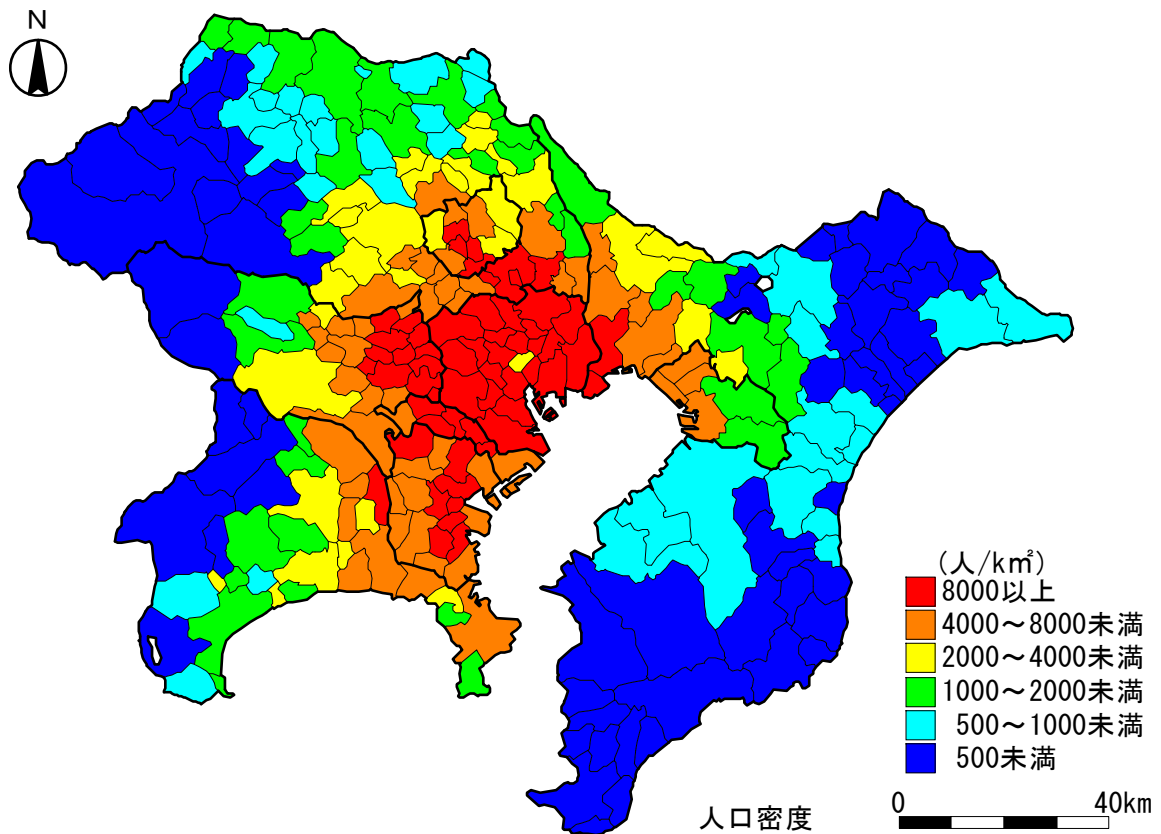


第3章 MANDARA で地図をつくろう②

一階級区分を考える一

本章では、地図データの構造と階級区分について考えていきます。たとえば、2色で色分けするのか4色で色分けするのかによって階級区分図から受ける印象や読みとれる内容に大きな違いが生じるように、階級区分は階級区分図の出来を大きく左右する重要な要素です。本章では、とくに階級区分の基本的な考え方について理解しましょう。

