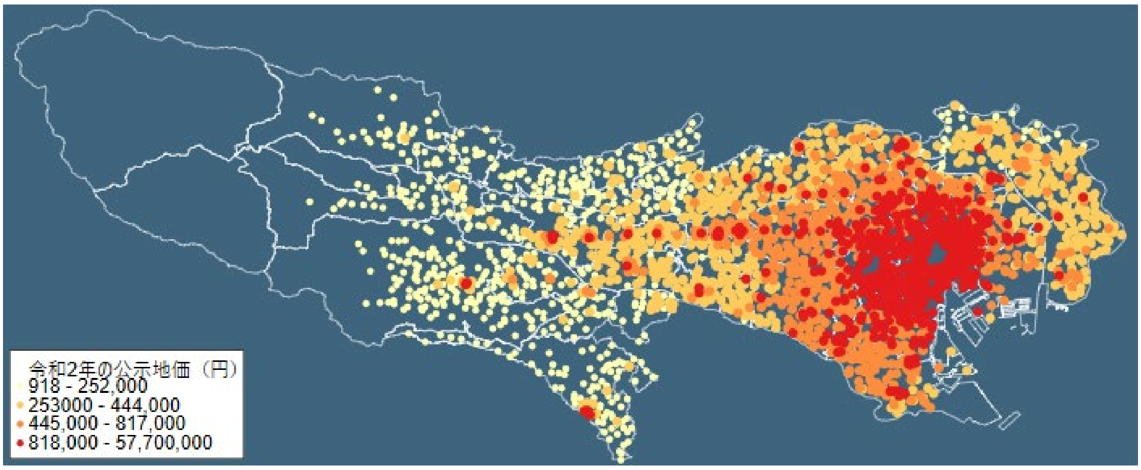
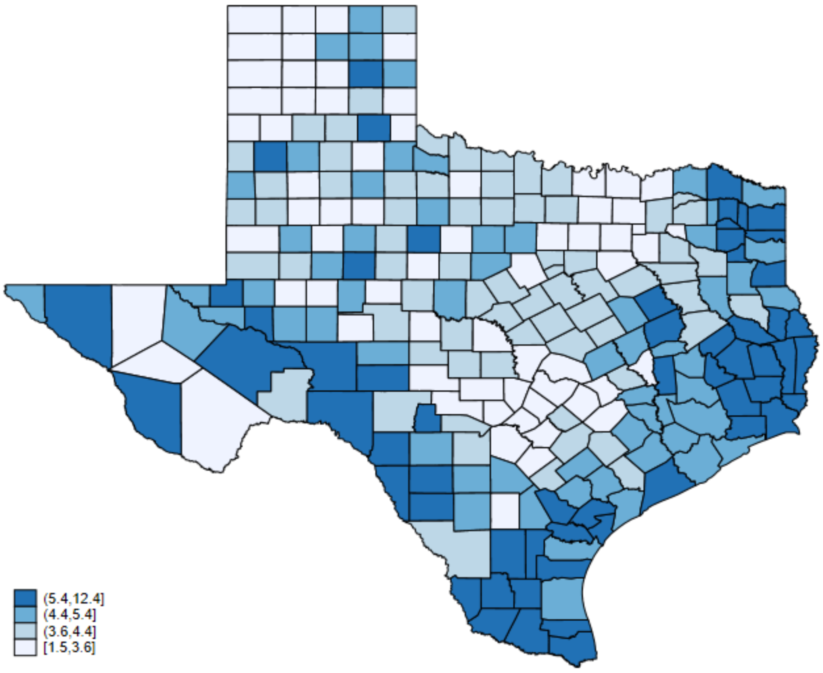
空間自己回帰(SAR)モデル

第一部　地理情報システム(GIS)と Stata - 東京都の地価データを地図上に表示



第二部　空間自己回帰モデル - 失業率と大学卒業率の関係をモデル化



LightStone® 株式会社ライトストーン

第一部　地理情報システム(GIS)と Stata - 東京都の地価データを地図上に表示

はじめに

- 地理情報システム(GIS : Geographic Information System)とは、「地理的位置を手がかりに、位置に関する情報を持ったデータ（空間データ）を総合的に管理・加工し、視覚的に表示し、高度な分析や迅速な判断を可能にする技術」です。

