 倍して b を加えたものに等しい.

算術平均の性質２：

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) = 0.$$

（証明）

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n\bar{x}.$$

ここで，$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ なので，$n\bar{x} = \sum_{i=1}^{n} x_i$ である．

よって，

$$\sum_{i=1}^{n} x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0.$$

したがって，

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) = \frac{1}{n}\cdot 0 = 0. \qquad \text{（証明終）}$$

$$\Downarrow$$

すべての観測値からその算術平均を引いて，「観測値 − 算術平均」を作ると，その「観測値 − 算術平均」の算術平均は 0 になる．

付録：分散の性質

a, b を定数とし，

$$w_i = ax_i + b$$

として，σ_x^2 を $(x_1, x_2, \cdots, x_n)$ の分散，σ_w^2 を $(w_1, w_2, \cdots, w_n)$ の分散とすると，

$$\sigma_w^2 = a^2\sigma_x^2.$$

（証明）

$\bar{w}$ を w_i の算術平均とすると，

$$\sigma_w^2 = \frac{1}{n}\sum_{i=1}^{n}(w_i - \bar{w})^2.$$

ここで，$\bar{w} = a\bar{x} + b$ なので，

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(w_i - \bar{w})^2 &= \frac{1}{n}\sum_{i=1}^{n}\left\{ax_i + b - (a\bar{x} + b)\right\}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(ax_i + b - a\bar{x} - b)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{a(x_i - \bar{x})\right\}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{a^2(x_i - \bar{x})^2\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n} \cdot a^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \\
&= a^2 \cdot \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \\
&= a^2 \sigma_x^2. \qquad \text{（証明終）}
\end{aligned}
$$

$$\Downarrow$$

すべての観測値を a 倍して b を加えたものの分散は，もとの分散を a^2 倍したものに等しい.