

第2回：データの加工・整理（1）

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

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



以下のような分析ができる。

- ▶ 最寄り駅までの所要時間が1分長くなると、中古マンションの価格は平均的に何円変化するか？
- ▶ 築年数が1年増えると、中古マンションの価格は平均的に何円変化するか？
- ▶ 面積が $1m^2$ 広がると、中古マンションの価格は平均的に何円変化するか？
- ▶ ワンルームの中古マンションはそれ以外の中古マンションに比べ、価格が平均的に何円異なるか？

実習 1

1. 前回ダウンロードした
13112_20101_20101.csv を Excel で開く.
2. 「データ」 → 「フィルター」と操作.
3. セル B1 (種類) の右下のボタンをクリックし、「宅地 (土地)」と「宅地 (土地と建物)」のチェックを外して「中古マンション等」のみチェックされている状態にする.
 - ▶ 「種類」が「中古マンション等」となっている行 (204 個の個体) のみ抽出される.
4. 前回作成した, setagayaapartment.xlsx を開く.

5. 13112_20101_20101.csv の「セル J7 からセル J508 まで」の範囲（取引価格）をコピー。
 - ▶ 抽出中の 204 個のみコピーされる。
6. setagayaapartment.xlsx のセル B2 を選択し、貼り付け。
 - ▶ コピーした抽出中の 204 個が上から順に間を空けず貼り付けされる。
7. 13112_20101_20101.csv の「セル I7 からセル I508 まで」の範囲（最寄駅：距離（分））をコピー。
 - ▶ 抽出中の 204 個のみコピーされる。
8. setagayaapartment.xlsx のセル C2 を選択し、貼り付け。
 - ▶ コピーした抽出中の 204 個が上から順に間を空けず貼り付けされる。

9. 13112_20101_20101.csv の「セル R7 からセル R508 まで」の範囲（建築年）をコピー。
 - ▶ 抽出中の 204 個のみコピーされる。
10. setagayaapartment.xlsx のセル D2 を選択し、貼り付け。
 - ▶ コピーした抽出中の 204 個が上から順に間を空けず貼り付けされる。
11. 13112_20101_20101.csv の「セル M7 からセル M508 まで」の範囲（面積 (m^2)) をコピー。
 - ▶ 抽出中の 204 個のみコピーされる。
12. setagayaapartment.xlsx のセル E2 を選択し、貼り付け。
 - ▶ コピーした抽出中の 204 個が上から順に間を空けず貼り付けされる。

13. 13112_20101_20101.csv の「セル L7 からセル L508 まで」の範囲（間取り）をコピー。
 - ▶ 抽出中の 204 個のみコピーされる。
14. setagayaapartment.xlsx のセル F2 を選択し，貼り付け。
 - ▶ コピーした抽出中の 204 個が上から順に間を空けず貼り付けされる。

15. 以下を参考にして、setagayaapartment.xlsx の D 列の、和暦の建築年を全て西暦（数字のみ、「年」は不要）に直す。

- ▶ 昭和 42 年 ⇒ 1967
- ▶ 昭和 45 年 ⇒ 1970
- ▶ 昭和 50 年 ⇒ 1975
- ▶ 昭和 55 年 ⇒ 1980
- ▶ 昭和 60 年 ⇒ 1985
- ▶ 昭和 64 年 ⇒ 1989
- ▶ 平成 2 年 ⇒ 1990
- ▶ 平成 5 年 ⇒ 1993
- ▶ 平成 10 年 ⇒ 1998
- ▶ 平成 15 年 ⇒ 2003
- ▶ 平成 20 年 ⇒ 2008
- ▶ 平成 21 年 ⇒ 2009

16. setagayaapartment.xlsx を上書き保存。本日の作業はここまで。

建築年の和暦を西暦に直す理由

- ▶ 「築年数が1年増えると、中古マンションの価格は平均的に何円変化するか？」を検証するには、**築年数**のデータが必要.
- ▶ データとして入手できたのは、築年数ではなく**建築年**.

⇒ データの観測時点（このデータは2010年）から建築年を引けば築年数を導出できる.

- ▶ 建築年は和暦として観測されている.
- ▶ 昭和時代に建築されたものと平成時代に建築されたものが混在.

⇒ 築年数導出の計算の際に不都合.

⇒ 建築年を西暦に直す.