

第1回：マイクロデータの探し方と 加工・整理

北村 友宏

2020年5月8日

本日の内容

1. ミクロデータの探し方
2. 分析例・データの入手
3. Excel でのデータの加工・整理

マイクロデータとは

- ▶ **マイクロデータ**：個人，家計，事業所，企業などの観測単位からなるデータ

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, 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

- ▶ 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。

研究助成室での申請により利用できるミ クロデータの例

- ▶ 日経 NEEDS Financial QUEST の企業財務データ
 - ▶ 経済学研究科研究助成室で申請
- ▶ 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
 - ▶ 神戸大学マイ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、以下のデータを用いた実証分析を例に、gretl を用いた分析の解説・実習を行う。

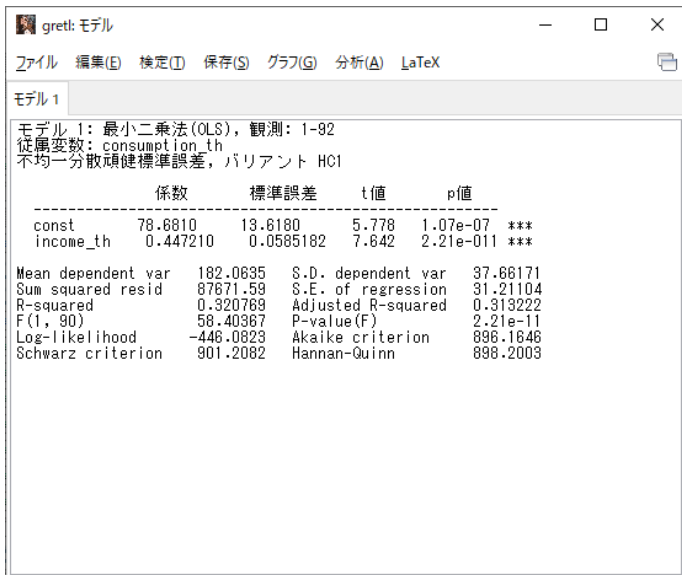
- ▶ **データ**：総務省『全国消費実態調査』2009年版の単身勤労世帯の1ヶ月間の可処分所得と消費支出（都道府県別・男女別平均）
 - ▶ これは厳密にはマイクロデータではなく**集計データ**だが、性質や扱い方がマイクロデータとほぼ同じ。
⇒ この授業では「マイクロデータを都道府県別や市町村別に集計したもの」を「広義のマイクロデータ」とする。
- ▶ **分析**：（ケインズ型）消費関数

$$c_i = \beta_0 + \beta_1 y_i + u_i$$

の推定

- ▶ **参考**：山本勲（2015）『実証分析のための計量経済学：正しい手法と結果の読み方』中央経済社。

gretl で出力された消費関数の推定結果



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: 1-92
従属変数: consumption_th
不均一分散頑健標準誤差, バリエーション HC1

	係数	標準誤差	t値	p値	
const	78.6810	13.6180	5.778	1.07e-07	***
income_th	0.447210	0.0585182	7.642	2.21e-011	***

Mean dependent var	182.0635	S.D. dependent var	37.66171
Sum squared resid	87671.59	S.E. of regression	31.21104
R-squared	0.320769	Adjusted R-squared	0.313222
F(1, 90)	58.40367	P-value(F)	2.21e-11
Log-likelihood	-446.0823	Akaike criterion	898.1646
Schwarz criterion	901.2082	Hannan-Quinn	898.2003

- ▶ 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

1. デスクトップにフォルダを作成し、名前を2020 ミクロデータ分析 1 とする。
 - ▶ 全角ローマ数字「I」は使わないほうが良い（文字化けするため）。
2. 政府統計の総合窓口 e-Stat の HP (<https://www.e-stat.go.jp/>) にアクセス。
3. 「統計データを探す」→「分野」→「企業・家計・経済」→「全国消費実態調査」→平成 21 年全国消費実態調査の「ファイル | 件数 更新日」で、「2,567 件 2013-03-28」と記載されているセルをクリック→家計収支編 [80 件] の「報告書非掲載表 [6 件]」と辿る。

4. 表番号 2-1 の男女, 地域別 1 世帯当たり 1 か月間の収入と支出の勤労者世帯の「EXCEL」をクリックし, 2019 ミクロデータ分析 1 のフォルダに保存. ファイル名が b134-2.xls となっていることを確認.