完成図



使用データ

地図データ：MANDARA 付属の市区町村別地図

日本市町村.MPF

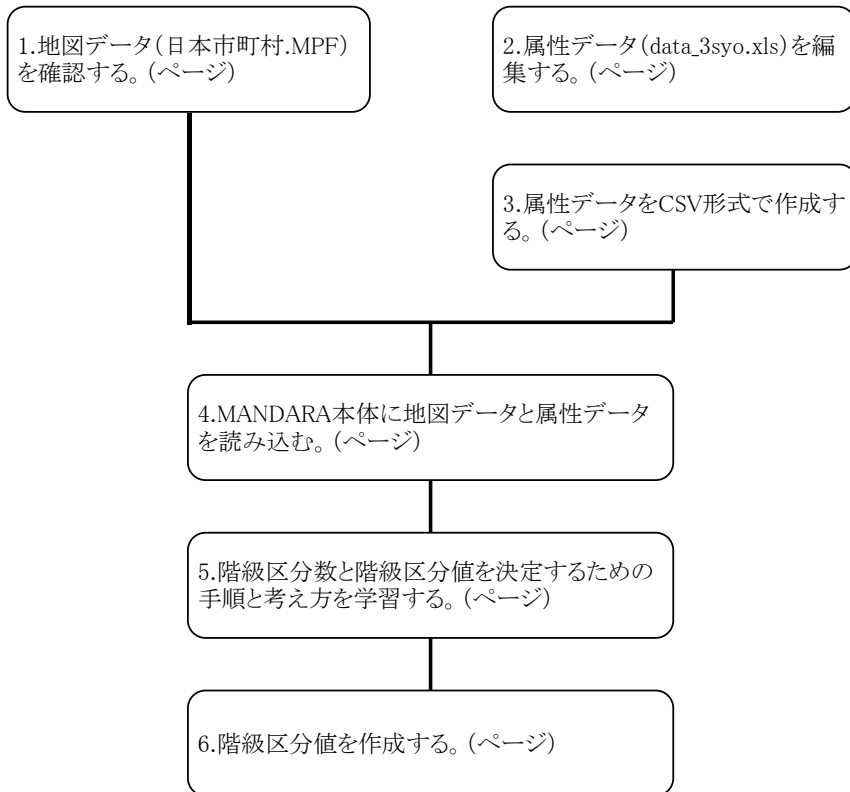
※MANDARA に付属

属性データ：南関東の市区町村別人口密度など

data_3syo.xls

※「3syo」フォルダより利用

作業の流れ



3.1 日本市町村.MPF の確認

それでは、地図データとして用いる「日本市町村.MPF」がどのような地図なのか、確認しましょう。MANDARA を起動させて下さい。マップエディタで確認します。

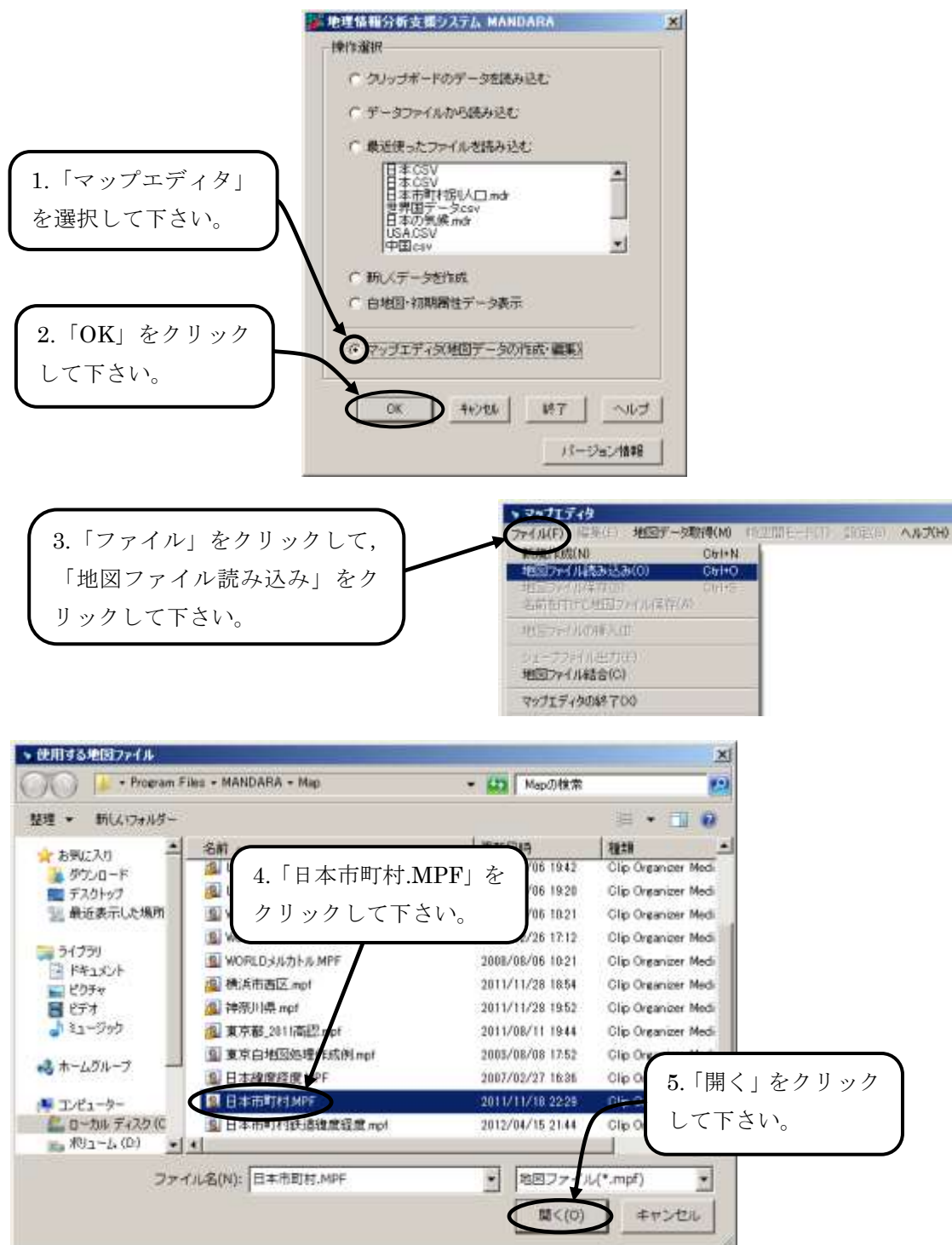


図 3.1.1 「日本市町村.MPF」をマップエディタで開く

続いて、佐渡島が画面の中央にくるように、地図を拡大しましょう。



図 3.1.2 地図の拡大とスクロール

さて、佐渡島が画面の中心に表示されているでしょうか。次に、この「日本市町村.MPF」がどのような構造を持っている地図なのか、みてみましょう。

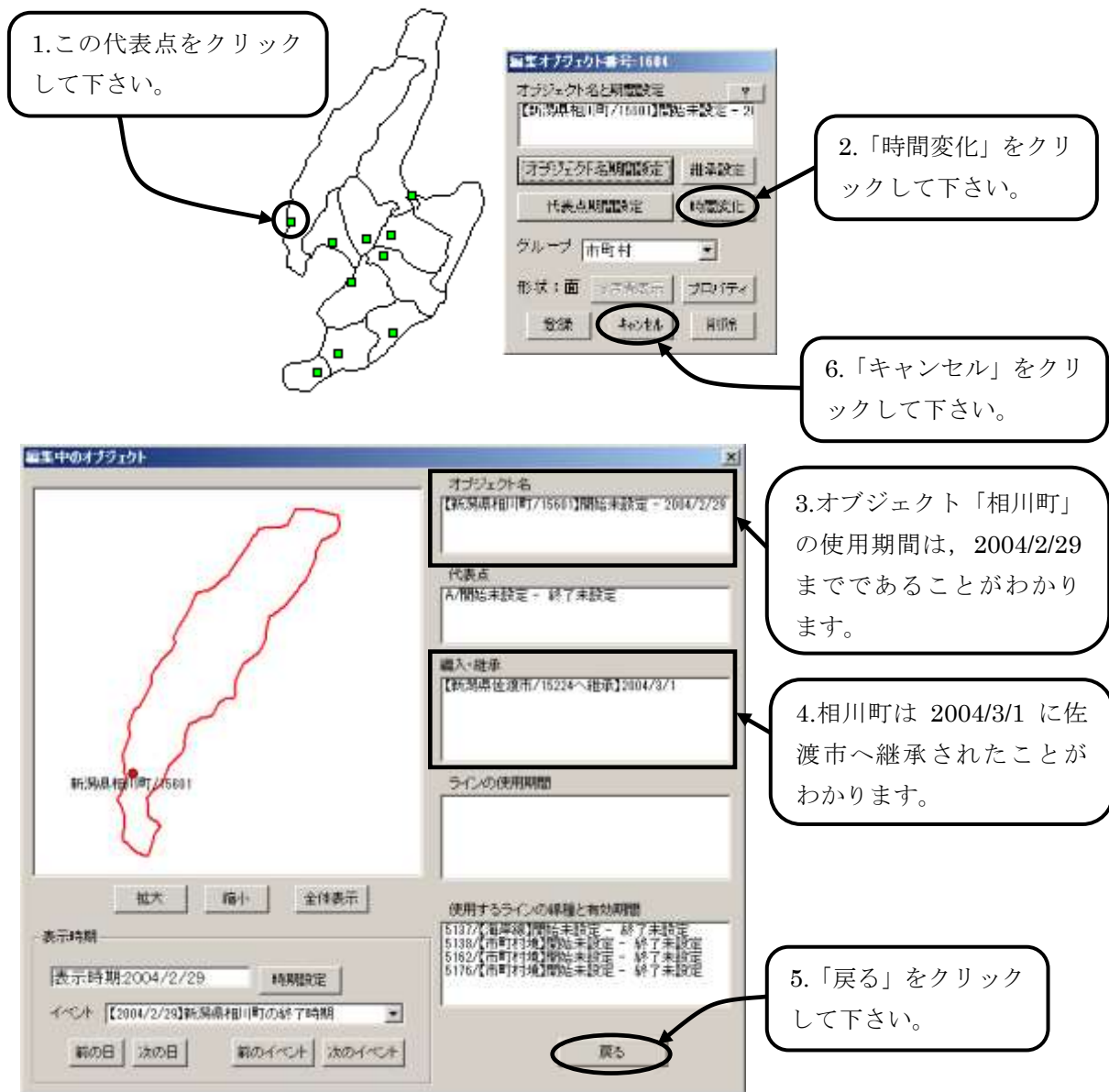


図 3.1.3 日本市町村.MPF は時間の概念を持つ

このように、「日本市町村.MPF」は、時間の概念を持つ地図であることがわかります。したがって、たとえば、「1990年の国勢調査データを用いて階級区分図を作成したい」ときも、属性データの編集の際に、時間の設定をきちんと行えば、1990年当時の市区町村で階級区分図を作成できるようになっています（1960年以降が対象）。なお、属性データ編集時の時間の設定については、このあとの3.2節で説明します。

また、今後、市町村合併が行われる可能性があります。日本市町村.MPFの最新版が必要なときは、MANDARAのホームページ (<http://ktgis.net/mandara/>) からダウンロードして利用することができます。

さて、もう少し、「日本市町村.MPF」を詳しくみてみましょう。画面上では、1枚の地図に見えますが、実は、5枚の地図（オブジェクトグループ）に分割して考えることもできます。

