第２回：「ミクロデータ分析Ｉ」の復習（2）

北村　友宏

2020 年 10 月 9 日

本日の内容

1. Excel でのデータの加工・整理
2. gretl でのデータの取り込み
3. gretl での記述統計の出力

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 1

1. 前回作成した，orangetokyo.xlsx を開く．
2. まず，築地市場で取引されたみかんの数量と価格のデータを，元のデータから「orangetokyo.xlsx」にコピー・貼り付けする．
   前回ダウンロードした g002-22-127.xls を開き，「セル J18 からセル K18 まで」の範囲（みかんの 1 月の数量と価格）をコピー．
3. orangetokyo.xlsx のセル D2 を選択し，貼り付け．
4. g002-22-127.xls の「セル L18 からセル M18 まで」の範囲（みかんの 2 月の数量と価格）をコピー．
5. orangetokyo.xlsx のセル D3 を選択し，貼り付け．

6. g002-22-127.xls の「セル N18 からセル O18 まで」の範囲（みかんの 3 月の数量と価格）をコピー.
7. orangetokyo.xlsx のセル D4 を選択し，貼り付け.
8. g002-22-127.xls の「セル P18 からセル Q18 まで」の範囲（みかんの 4 月の数量と価格）をコピー.
9. orangetokyo.xlsx のセル D5 を選択し，貼り付け.
10. g002-22-127.xls の「セル R18 からセル S18 まで」の範囲（みかんの 5 月の数量と価格）をコピー.
11. orangetokyo.xlsx のセル D6 を選択し，貼り付け.

12. g002-22-127.xls の「セル T18 からセル U18 まで」の範囲（みかんの 6 月の数量と価格）をコピー.
13. orangetokyo.xlsx のセル D7 を選択し，貼り付け.
14. g002-22-127.xls の「セル V18 からセル W18 まで」の範囲（みかんの 7 月の数量と価格）をコピー.
15. orangetokyo.xlsx のセル D8 を選択し，貼り付け.
16. g002-22-127.xls の「セル X18 からセル Y18 まで」の範囲（みかんの 8 月の数量と価格）をコピー.
17. orangetokyo.xlsx のセル D9 を選択し，貼り付け.

18. g002-22-127.xls の「セル Z18 からセル AA18 まで」の範囲（みかんの 9 月の数量と価格）をコピー.
19. orangetokyo.xlsx のセル D10 を選択し，貼り付け.
20. g002-22-127.xls の「セル AB18 からセル AC18 まで」の範囲（みかんの 10 月の数量と価格）をコピー.
21. orangetokyo.xlsx のセル D11 を選択し，貼り付け.
22. g002-22-127.xls の「セル AD18 からセル AE18 まで」の範囲（みかんの 11 月の数量と価格）をコピー.
23. orangetokyo.xlsx のセル D12 を選択し，貼り付け.

24. g002-22-127.xls の「セル AF18 からセル AG18 まで」の範囲（みかんの 12 月の数量と価格）をコピー.
25. orangetokyo.xlsx のセル D13 を選択し，貼り付け.
26. 続いて，大田市場で取引されたみかんの数量と価格のデータを，元のデータから「orangetokyo.xlsx」にコピー・貼り付けする. 前回ダウンロードした g002-22-128.xls を開き，「セル J18 からセル K18 まで」の範囲（みかんの 1 月の数量と価格）をコピー.
27. orangetokyo.xlsx のセル D14 を選択し，貼り付け.

28. 2 月から 12 月についても，先ほどの g002-22-127.xls と同様の作業を行い，それぞれの月の数量と価格の数値を「orangetokyo.xlsx」にコピー・貼り付け.
29. g002-22-129.xls から g002-22-135.xls についても同様の作業を行い，多摩ニュータウン市場までの数量と価格の数値を「orangetokyo.xlsx」にコピー・貼り付け.
    - 北足立市場（KitaAdachi）は 26 行目から 37 行目，葛西市場（Kasai）は 38 行目から 49 行目，豊島市場（Toshima）は 50 行目から 61 行目，淀橋市場（Yodobashi）は 62 行目から 73 行目，世田谷市場（Setagaya）は 74 行目から 85 行目，板橋市場（Itabashi）は 86 行目から 97 行目，多摩ニュータウン市場（TamaNewTown）は 98 行目から 109 行目.
30. orangetokyo.xlsx を上書き保存して閉じる.