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。

ミクロ・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**ミクロデータ (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

すでに説明したように，この授業では「ミクロデータを都道府県別や市町村別に集計したものを」「広義のミクロデータ」とする。



先ほどの実習でダウンロードした『全国消費実態調査』2009年版の都道府県別・男女別の所得と消費のデータは、

- ▶ 横断面データ
- ▶ 広義のミクロデータ（厳密には集計データ）

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	prefecture	income	consumption	male	female
2	1	Hokkaido	227,349	155,491	1	0
3	2	Aomori	233,967	175,207	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
48	47	Okinawa	214,233	137,726	1	0
49	1	Hokkaido	207,155	172,835	0	1
50	2	Aomori	169,422	143,179	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
95	47	Okinawa	144,644	139,716	0	1

実習 2

1. Excel を起動し，消費 2009.xlsx という名前で 2019 ミクロデータ分析 1 フォルダに保存.
2. セル A1 に id，セル B1 に prefecture，セル C1 に income，セル D1 に consumption，セル E1 に male，セル F1 に female と入力.
3. 「セル A2 からセル A48 まで」のセル 47 個にそれぞれ半角で 1, 2, ..., 47 の通し番号を入力.
4. b134-2.xls を開き，シート B134-2 の「セル P11 からセル BJ11 まで」の範囲をコピー.
5. 消費 2009.xlsx のセル B2 を選択し，行列を入れ替えて貼り付け.

6. 消費 2009.xlsx の B 列の都道府県名を全てローマ字表記に直す.
 - ▶ 2 行目から順に,
Hokkaido, Aomori, Iwate, Miyagi, Akita,
Yamagata, Fukushima, Ibaraki, Tochigi,
Gumma, Saitama, Chiba, Tokyo, Kanagawa,
Niigata, Toyama, Ishikawa, Fukui,
Yamanashi, Nagano, Gifu, Shizuoka,
Aichi, Mie, Shiga, Kyoto, Osaka, Hyogo,
Nara, Wakayama, Tottori, Shimane,
Okayama, Hiroshima, Yamaguchi,
Tokushima, Kagawa, Ehime, Kochi,
Fukuoka, Saga, Nagasaki, Kumamoto,
Oita, Miyazaki, Kagoshima, Okinawa
とする.
7. 消費 2009.xlsx の「セル A2 からセル B48 まで」の範囲をコピーし, セル A49 を選択して貼り付け.

8. b134-2.xls のシート B134-2 の「セル P37 からセル BJ37 まで」の範囲をコピー.
9. 消費 2009.xlsx のセル C2 を選択し、**行列を入れ替えて**貼り付け.
10. b134-2.xls のシート B134-2 の「セル P39 からセル BJ39 まで」の範囲をコピー.
11. 消費 2009.xlsx のセル D2 を選択し、**行列を入れ替えて**貼り付け.
12. 消費 2009.xlsx の「セル E2 からセル E48 まで」の全てのセルに 1 を入力.
13. 消費 2009.xlsx の「セル F2 からセル F48 まで」の全てのセルに 0 を入力.

14. b134-2.xls のシート B134-2 (2) の「セル P37 からセル BJ37 まで」の範囲をコピー.
15. 消費 2009.xlsx のセル C49 を選択し、行列を入れ替えて貼り付け.
16. b134-2.xls のシート B134-2 (2) の「セル P39 からセル BJ39 まで」の範囲をコピー.
17. 消費 2009.xlsx のセル D49 を選択し、行列を入れ替えて貼り付け.
18. 消費 2009.xlsx の「セル E49 からセル E95 まで」の全てのセルに 0 を入力.
19. 消費 2009.xlsx の「セル F49 からセル F95 まで」の全てのセルに 1 を入力.

20. 消費 2009.xlsx を上書き保存. 本日の作業はここまで.

- ▶ このデータでは「富山県の男性」は全て欠損値 (missing value) . "-" と表示されている.
- ▶ gretl で読み込む場合, 欠損値は "-" としたままでよい.