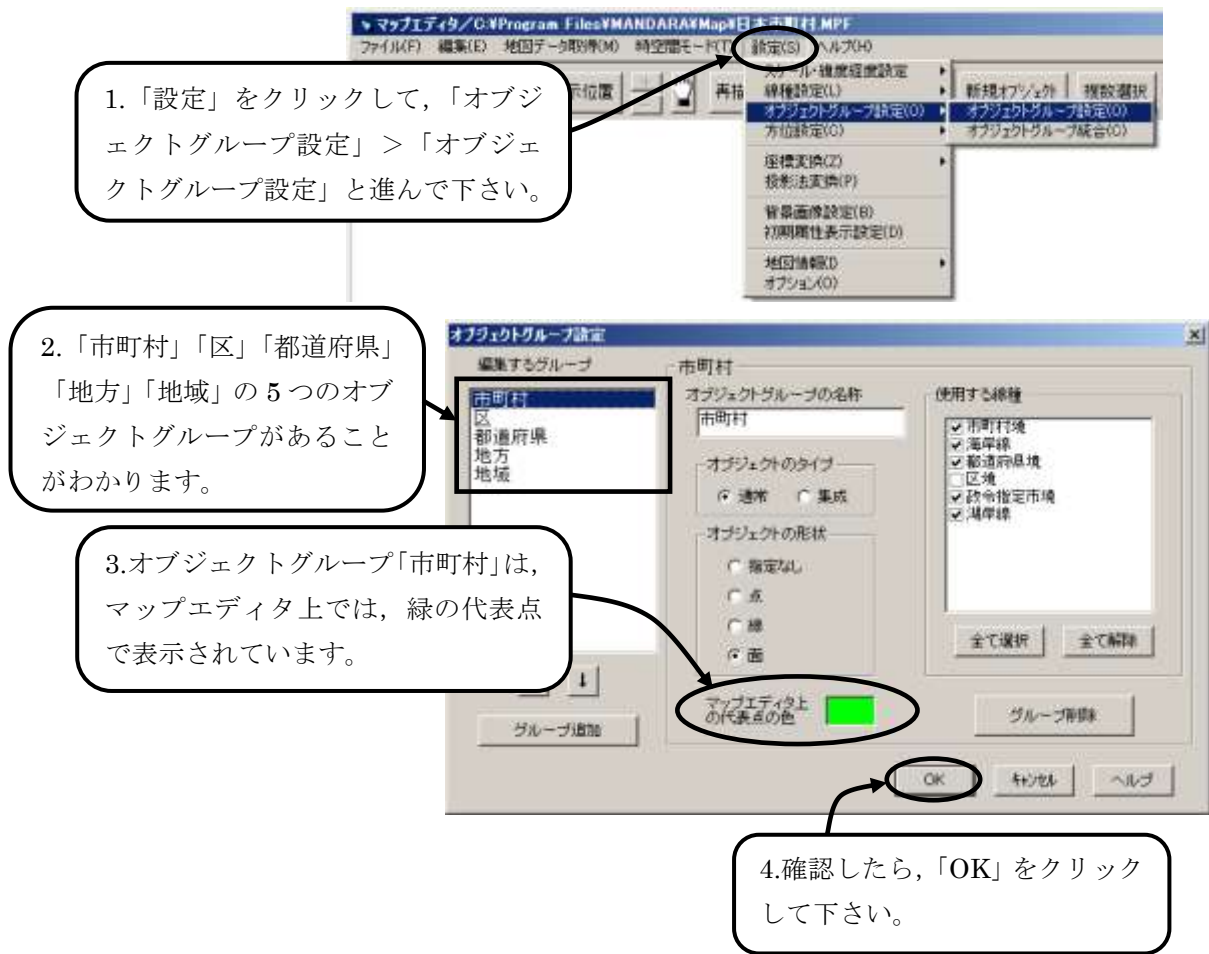


図 3.1.4 日本市町村.MPF の5つのオブジェクトグループ

それぞれのオブジェクトグループを、1つの「レイヤ」にたとえると、次のように考えることもできます。

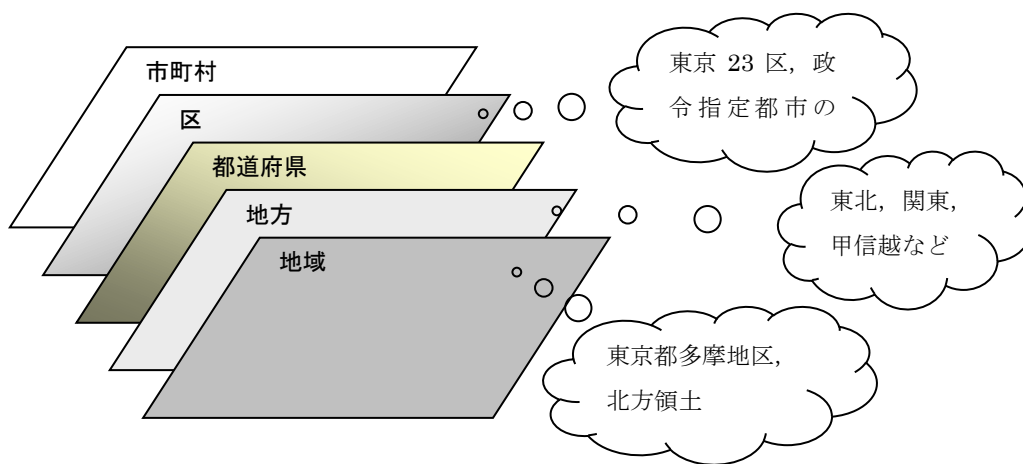


図 3.1.5 日本市町村.MPF を構成する5つのレイヤ

このように、第3章で用いる「日本市町村.MPF」は、時間の概念と5つのオブジェクトグループ(レイヤ)から構成されている地図データであることがわかります。

3.2 属性データの作成と MANDARA への読み込み

では、属性データを作成しましょう。「data_3syo.xls」を開いて、次の点を確認して下さい。

1.地図データには、「日本市町村.MPF」を指定します。

2.MANDARA タグ「TIME」は、地図に時間設定がある場合に限り指定します。

3.本章で作成する階級区分図は、2005年の国勢調査のデータを用いますので、2005年10月1日と設定します。

	A	B	C	D	E	F	G
1	MAP	日本市町村					
2	TIME	2005	10	1			
3	TITLE	人口(2005)	人口(2000)	人口増加数	人口増加率	面積	人口密度
4	UNIT	人	人	人	%	km ²	人/km ²
5	東京都			57			
6	東京都			258			
7	東京都			264			
8	東京都			188			
9	東京都文京区	189632	176017	13615	7.7	11.31	16766.8
10	東京都台東区	165186	156325	8861	5.7	10.08	16387.5

図 3.2.1 data_3syo.xls

「TIME」という新しい MANDARA タグが登場します。「TIME」は、時間設定のある地図データ (MANDARA の公式マニュアルでは、「時空間モードで作成された地図」とよんでいます。) を用いるときに限って、属性データの時期を指定するために指定します。前節で確認したとおり、「日本市町村.MPF」は、時間の概念を持つ地図データです。したがって、「TIME」タグで時間の設定をする必要があります。文法は次の通りです。

- ・ TIME 年 (西暦) 月 日
 - ※ 月日は省略可能です。省略した場合は、1月1日に自動設定されます。
 - ※ 「TIME」タグは、「LAYER」タグよりも下に、オブジェクト名よりも上に置きます。

第2章では、属性データをクリップボードにコピーして読み込む方法を紹介しましたが、CSV形式で保存したファイルを読み込むことも可能です。それでは、「data_3syo.xls」をCSV形式で保存しましょう。