  出典）国土地理院　https://www.gsi.go.jp/GIS/whatisgis.html

- Stata 15 以降のバージョンは地理情報システム(GIS)を正式にサポートしています。地図データのインポートと加工、グラフ化、空間自己回帰(SAR : spatial autoregressive) モデルのためのコマンドが用意されています。
- 地図情報としては一般的に「シェープファイル」がよく利用されます。シェープファイルは図形情報と属性情報を持った地図データファイル群で、政府機関などのウェブサイトで配布されています。
- この例題では、東京都の地価データを地図上に表示する方法を紹介します。メニュー操作を基本としていますが、一部コマンド操作のものもあります。do ファイル part1.do に全ての操作のコマンドをまとめています。

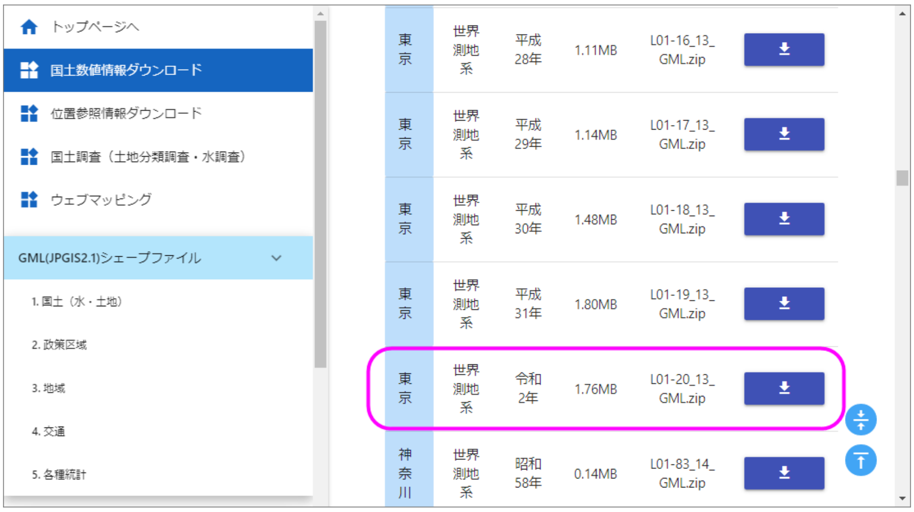
1. 地価データの準備
2. 地図データの準備
3. グラフ作成

LightStone® 株式会社ライトストーン

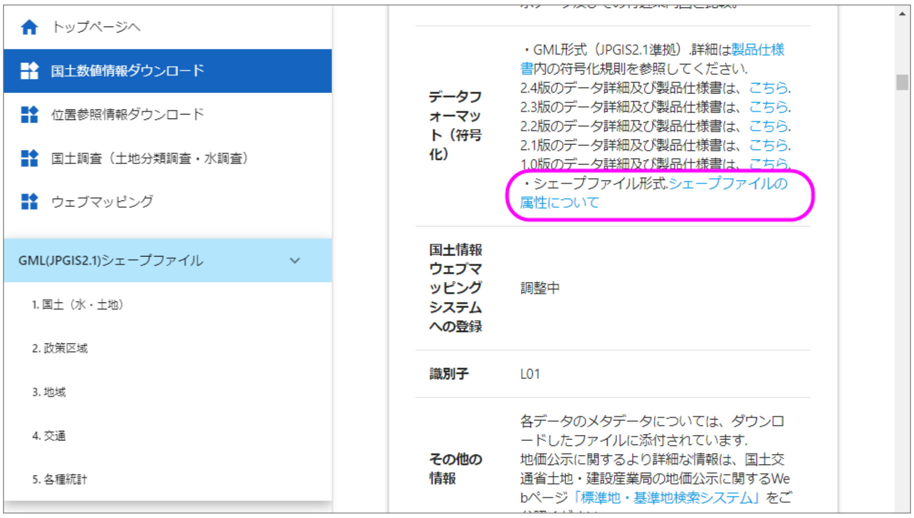
1. 地価データの準備

Step 1: 東京都の地価データのシェープファイルをダウンロードする

- こちらのページから令和 2 年の東京都の地価公示データをダウンロードします。これは各年 1 月 1 日時点の全国の標準地について、位置（点）、公示価格、利用現況、用途地域、地積等を GIS データとして整備したものです。
https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-L01-v2_5.html

