

# 第1回：「ミクロデータ分析Ⅰ」の 復習（1）

北村 友宏

2020年10月2日

# 本日の内容

1. ミクロデータの入手
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。

# 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、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
- ▶ 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，12月頃ま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**：農林水産省『青果物卸売市場調査報告』（平成22年版）の，東京都中央卸売市場（計9市場，12ヶ月間）で取引されたみかんの数量と価格（市場別・月別）
- ▶ **分析**：みかんの需要関数

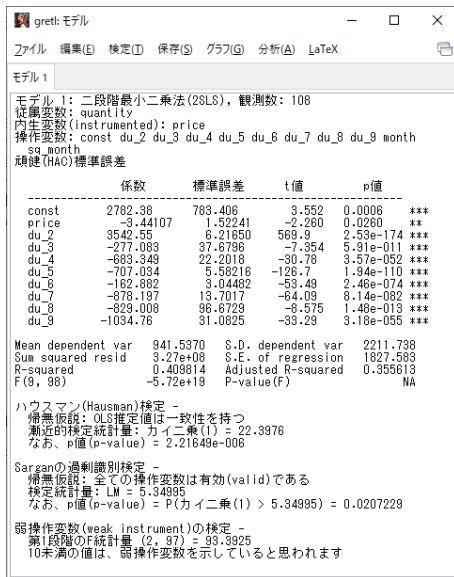
$$q_{it} = \beta_0 + \beta_1 p_{it} + u_{it}$$

の推定

- ▶ **参考**：鹿野繁樹（2015）『新しい計量経済学—データで因果関係に迫る』日本評論社．



# 出力されたみかんの需要関数の推定結果



- ▶ 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

「操作変数法」のトピックの準備も兼ねて、「マイクロデータ分析1」の復習として、みかんの需要関数を推定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2020 ミクロデータ分析 2 とする。
  - ▶ 全角ローマ数字「II」は使わないほうが良い（文字化けするため）。
2. 農林水産省 HP (<http://www.maff.go.jp/>) にアクセス。
3. 「統計情報」 → 「卸売市場に関する統計」 → 「青果物卸売市場調査」 → 「青果物卸売市場調査報告」の「調査結果データ〔Excel：e-Stat〕」の「平成22年」と辿る。
  - ▶ 青果物卸売市場調査報告は、「(産地別)」の付いていないほう。

4. 「**果実**の卸売数量・価額・価格」の表番号 **2-3** の、「東京都中央卸売市場」の「築地市場」, 「大田市場」, 「北足立市場」, 「葛西市場」, 「豊島市場」, 「淀橋市場」, 「世田谷市場」, 「板橋市場」, 「多摩ニュータウン市場」の計 **9** 市場について「EXCEL」をクリックし, 2020 ミクロデータ分析 2 フォルダに保存. ファイル名が g002-22-127.xls から g002-22-135.xls までの連番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。

# 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『青果物卸売市場調査報告』（平成 22 年版）の，東京都中央卸売市場（計 9 市場，12 ヶ月間）で取引されたみかんの数量と価格（市場別・月別）のデータは，

- ▶ パネルデータ
- ▶ ミクロデータ

- ▶ 期末レポートでは、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、結果およびその解釈をレポートとしてまとめてもらう予定。
- ▶ 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。



- ▶ 期末レポートでも、「広義のマイクロデータ」を用いてよい。
- ▶ ただし、期末レポート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
1	id	name	month	quantity	price
2	1	Tsukiji	1	2118	165
⋮	⋮	⋮	⋮	⋮	⋮
7	1	Tsukiji	6	80	1012
⋮	⋮	⋮	⋮	⋮	⋮
13	1	Tsukiji	12	3848	270
14	2	Ota	1	8281	173
⋮	⋮	⋮	⋮	⋮	⋮
19	2	Ota	6	630	904
⋮	⋮	⋮	⋮	⋮	⋮
25	2	Ota	12	15913	268

## 実習 2

1. Excel を起動し， orangetokyo.xlsx という名前で 2020 ミクロデータ分析 2 フォルダに保存.
2. セル A1 に id, セル B1 に name, セル C1 に month, セル D1 に quantity, セル E1 に price と入力.
3. 「セル A2 からセル A13 まで」のセル 12 個全てに 1 と入力.
4. 「セル A14 からセル A25 まで」のセル 12 個全てに 2 と入力.
5. 「セル A26 からセル A37 まで」のセル 12 個全てに 3 と入力.
6. 「セル A38 からセル A49 まで」のセル 12 個全てに 4 と入力.

7. 「セル A50 からセル A61 まで」のセル 12 個全てに 5 と入力.
8. 「セル A62 からセル A73 まで」のセル 12 個全てに 6 と入力.
9. 「セル A74 からセル A85 まで」のセル 12 個全てに 7 と入力.
10. 「セル A86 からセル A97 まで」のセル 12 個全てに 8 と入力.
11. 「セル A98 からセル A109 まで」のセル 12 個全てに 9 と入力.

12. 「セル B2 からセル B13 まで」のセル 12 個全てに Tsukiji と入力.
13. 「セル B14 からセル B25 まで」のセル 12 個全てに Ota と入力.
14. 「セル B26 からセル B37 まで」のセル 12 個全てに KitaAdachi と入力.
15. 「セル B38 からセル B49 まで」のセル 12 個全てに Kasai と入力.
16. 「セル B50 からセル B61 まで」のセル 12 個全てに Toshima と入力.

17. 「セル B62 からセル B73 まで」のセル 12 個全てに Yodobashi と入力.
18. 「セル B74 からセル B85 まで」のセル 12 個全てに Setagaya と入力.
19. 「セル B86 からセル B97 まで」のセル 12 個全てに Itabashi と入力.
20. 「セル B98 からセル B109 まで」のセル 12 個全てに TamaNewTown と入力.

21. 「セル C2 からセル C13 まで」のセル 12 個にそれぞれ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 の通し番号を入力.
22. 「セル C14 からセル C25 まで」のセル 12 個にそれぞれ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 の通し番号を入力.
23. 「セル C26 からセル C37 まで」のセル 12 個にそれぞれ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 の通し番号を入力.
24. 「セル C38 からセル C49 まで」のセル 12 個にそれぞれ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 の通し番号を入力.
25. 「セル C50 からセル C61 まで」のセル 12 個にそれぞれ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 の通し番号を入力.

26. 「セル C62 からセル C73 まで」のセル 12 個にそれぞれ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 の通し番号を入力.
27. 「セル C74 からセル C85 まで」のセル 12 個にそれぞれ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 の通し番号を入力.
28. 「セル C86 からセル C97 まで」のセル 12 個にそれぞれ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 の通し番号を入力.
29. 「セル C98 からセル C109 まで」のセル 12 個にそれぞれ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 の通し番号を入力.
30. orangetokyo.xlsx を上書き保存. 本日の作業はここまで.