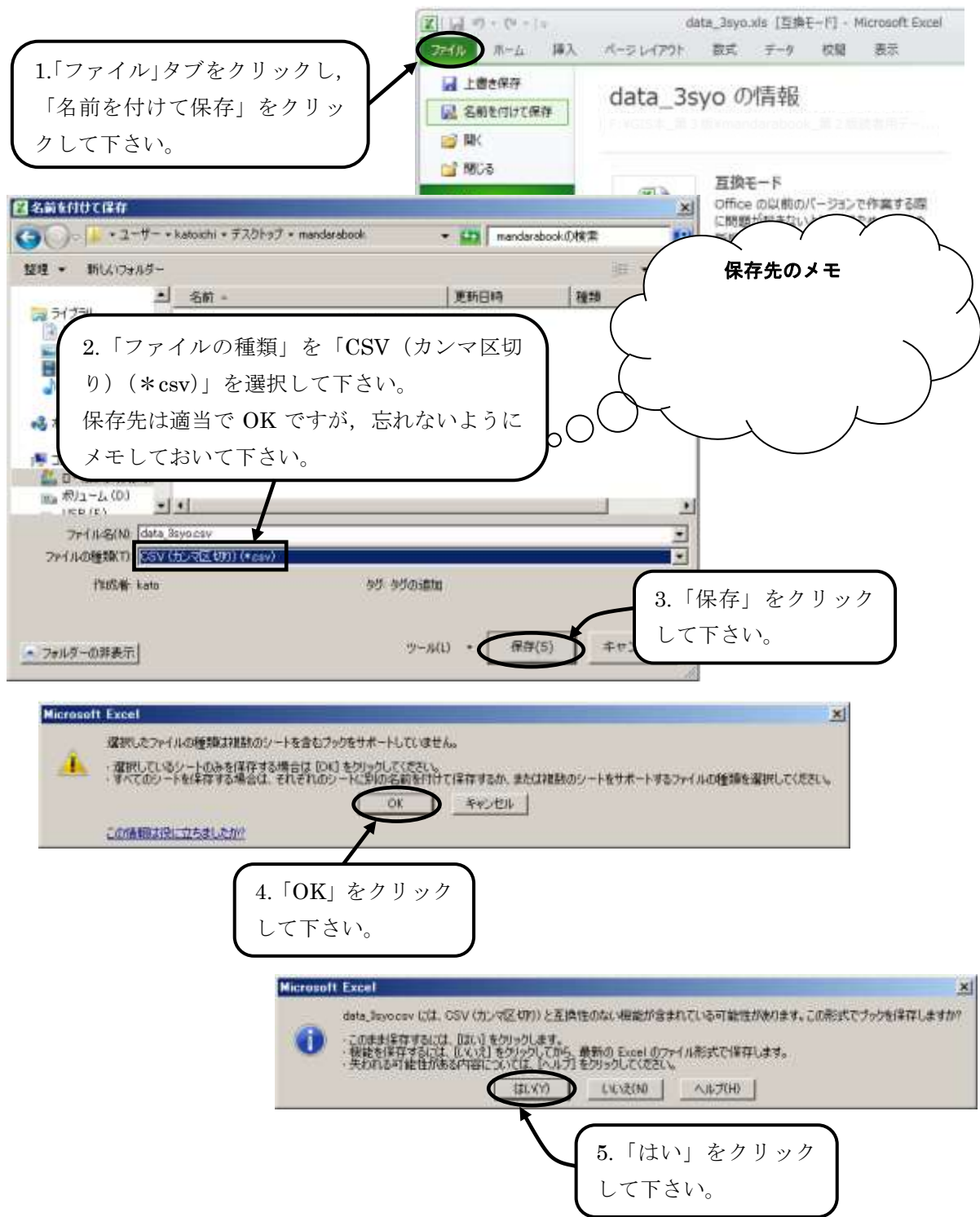


図 3.2.1 data_3syocsv.xls を CSV 形式で保存する

「CSV」とは？

「CSV」とは、個々のテキストデータをカンマ (,) で区切ってあるファイル形式のことをいいます。次ページの図は、エクセルで作成した CSV 形式のファイルを Windows の「メモ帳」で開いたものです。カンマで区切られていることを確認して下さい。……次ページに続く

次に、「data_3syo.csv」をMANDARAに読み込みましょう。MANDARAを起動して下さい。

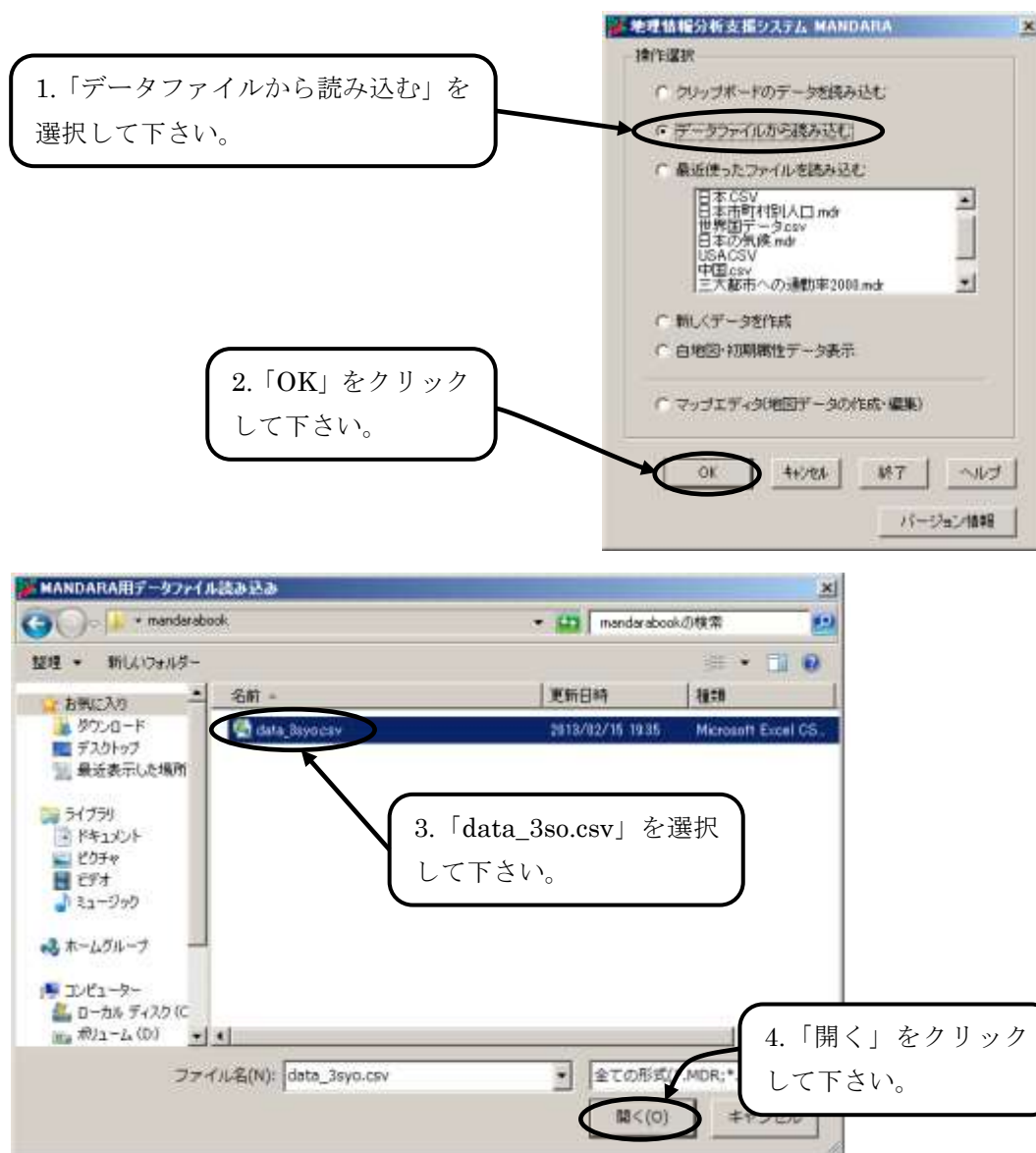


図 3.2.2 data_3syo.csv を MANDARA に読み込む

「CSV」とは？……前ページからの続き

	A	B	C	D
1	後藤	真太郎	谷	謙二
2	酒井	聡一	坪井	壺太郎

→ 後藤,真太郎,谷,謙二
酒井,聡一,坪井,壺太郎

CSV形式のファイルの編集には、通常、エクセルを利用します。エクセルはカンマをデータの区切りとして認識し、カンマで区切られた個々のデータをセルに1つずつ格納します。したがって、CSVファイルはエクセル上で、通常のエクセル形式のファイルと同様に、編集することができます。ただし、CSV形式で保存する場合、データはすべてテキストになりますので、計算結果は保存されますが、計算式は失われてしまいます。こうしたことを未然に防ぐため、CSV形式で保存する際には、注意を促すメッセージが表示されるようになっています。

3.3 階級区分を考える

本章の冒頭でもお話ししたように、階級区分によって図から受ける印象は大きく変わってきます。それでは、階級区分をどのように考えればよいのでしょうか。一般的な手順は、次の通りです。

- ① 統計データを何段階に分けるか（階級分割数）。
- ② 統計データをどの値で区切るか（階級区分値）。
- ③ どのような色（模様）を割りあてるか。・・・第2章のコラム「配色」（ページ）参照。

このなかで、最も判断が難しいのは、②の階級区分値です。そこで、本節では、階級区分値を中心に説明していきます。

3.3.1 階級分割（区分）数

それでは、まず階級分割数から検討していきましょう。菅野ほか（1987）は、「分布事象を理解しやすくするために統計値をいくつかの階級に分けて、模様あるいは色彩で表現するが、一般の読者の模様識別能力はせいぜい8～10とされているし、色彩の識別能力も11ぐらいが限度とされているので、階級の数はそれによって制約される。実際には、6階級くらいが適当であろう。また、少なくとも4階級は必要である。」りと述べています。ただし、統計データの個数や精度、階級区分図の作成目的などによって、一概にそうしなければいけないというわけではありません。

3.3.2 階級区分値

いよいよ階級区分値の検討に入ります。まず、度数分布表（または、ヒストグラム）を作成して、統計データの分布を確認してみましょう。ヒストグラムとは、度数分布表を棒グラフの形式で表したものと考えて下さい（厳密にはもう少し細かい条件があるのですが・・・）。度数分布表はエクセルでも作成できますが、MANDARAでも作成できます。ここでは、MANDARAによる方法を紹介します。