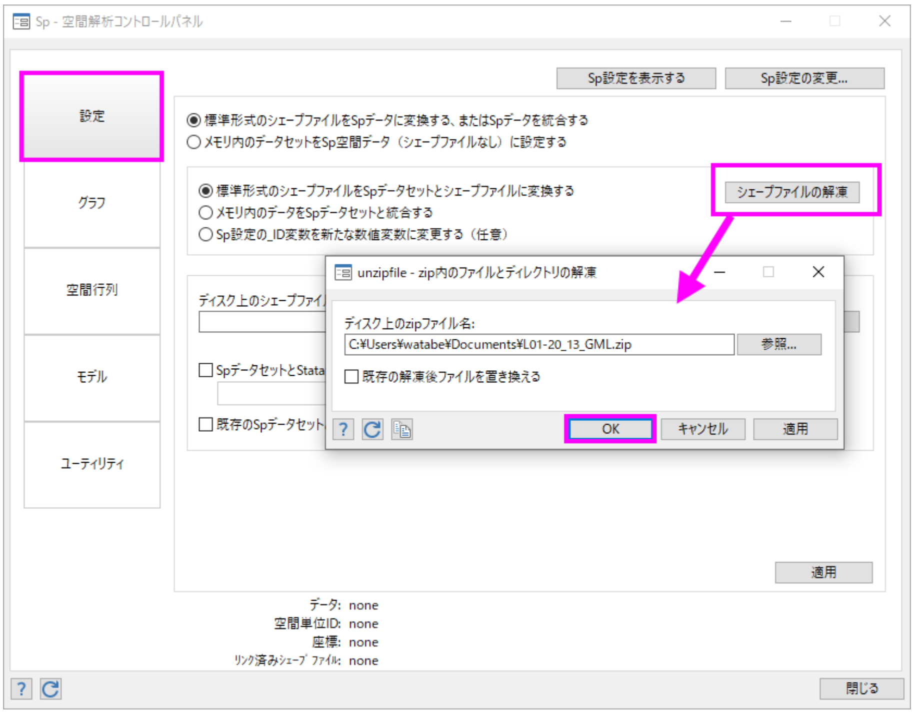


ダウンロードしたファイル L01-20_13_GML.zip を現在の作業フォルダにコピーします。また、「データフォーマット（符号化）」欄にある「**シェープファイルの属性について**」の Excel ファイルをダウンロードします。これはシェープファイルの属性名の対応表です。



Step 2: シェープファイルを Stata 形式に変換する

● ダウンロードしたシェープファイルを解凍（展開）します。Stata を起動し、メニューから「**統計 > 空間自己回帰モデル**」を選択します。「空間解析コントロールパネル」で次のように設定します。