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．

実習 1

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．
※今年度前期の「ミクロデータ分析 I」など，gretl を使う他の授業科目を受講しており，すでに自分の PC に gretl をインストールしていても，2020 年 8 月 6 日以前にダウンロード・インストールした場合は再度，最新バージョンをダウンロードし，再インストールすること．

1. gretl の公式 HP（http://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．

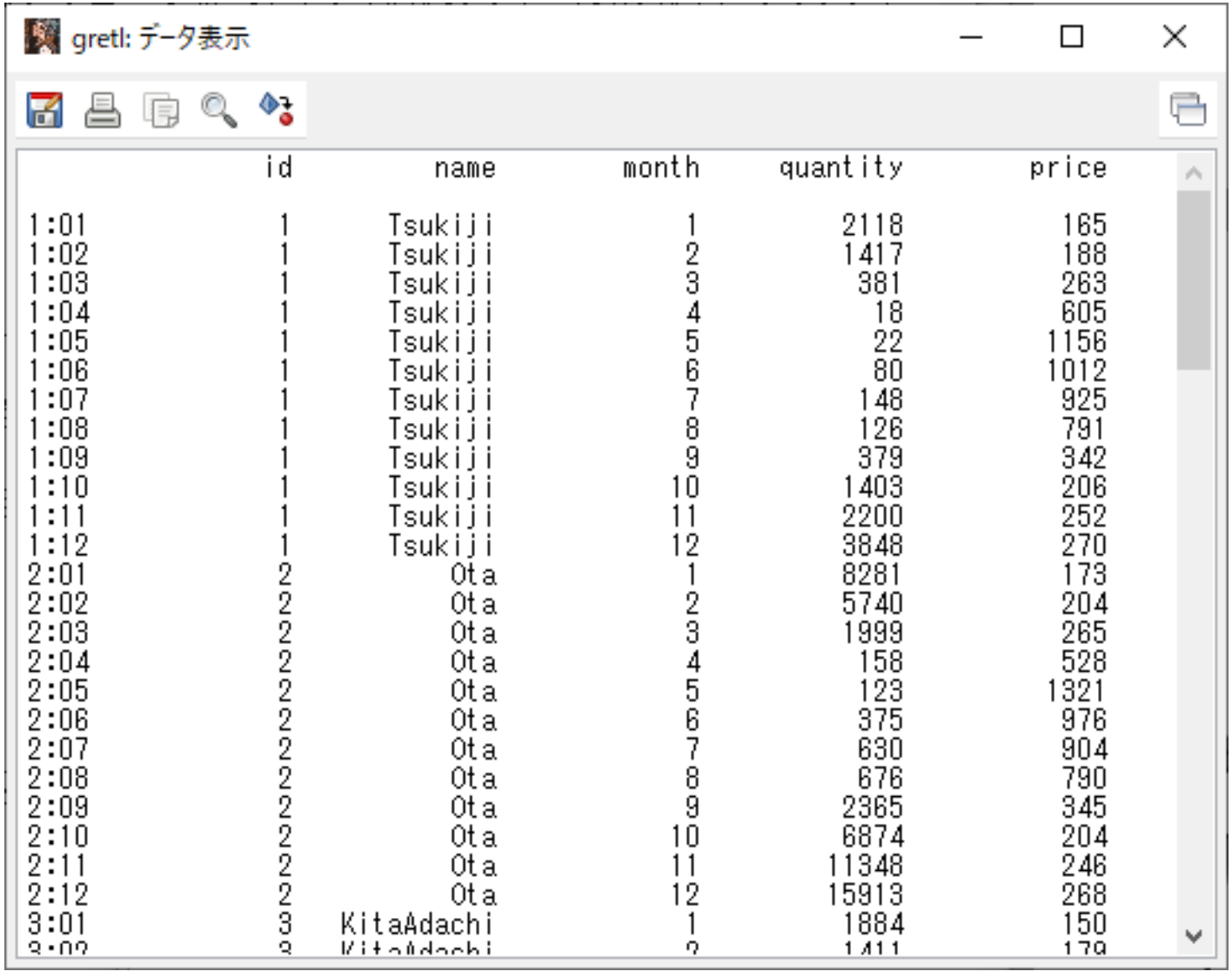
3. latest release にあるリンクをクリックしてインストールファイルを保存.
   - Windows の場合：
     最近の PC はほとんど 64bit 版なので,
     gretl-2020d-64.exe を選んでも問題ない場合が多い．自分の PC が 32bit 版であれば,
     gretl-2020d-32.exe を選ぶ．解凍ソフト（7-Zip や Lhaplus など）を持っていれば,
     gretl-2020d-win32.zip を選んでもよい.
   - Mac の場合：
     gretl-2020d-quartz.pkg を選ぶ.
4. 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.

