

第 1 回：マイクロデータの探し方

北村 友宏

2020 年 10 月 2 日

本日の内容

1. ミクロデータの探し方
2. Excel でのデータの加工・整理

マイクロデータの入手方法

- ▶ 官庁ホームページからダウンロード
 - ▶ e.g., 総務省 HP, 農林水産省 HP, 国土交通省 HP, 政府統計の総合窓口 e-Stat の HP
- ▶ 図書館にて CD-ROM や紙媒体資料を利用
 - ▶ e.g., 神戸大学附属図書館（社会科学系図書館や海事科学分館）
- ▶ 研究助成室で申請

官庁 HP から入手できるマイクロデータの例

- ▶ 農林水産省『青果物卸売市場調査報告』
- ▶ 国土交通省『鉄道統計年報』
- ▶ 総務省『地方公営企業年鑑』
 - ▶ 交通，上下水道，病院等の事業者別データ
- ▶ 経済産業省資源エネルギー庁『電力調査統計』

図書館から入手できるマイクロデータの例

- ▶ 国土交通省『鉄道統計年報』
- ▶ 日本航空協会『航空統計要覧』
- ▶ 全国空港ビル協会『全国空港ターミナルビル要覧』
- ▶ 総務省『地方公営企業年鑑』

CD-ROM 収録の電子データは、社会科学系図書館の場合はメインカウンターのスタッフに申し出れば利用できる。

申請により利用できるミクロデータの例

- ▶ 日経 NEEDS Financial QUEST の，日本の上場企業財務データ
 - ▶ 経済学研究科研究助成室で申請
- ▶ 企業データベース Orbis の，世界各国の企業財務データ
 - ▶ 経営学研究科情報活用支援室で申請
- ▶ 総務省統計の匿名データ
 - ▶ 経済学研究科研究助成室で申請
 - ▶ 神戸大学ミクロデータセンター KUMiC が利用促進の取り組みをしている。

特定の個人等の識別ができないように調査票情報を加工したデータを匿名データという。

研究助成室での申請により利用できる匿名データ

- ▶ 『全国消費実態調査』
- ▶ 『社会生活基本調査』
- ▶ 『就業構造基本調査』
- ▶ 『住宅・土地統計調査』
- ▶ 『労働力調査』
- ▶ 『国勢調査』

gretl を用いた分析例

この授業では、教科書（鹿野繁樹・2015『新しい計量経済学』）に掲載されている実証分析を例に、gretl を用いた分析の解説・実習を行う。

- ▶ **データ**：国土交通省 土地総合情報システム『不動産取引価格情報検索』の、2010年第1四半期に東京都世田谷区で取引された中古マンションの価格と最寄り駅までの所要時間（物件別）
- ▶ **分析**：「駅へのアクセスのよさがマンション価値に与える影響」を分析するためのモデル（教科書 pp.111-112）

$$price_i = \beta_0 + \beta_1 minutes_i + u_i$$

の推定

gretl で出力されたモデル推定結果

gretl: モデル

ファイル 編集(E) 検定(D) 保存(S) グラフ(G) 分析(A) LaTeX

モデル 1

モデル 1: 最小二乗法(OLS), 観測: 1-194
従属変数: price_10th
不均一分散頑健標準誤差, バリエーション HCl

	係数	標準誤差	t値	p値	
const	3092.68	245.524	12.60	6.55e-027	***
minutes	74.5608	22.0194	3.386	0.0009	***

Mean dependent var 3762.577 S.D. dependent var 2150.961
Sum squared resid 8.62e+08 S.E. of regression 2118.252
R-squared 0.035207 Adjusted R-squared 0.030182
F(1, 192) 11.46597 P-value(F) 0.000860
Log-likelihood -1759.988 Akaike criterion 3523.976
Schwarz criterion 3530.512 Hannan-Quinn 3526.623

- ▶ 前スライドの結果が出力できるようになり，その結果を読み取って解釈できるようになることが目標.
- ▶ 出力結果の読み方は，後の授業で説明する.