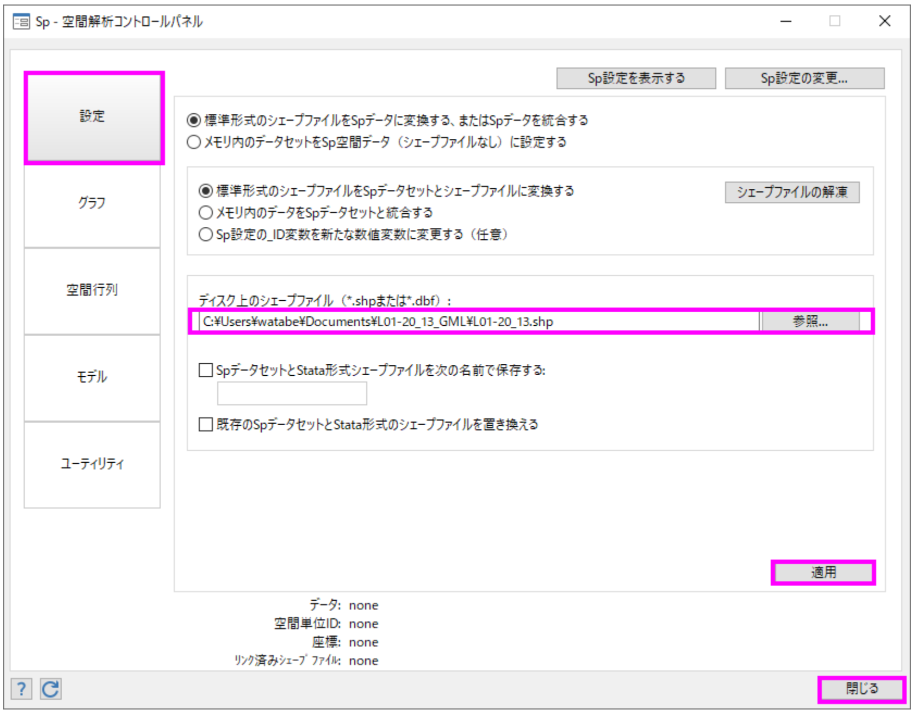


結果画面に解凍結果が表示されます。

```
  inflating: L01-20_13_GML/KS-META-L01-20_13.xml
  inflating: L01-20_13_GML/L01-20_13.dbf
  inflating: L01-20_13_GML/L01-20_13.geojson
  inflating: L01-20_13_GML/L01-20_13.prj
  inflating: L01-20_13_GML/L01-20_13.shp
  inflating: L01-20_13_GML/L01-20_13.shx
  inflating: L01-20_13_GML/L01-20_13.xml

successfully unzipped C:\Users\watabe\Documents\L01-20_13_GML.zip to current directory
total processed:  7
        skipped:  0
      extracted:  7
```


LightStone® 株式会社ライトストーン

- シェープファイルを Stata のデータ形式に変換します。次のように設定して、「空間解析コントロールパネル」を閉じます。




```
(importing .shp file)
(importing .dbf file)
(creating _ID spatial-unit id)
(creating _CX coordinate)
(creating _CY coordinate)

file L01-20_13_shp.dta created
file L01-20_13.dta     created
```

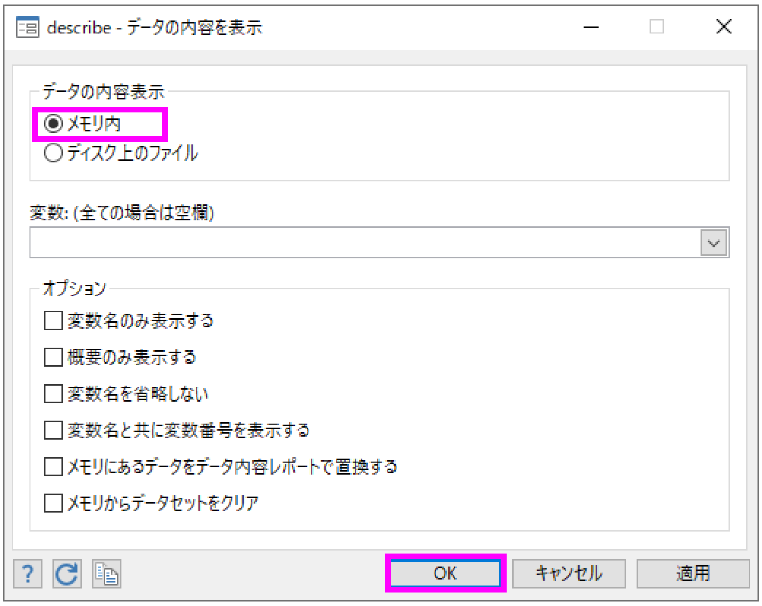
- 作成された Stata 形式のシェープファイル L01-20_13.dta には日本語が含まれているため、適切なエンコードを設定して変換します。（コマンド操作のみ）

```
unicode encoding set "Shift_JIS"


unicode translate "L01-20_13.dta"
```

```
(using Shift_JIS encoding)
(Directory ./bak.stunicode created; please do not delete)

File summary (before starting):
      1  file(s) specified
      1  file(s) to be examined ...

File L01-20_13.dta (Stata dataset)
    all variable names okay, ASCII
    all data labels okay, ASCII
    all variable labels okay, ASCII
    all characteristic names okay, ASCII
    all characteristic contents okay, ASCII
    111 str# variables okay, ASCII
      0 str# variables okay, already UTF-8
     19 str# variables translated
        ------------------------------------------------------------------------------------------
        File successfully translated

File summary:
    all files successfully translated
```