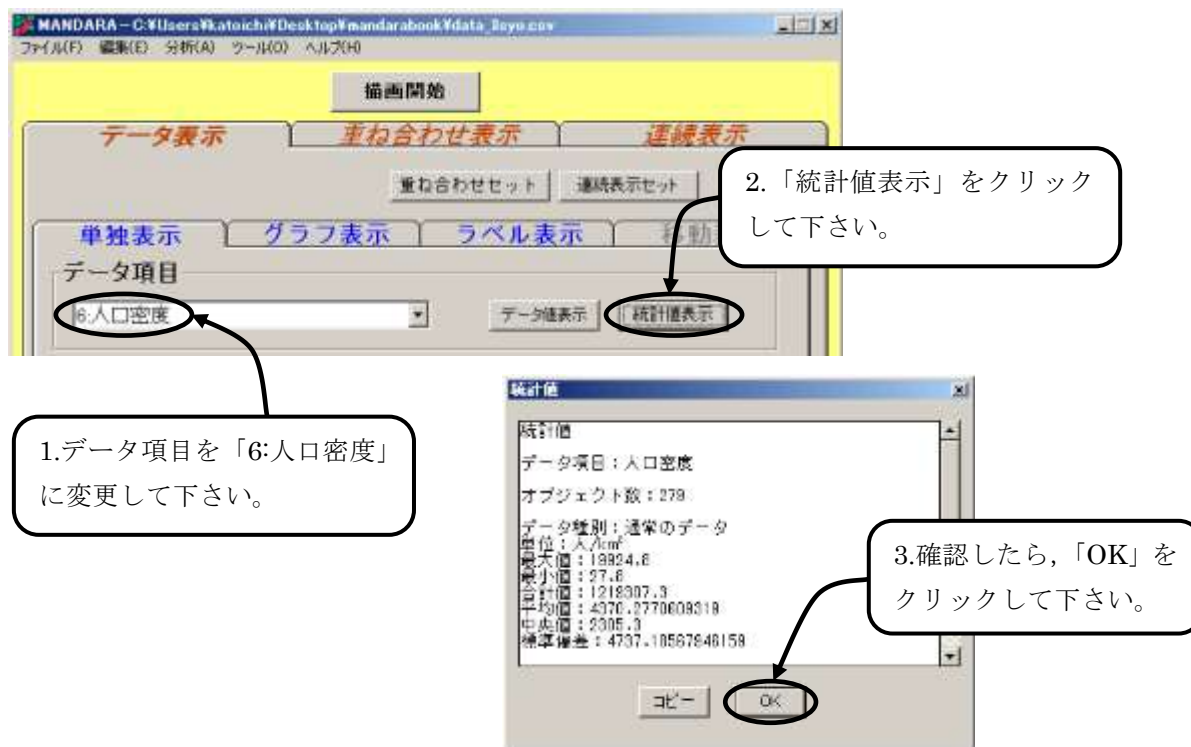


図 3.3.2.1 統計値の確認

度数分布表を作成するために必要な情報は、最大値と最小値の2つです。また、度数分布表を評価するために必要な情報は、平均値とオブジェクト数の2つです。

表 3.3.2.1 度数分布表を作成・評価するためのパラメータ

最大値	19,924.8 人/km ²	
最小値	27.8 人/km ²	
平均値	4,370.3 人/km ²	小数第2位を四捨五入
オブジェクト数	279	市区町村数

さて、度数分布表を作成しましょう。最大値、最小値に注意して階級区分すると、階級分割数を20、階級区分値は1000人/km²刻みの等間隔で作成するのがちょうどよいでしょう。

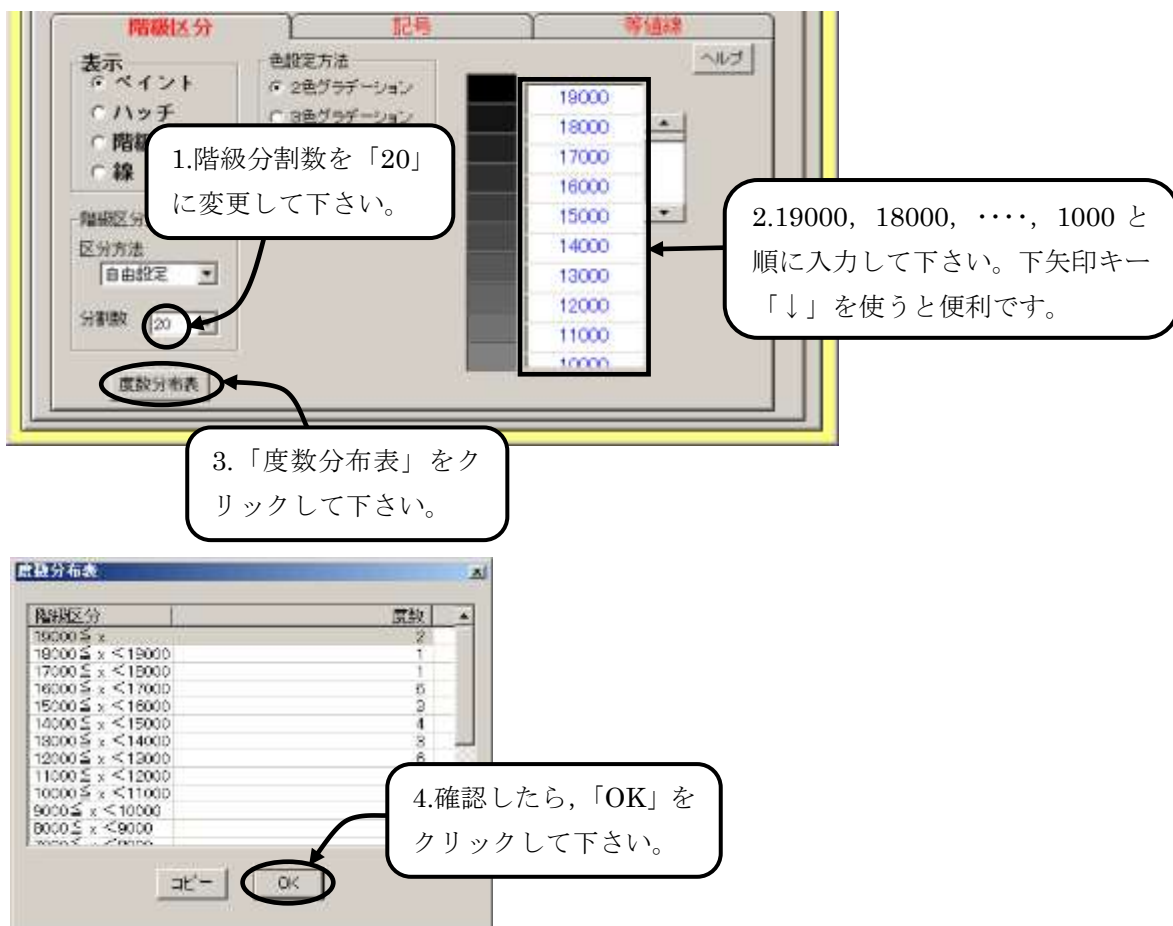


図 3.3.2.2 度数分布表の作成

次ページの図 3.3.2.3 は、度数分布表をヒストグラムにしたものです。このヒストグラムをみると、度数が最大なのは、1000人/km²未満であることがわかります。平均が約4,370人/km²ですから、平均よりもだいぶ少ない階級に集中していることがわかります。また、人口密度が高くなるにつれて、徐々に度数が少なくなっていく傾向も読みとれます。