実習 2

1. 先ほどの実習でインストールした gretl を起動.
2. orangetokyo.xlsx を，gretl の画面にドラッグ・アンド・ドロップ.
3. 出てきたダイアログボックスの，インポートを開始する場所: の列: と行: がともに 1 になっていることを確認し，「OK」をクリック.
4. 「インポート可能なシートを 1 個見つけました」で始まるメッセージが表示されるので，「閉じる」をクリックすると，データが読み込まれる.

5. 「インポートされたデータは・・・(中略)・・・解釈し直しますか？」というメッセージが表示されるので，「はい」をクリック．
6. 出てきたダイアログボックスの選択肢のうち，「パネル」をクリックして選択し，「進む」をクリック．
   - 作成した orangetokyo.xlsx は複数個体（市場）・複数時点（月）のパネルデータ．
7. 「インデックス変数を使用する」をクリックして選択し，「進む」をクリック．
8. ユニット（グループ）インデックス変数は「id」を，タイム・インデックス変数は「month」を選び，「進む」をクリック．
   - orangetokyo.xlsx において，変数「id」は市場番号を，変数「month」は時点番号（月）を表す．

9. 「パネルデータ (時系列データを重ねた構造)
9 個のクロスセクション・ユニットが、%d 期観測されたデータ」と表示されていることを確認し，「適用」をクリックすると，データが読み込まれる.
   - 「%d」が文字化けしているが，読み込みに支障はない.
10. 「id」から「price」までの 5 つをドラッグして選択し，その上で右クリック→「データ（値）を表示」と操作すると，全変数の観測値リストが新規ウィンドウにて表示される.



	id	name	month	quantity	price
1:01	1	Tsukiji	1	2118	165
1:02	1	Tsukiji	2	1417	188
1:03	1	Tsukiji	3	381	263
1:04	1	Tsukiji	4	18	605
1:05	1	Tsukiji	5	22	1156
1:06	1	Tsukiji	6	80	1012
1:07	1	Tsukiji	7	148	925
1:08	1	Tsukiji	8	126	791
1:09	1	Tsukiji	9	379	342
1:10	1	Tsukiji	10	1403	206
1:11	1	Tsukiji	11	2200	252
1:12	1	Tsukiji	12	3848	270
2:01	2	Ota	1	8281	173
2:02	2	Ota	2	5740	204
2:03	2	Ota	3	1999	265
2:04	2	Ota	4	158	528
2:05	2	Ota	5	123	1321
2:06	2	Ota	6	375	976
2:07	2	Ota	7	630	904
2:08	2	Ota	8	676	790
2:09	2	Ota	9	2365	345
2:10	2	Ota	10	6874	204
2:11	2	Ota	11	11348	246
2:12	2	Ota	12	15913	268
3:01	3	KitaAdachi	1	1884	150
3:02	3	KitaAdachi	2	1411	179

このような画面が表示されれば成功．確認したら閉じる．

※もし数字が違っていたら，データセット（orangetokyo.xlsx）の作成の際にミスをしているということなので，前回の講義スライドを参照してデータセットの作成からやり直すこと.

11. メニューバーから「ファイル」→「データに名前を付けて保存」と操作し，orangetokyo.gdt という名前で「2020 ミクロデータ分析 2」フォルダに保存.