- メニューから「**ファイル > 開く**」を選択して、L01-20_13.dta を開きます。

- データの内容を確認します。メニューから「**データ > データの内容表示 > メモリ/ファイル内のデータの内容表示**」を選択して、次のように設定します。

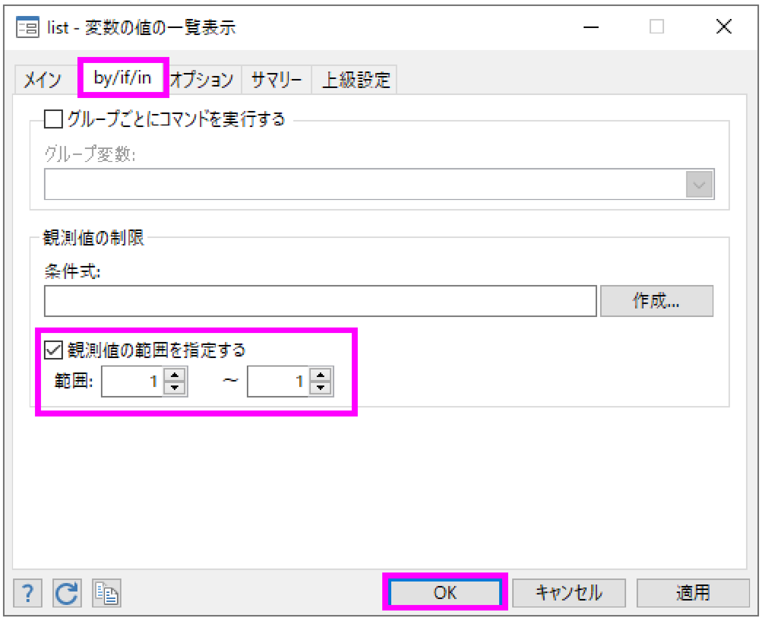


次のように結果が表示され、非常に多くの情報（シェープファイルの属性）が含まれていることがわかります。先頭の_ID は空間 ID、_CX は X 座標、_CY は Y 座標です。storage type 欄より、これらは数値変数としてインポートされていることがわかります。その他の変数は

文字列変数(str)としてインポートされています。

```
Contains data from C:\Users\watabe\Documents\L01-20_13.dta
  obs:         2,602
 vars:           133                          10 Mar 2021 09:46

              storage   display    value
variable name   type    format     label      variable label

_ID             int     %12.0g                Spatial-unit ID
_CX             double  %10.0g                x-coordinate of area centroid
_CY             double  %10.0g                y-coordinate of area centroid
L01_001         str3    %9s                   L01_001
L01_002         str3    %9s                   L01_002
L01_003         str3    %9s                   L01_003
L01_004         str3    %9s                   L01_004
L01_005         str4    %9s                   L01_005
L01_006         str8    %9s                   L01_006
L01_007         str1    %9s                   L01_007
L01_008         str5    %9s                   L01_008
                str5
```

- データの 1 行目の内容を確認します。メニューから「データ ＞ データの内容表示 ＞ データの一覧表示」を選択します。「by/if/in」タブで次のように設定します。



次のように結果が表示されます。

LightStone® 株式会社ライトストーン

1.	_ID	_CX	_CY	L01_001	L01_002	L01_003	L01_004	L01_005	L01_006
	1	139.74481	35.69014	000	001	000	001	2020	3160000

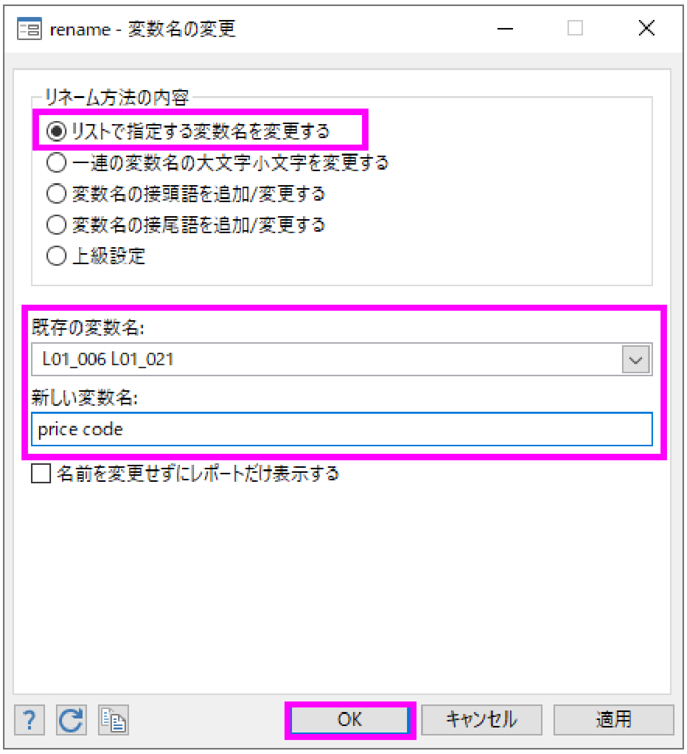
L01_007	L01_008	L01_009	L01_010	L01_011	L01_012	L01_013	L01_014
1	false	false	false	false	false	false	false