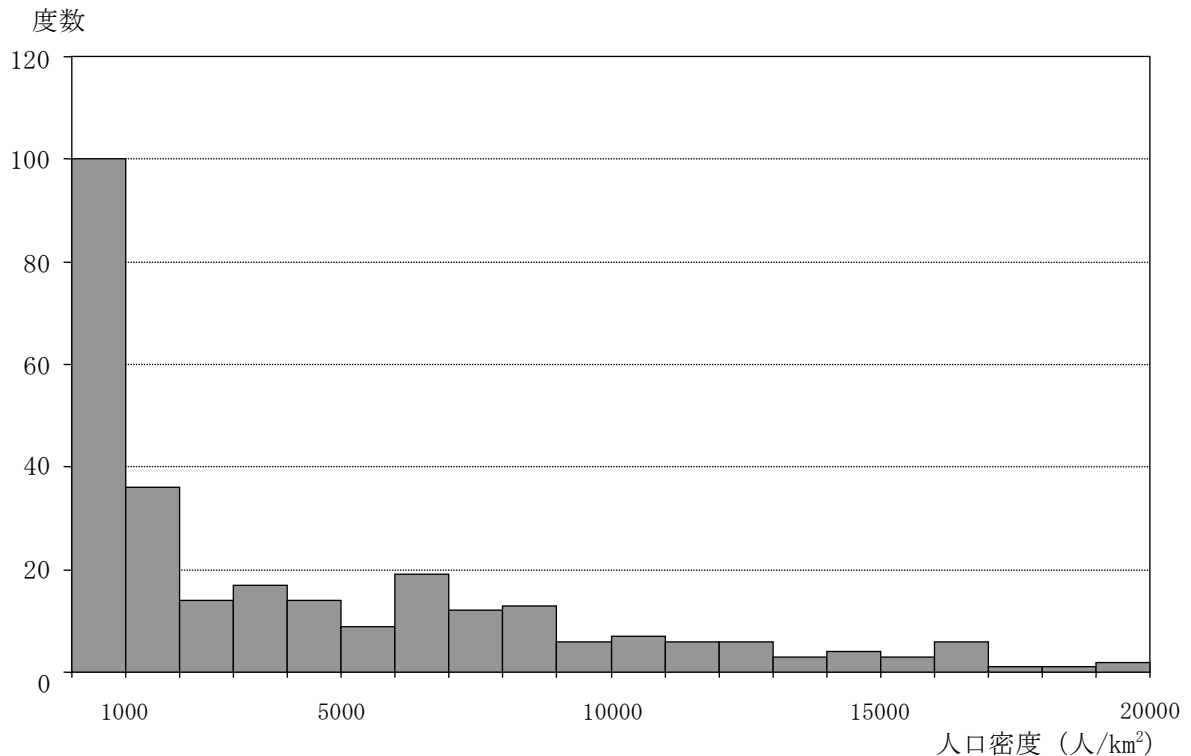


図 3.3.2.3 ヒストグラム

さて、階級区分値を決定する方法は、いくつか方法が知られています。ここでは、(1)等間隔の区分、(2)平均値を用いる区分、(3)等比数列を用いる区分、(4)分位数（データを等分する区分）の4つの方法を紹介します。さらに詳しく知りたい方は、章末の引用・参考文献を参照して下さい。

(1) 等間隔の区分

統計データの最大値から最小値の差をとり、これを階級分割数で割ると、等間隔の区分になります。場合によっては、区切りのよい数に丸めることもあります。ここでは、階級分割数を6として、実際に計算してみましょう。

$$\text{最大値} - \text{最小値} = 19,924.8 - 27.8 = 19,897.0$$

$$\text{間隔} = 19897.0 \div 6 \approx 3,316.167$$

階級区分は、小さい方から $27.8(\text{最小値}) + 3,316.167 = \underline{3,343.967}$, $3,343.967 + 3,316.167 = \underline{6,660.133}$, $6,660.133 + 3,316.167 = \underline{9,976.3}$, $9,976.3 + 3,316.167 = \underline{13,292.47}$, $13,292.47 + 3,316.167 = \underline{16,608.63}$ となります。しかし、このままでは区切りが悪いので、階級区分の下限を $3,000 \text{ 人/km}^2$ 、間隔を $3,000 \text{ 人/km}^2$ にして、階級区分図を作成してみましょう。

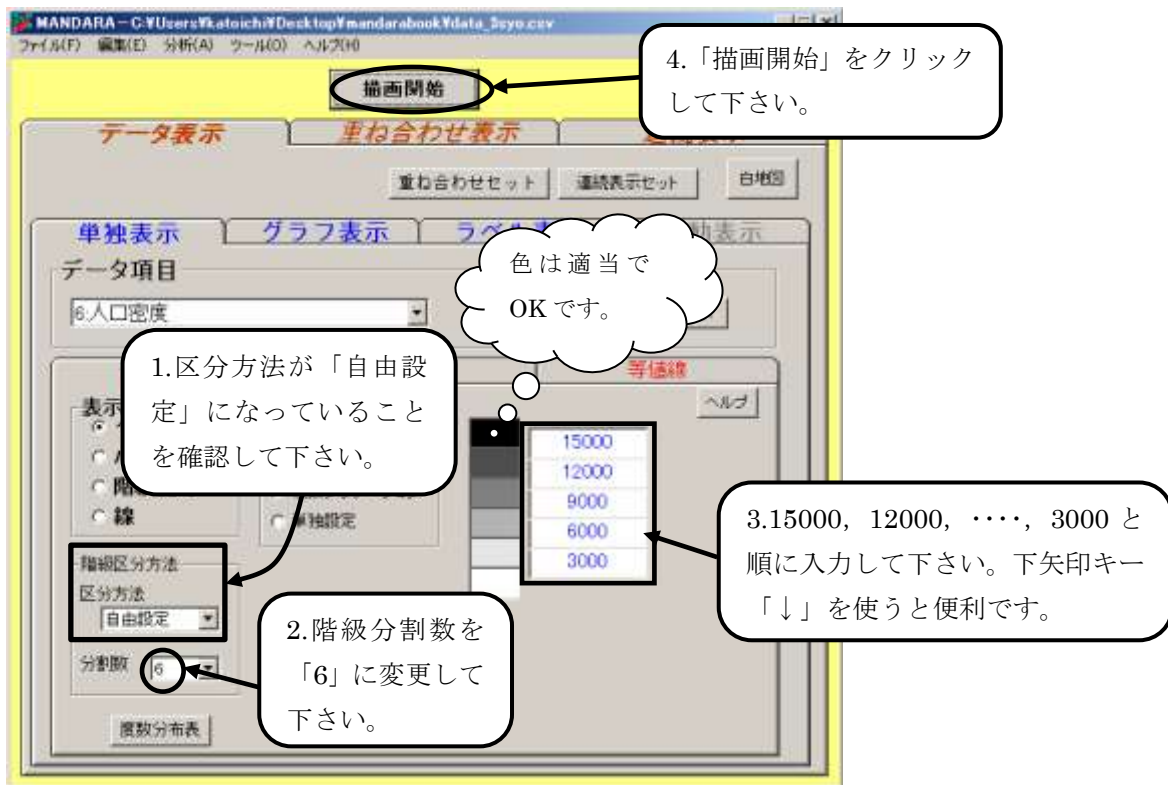


図 3.3.2.4 等間隔の区分

前ページのヒストグラムで確認したとおり、3000 人/km² 未満の階級に多くの市区町村が含まれてしまうという問題があります。

(2) 平均値を用いる区分

統計データ全体の平均値を求め、これを境にデータを 2 つのグループに分けます。次に、それぞれのグループの平均値を求め、これを境にデータを 2 つのグループに分けます。こうして、4 つのグループができます。これを繰り返していけば、8, 16, …, 2ⁿ 個の階級分割ができます。ここでは、階級分割数を 4 として階級区分値を計算してみましょう。それでは、「data_3syo.xls」を開いて下さい。必要なデータは、このファイルの「Sheet2」のワークシートにあります。

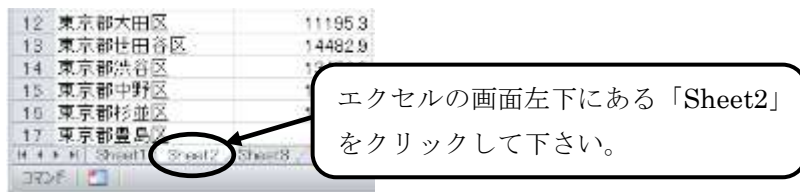


図 3.3.2.5 「Sheet2」のワークシートを開く

手順は次の通りです。①統計データを降順（大きい順に）に並べ替える。②全体の平均値を求める。③平均値を境に、2 つのグループに分ける。④それぞれのグループの平均値を求める。
それでは、実際に計算してみましょう。