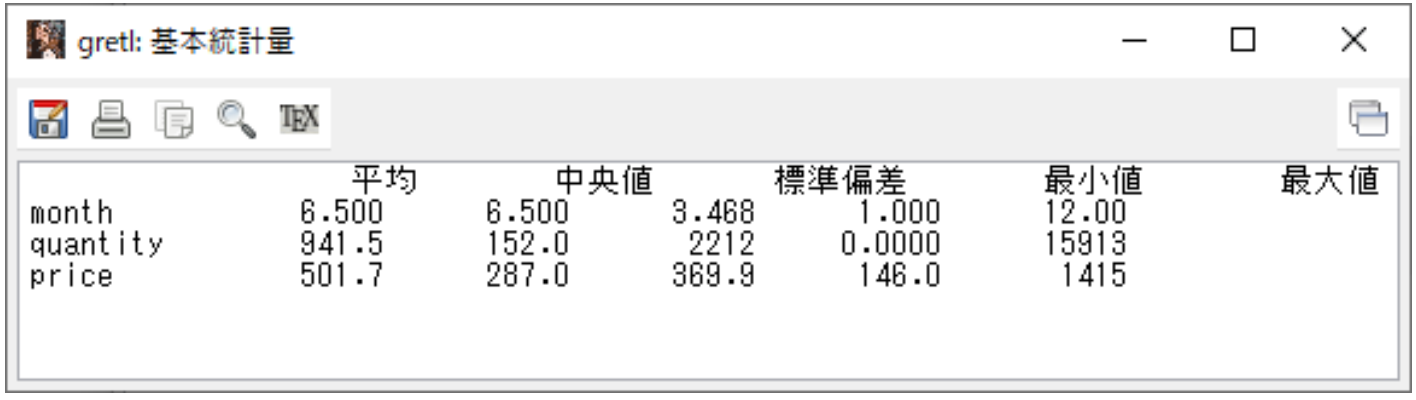
記述統計

- データセットを読み込んだ gretl の画面上で，記述統計を出力したい変数を選択し，右クリック→「基本統計量」と操作し，「主要な統計量を表示する」が選ばれている状態で「OK」をクリックすると，選んだ変数の，平均 (mean)，中央値 (median)，標準偏差 (standard deviation)，最小値 (minimum)，最大値 (maximum) が表示される．
  - 「記述統計」は，「基本統計量」や「要約統計量」ともいう．

- 平均
  - $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$
- 中央値
  - 観測値を小さい順に並べたときに中央に来る値.
  - 観測値数 n が偶数の場合は中央で隣り合う 2 つの値の平均値.
- 標準偏差
  - $s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}.$
- 最小値
  - $\min\{x_i\}.$
- 最大値
  - $\max\{x_i\}.$

実習 3

1. 「month」から「price」までの 3 つをドラッグして選択し，その上で右クリック→「基本統計量」と操作.
2. 「主要な統計量を表示する」が選ばれている状態で「OK」をクリックすると，選択した変数の記述統計 5 種類が表示される.
   - 最新バージョン（2020 年 8 月 6 日版）では，この表示が日本語化されている.



	平均	中央値	標準偏差	最小値	最大値
month	6.500	6.500	3.468	1.000	12.00
quantity	941.5	152.0	2212	0.0000	15913
price	501.7	287.0	369.9	146.0	1415

このような画面が表示されれば成功.
Mac の PC では，小数点以下の表示桁数が異なっている場合がある.
最新バージョン（2020 年 8 月 6 日版）では，上の画像のように統計量名が全て日本語で表示される.

- 統計量の名前の位置がズレていて見づらいが,
  各変数について出力された数字は左から平均,
  中央値，標準偏差，最小値，最大値の順.

まだ作業があるので,「gretl: 基本統計量」のウィンドウはまだ閉じない！

3. 表示されている記述統計の画面上で右クリック→「名前を付けて保存...」と操作.
4. 出てきたダイアログボックスの,「標準テキスト」を選び,「OK」をクリック.
5. 記述統計 10 月 9 日.txt という名前で「2020 ミクロデータ分析 2」フォルダに保存. 本日の作業はここまで.