実習 1

教科書『新しい計量経済学』に載っている、「駅へのアクセスのよさがマンション価値に与える影響」を分析するためのデータの入手・加工・整理を行う。

1. デスクトップにフォルダを作成し、名前を 2020microdatag とする。
2. 国土交通省 土地総合情報システム HP (<https://www.land.mlit.go.jp/webland/>) にアクセス。
3. 「不動産取引価格情報検索」をクリック。
4. 「時期を選ぶ」の「ダウンロードの場合 (2005年7月～直近)」の「ダウンロード」をクリック。

5. 「取引時期 From:」と「取引時期 To:」はどちらも「2010 年第 1 四半期」を選択.
6. 「都道府県:」は「東京都」を, 「市区町村:」は「世田谷区」を選択.
7. 「ダウンロード」をクリックし, 2020microdatag フォルダに保存. zip 形式で圧縮されたファイルが保存される. ファイル名が 13112_20101_20101.zip となっていることを確認.
8. 解凍ソフトを用いて 13112_20101_20101.zip を解凍. すると, 13112_20101_20101.csv が出現する.
 - ▶ Windows の場合は例えば 7-Zip, Mac の場合は例えば Keka などの解凍ソフトが必要.

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., 兵庫県のみ, 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

ミクロデータは通常, 横断面データまたはパネルデータに含まれる。

マイクロ・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**マイクロデータ (micro data)** という。
 - ▶ 個人の所得，消費
- ▶ ミクロデータを市町村，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 (aggregate data)** という。
 - ▶ 個人の所得，消費の各都道府県における平均



先ほどの実習でダウンロードした東京都世田谷区の中古マンション（物件別）のデータは、

- ▶ 横断面データ
- ▶ ミクロデータ

- ▶ 学期末に課すタームペーパーでは、各自でマイクロデータを探して入手・加工し、授業で取り上げた手法を用いて分析し、研究背景・推定モデル・結果・考察を、9 ページ以内の論文にまとめて提出する。
- ▶ 現在マイクロデータを用いた実証研究に基づく修士論文を執筆している学生は、修士論文のうち、研究背景・推定モデル・結果・考察の部分を中心に9 ページ以内にまとめたものを提出してもよい。
- ▶ 詳細は追ってアナウンスする。

- ▶ この授業では「マイクロデータを都道府県別や市町村別に集計したもの」を「広義のマイクロデータ」とする。



- ▶ タームペーパーでも、「広義のマイクロデータ」を用いてよい。
- ▶ ただし、タームペーパーで**国単位の集計データ**や、**単一の個体**（例えば兵庫県のみ）の**時系列データ**などを用いた場合、**大きく減点**され、**単位を取得できない**場合がある。

データの加工・整理方法

入手したデータは，そのままでは統計解析ソフトを用いた分析には使えない。

そこで，以下の加工・整理をする。

- ▶ Excel ファイルの 1 行目は変数名
- ▶ 2 行目は，1 番目の個体の各変数の数値
- ▶ 3 行目は 2 番目の個体，4 行目は 3 番目の個体，…
- ▶ 変数名を含め，セルは**全て半角英数字で入力**する。
 - ▶ **理由** セルに全角日本語が入力された Excel ファイルを統計解析ソフトで読み込むと文字化けするから。

加工・整理後の Excel ファイルの形

	A	B	C	D	E	F
1	id	price	minutes	year	area	onekr
2	1	6200000	5	1984	15	1
3	2	37000000	3	1999	50	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
205	204	34000000	8	1986	60	0

実習 2

1. Excel を起動し，setagayaapartment.xlsx という名前で 2020microdatag フォルダに保存.
2. セル A1 に id，セル B1 に price，セル C1 に minutes，セル D1 に year，セル E1 に area，セル F1 に onekr と入力.
3. 「セル A2 からセル A205 まで」のセル 204 個にそれぞれ 1, 2, ..., 204 の通し番号を入力.
4. setagayaapartment.xlsx を上書き保存. 本日の作業はここまで.