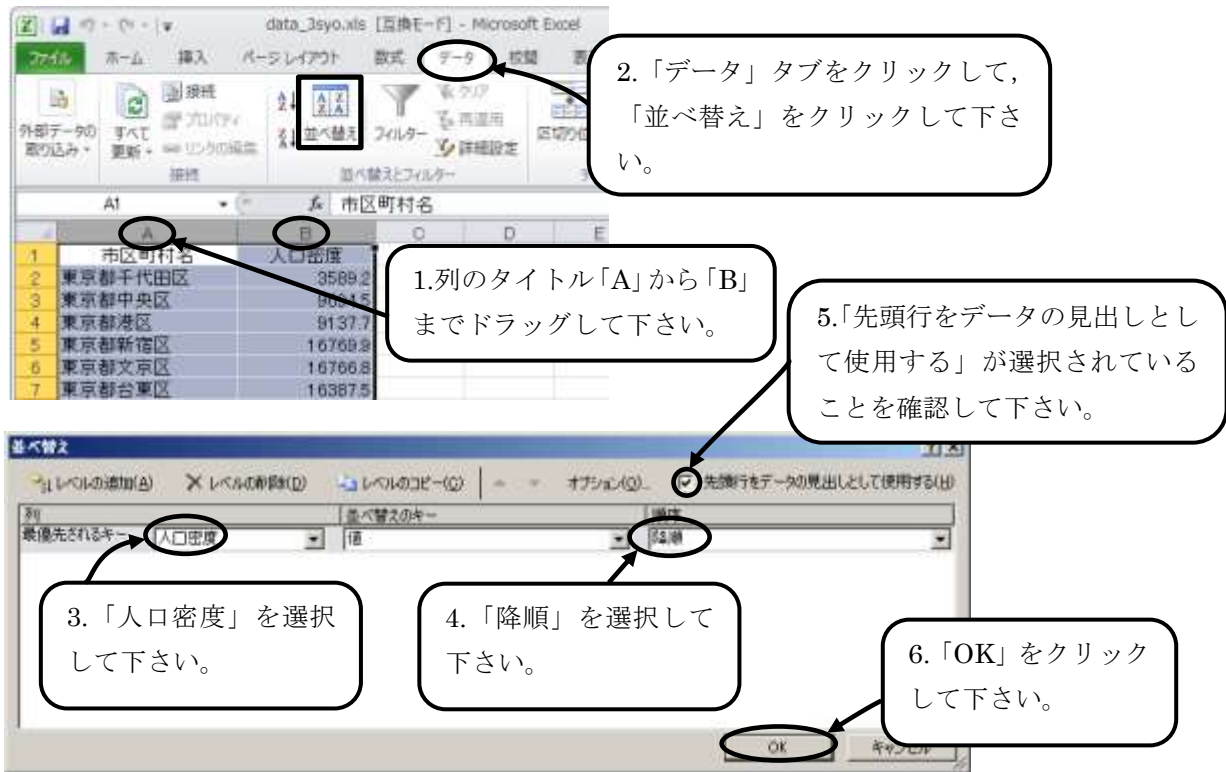


図 3.3.2.6 降順で並べ替え

次に、全体の平均値を計算しましょう。

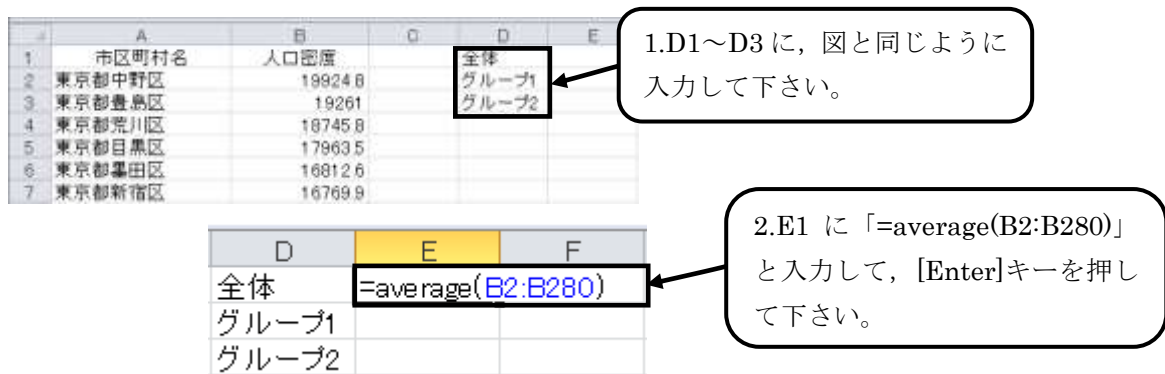


図 3.3.2.7 全体の平均値を求める

ここで、エクセルの関数「average」について紹介します。関数「average」は、平均を求める関数です。文法は次の通りです。なお、アルファベットと数字は大文字でも小文字でも構いませんが、必ず半角で入力して下さい。

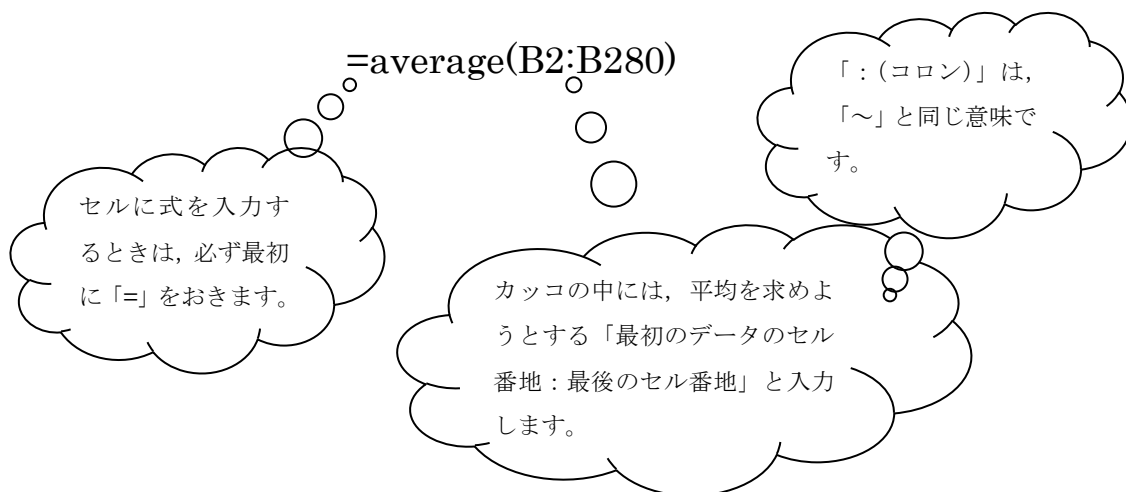


図 3.3.2.8 平均を求める関数「average」

全体の平均値が約 4,370 人/km²と求まりましたので、平均値より大きいグループを「グループ 1」、平均値より小さいグループを「グループ 2」として、各グループの平均値を E2, E3 にそれぞれ求めてみましょう。グループ 1 の平均値は約 9,598 人/km²、グループ 2 の平均値は約 1,264 人/km²となります。したがって階級区分値は、9,598 人/km²、4,370 人/km²、1,264 人/km²となります。

それでは、図 3.3.2.4 (ページ) を参考に、階級区分図を作成しましょう。都心部から郊外に向かうにつれて人口密度が低くなっていく様子が読みとりやすくなりました。

(3) 等比数列を用いる区分

統計データを階級区分することは、個々の統計データをグループ分けすることです。ですから、機械的に階級区分するのではなく、統計データの分布にふさわしい階級区分を考えることも大切です。そこで、図 3.3.2.3 (ページ) のヒストグラムをみてみましょう。統計データは、1,000 人/km²未満の階級に全体の約 3 割の市町村が含まれています。続く 1,000 人/km²以上 2,000 人/km²未満の階級になると急激に減少し、そして、これ以上の階級に含まれる市区町村の数は徐々に少なくなっていく傾向が読みとれます。

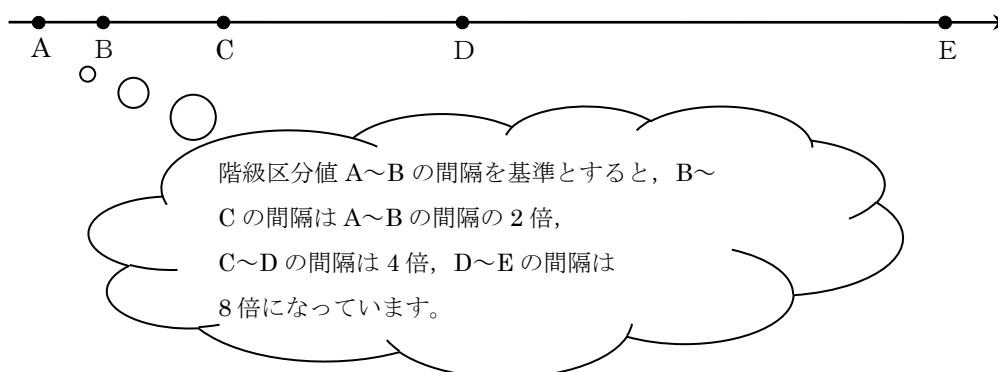


図 3.3.2.9 等比数列を用いる区分

たとえば、上の図 3.3.2.9 のように、階級区分値 A~B の間隔を基準としたとき、B~C の間隔を A~B の間隔の 2 倍、C~D の間隔を 4 倍、D~E の間隔を 8 倍となるように階級区分値を定めると、

等比数列を用いた階級区分になります。統計データの値が小さいうちは間隔が小さく、統計データの値が大きくなるにつれて間隔が広がります。したがって、統計データがどちらか一方に大きく偏った分布をしている場合には、適した方法といえるでしょう。

しかし、等比数列ですから、適切な初項と公比を決めなければなりません。目安としては、各階級区分に含まれる統計データの数がほぼ同数になることを目標にするとよいでしょう。そのためには、何度も何度も試行錯誤が必要ですが、MANDARA を使えば、大丈夫です！ここでは、階級分割数を6として考えることにします。

① 初項 200, 公比 3 のとき → $a_n = 200 \times 3^{n-1}$

n に 1~5 を代入すると、階級区分値が得られます。階級区分値は、200, 600, 1,800, 5,400, 16,200 になります。それでは、図 3.3.2.4 (ページ) を参考に、階級区分を設定して描画しましょう。

次に、度数分布表をみてみましょう。

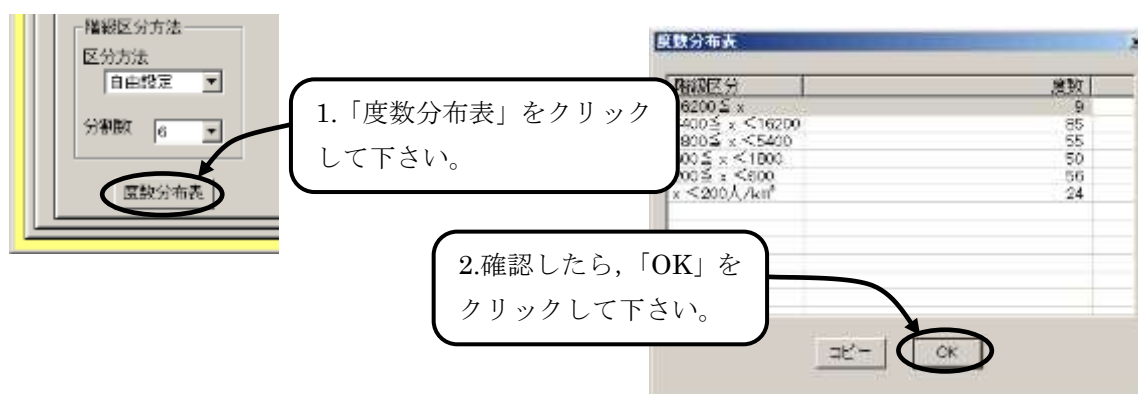


図 3.3.2.10 度数分布表を確認する

16,200 人/km² 以上の階級にわずか 9 つの区しか含まれないこと、さらに、5,400 人/km² 以上 16,200 人/km² 未満の階級に全市区町村の 3 割が含まれることになり、バランスを欠く階級区分図であるといえるでしょう。

② 初項 300, 公比 2.5 のとき → $a_n = 300 \times 2.5^{n-1}$

n に 1~5 を代入すると、階級区分値は、300, 750, 1,875, 4,688, 11,719 になります。それでは、階級区分図を作成し、度数分布表も確認しましょう。

最上位の階級である 11,719 人/km² 以上に 29 市区が含まれるようになり、先ほどの①よりもバランスのよい階級区分図であるといえるでしょう。

③ 初項 500, 公比 2 のとき → $a_n = 500 \times 2^{n-1}$

階級区分値は、500, 1,000, 2,000, 4,000, 8,000 になります。度数分布表をみると、8,000 人/km² 以上と 500 人/km² 未満の階級の度数（市区町村数）がほかの階級に比べて多くなっていますが、比較的均等に近づきました。①と②に比べるとバランスのよい階級区分図になっています。

(4) データ数を等分する区分 (分位数)

各階級に含まれる統計データの数が等しくなるように階級区分します。ただし、統計データはあらかじめ降順 (大きい順) に並べ替えておく必要があります。それでは、「data_3syo.xls」を開いて下さい。必要なデータは、このファイルの「Sheet3」のワークシートにあります。Sheet3の開き方は、図 3.3.2.5 (ページ) を参照して下さい。なお、Sheet3 の統計データはあらかじめ降順に並べ替えられています。

1. C2 と C3 に、それぞれ「1」「2」と入力して下さい。

2. C2 から C3 までドラッグして下さい。

3. C3 セルの右下にマウスの矢印をおいて下さい。すると、マウスの矢印が十字の形に変化します。

4. 下までドラッグして下さい。3~279の数字が自動的に入りました。

	A	B	C	D
1	市区町村名	人口密度		
2	東京都中野区	19924.8	1	
3	東京都豊島区	19261	2	
4	東京都荒川区	18745.8		
5	東京都目黒区	17963.5		
6	東京都墨田区	16812.6		
7	東京都新宿区	16769.9		

	A	B	C	D
1	市区町村名	人口密度		
2	東京都中野区	19924.8	1	
3	東京都豊島区	19261	2	
4	東京都荒川区	18745.8		
5	東京都目黒区	17963.5		
6	東京都墨田区	16812.6		
7	東京都新宿区	16769.9		

	A	B	C	D
1	市区町村名	人口密度		
2	東京都中野区	19924.8	1	
3	東京都豊島区	19261	2	
4	東京都荒川区	18745.8		
5	東京都目黒区	17963.5		
6	東京都墨田区	16812.6		
7	東京都新宿区	16769.9		

	A	B	C	D
267	神奈川県箱根町	153	266	
268	千葉県長南町	150.3	267	
269	千葉県富山町	142.6	268	
270	千葉県三芳村	137.3	269	
271	千葉県丸山町	123.3	270	
272	埼玉県秩父市	122.1	271	
273	埼玉県東秩父村	102.1	272	
274	千葉県大多喜町	89.7	273	
275	埼玉県小畷野町	84.5	274	
276	神奈川県山北町	56.3	275	
277	埼玉県神泉村	51.3	276	
278	神奈川県清川村	49.2	277	
279	東京都奥多摩町	29.9	278	
280	東京都檜原村	27.8	279	
281				
282				

図 3.3.2.11 連続データの作成 (フィル)

統計データの個数は 279 です。階級分割数を 6 とすれば、1 階級あたりの市区町村数は 46.5 になります。1 階級あたりの市区町村数は整数でなければならないので、ここでは、1 階級あたりの市区町村数を 47 とし、端数は最下位の階級で調整することにします。それでは、今、付けた番号が 47, 94, 141, 188, 235 番の人口密度が階級区分値になりますので、確認してみましょう。階級区分値は、8976.3, 5661.5, 2216.8, 782.8, 330.3 になります。

それでは、図 3.3.2.4 (ページ) を参考に、階級区分図を作成しましょう。(3)等比数列を用いる区分の③初項 500, 公比 2 のときと比較すると、連続性(同じ階級区分の区域が連続する)に若干欠けますが、まずまずの出来になっています。

いずれにしても、階級区分の方法は、どれが正しく、どれが間違いということはありません。階級区分図から「何を伝えたいのか」という作成目的に照らし合わせて、手順を踏んで階級区分すればよいのです。

MANDARA があらかじめ用意する階級区分方法

3.3.2 項では、「階級区分方法」を「自由設定」にして等間隔の区分、平均値を用いる区分、等比数列を用いる区分、データ数を等分する区分(分位数)の 4 つの方法について説明してきました。MANDARA では、あらかじめ「分位数」、「面積分位数」、「標準偏差」、「等間隔」による階級区分方法を用意しています。詳しくは、谷謙二著『フリー GIS ソフト MANDARA パーフェクトマスター』を参照して下さい。

a 面積分位数

各階級に含まれる地域の面積が等しくなるようにデータを区分する方法です。たとえば、階級分割数が 4 であれば、総面積の 25%、50%、75%にあたる値が階級区分値になります。ただし、面形状オブジェクトの場合のみ使用できます。

b 標準偏差を用いる区分

データが正規分布に近い場合には、標準偏差を用いる区分が使われることがあります。まず、平均値と標準偏差を求めます。次に、平均値に標準偏差を加えたり引いたりして階級区分値を決めます。平均値を \bar{x} 、標準偏差を σ とすると、MANDARA では、 $\bar{x} + \sigma$ 、 $\bar{x} + 1/2\sigma$ 、 \bar{x} 、 $\bar{x} - 1/2\sigma$ 、 $\bar{x} - \sigma$ と階級区分値を決定します。分割数は 6 で固定です。たとえば、平均値が 50 万円、標準偏差が 6 万円のデータがあれば、階級区分値を 56 万円、53 万円、50 万円、47 万円、44 万円とすれば、標準偏差を用いる区分になります。

引用文献

- 1) 菅野峰明・安仁屋政武・高阪宏行(1987):『地理学講座 第 2 巻 地理的情報の分析手法』, 古今書院, p.61.

参考文献

- 安仁屋政武(1987):『主題図作成の基礎』, 地人書房.
谷謙二(2011):『フリー GIS ソフト MANDARA パーフェクトマスター』, 古今書院.