L01_015	L01_016	L01_017	L01_018	L01_019	L01_020	L01_021	L01_022
false	false	false	false	false	false	13101	千代田

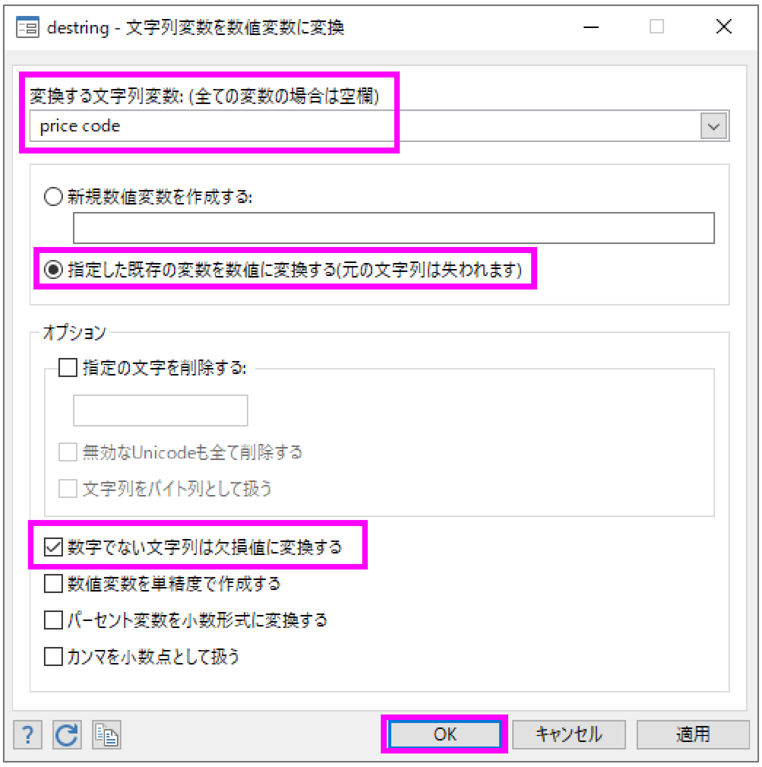
L01_023	L01_024	L01_025	L01_026	L01_027	L01_028
東京都　千代田区三番町６番２５	969	住宅，その他	共同住宅	SRC10	true

Step 1 でダウンロードした「**シェープファイルの属性について**」の Excel ファイルを使用して、対応番号と属性名を確認します。今回は **L01_006** の公示価格（円/m²）、**L01_021** の標準地行政コードを使用します。
※2021 年 3 月 10 日現在、対応表の内容に誤りがある様子のため、データの内容から判断して対応付けを行いました。

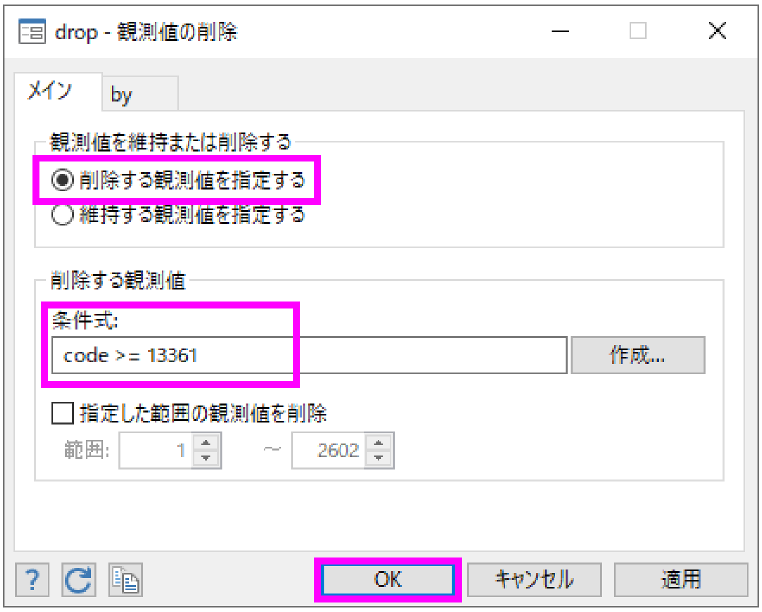
- 使用する変数（属性）に分かりやすい名前を付けます。メニューから「**データ ＞ データユーティリティ ＞変数のグループ名を変更**」を選択して、次のように設定します。



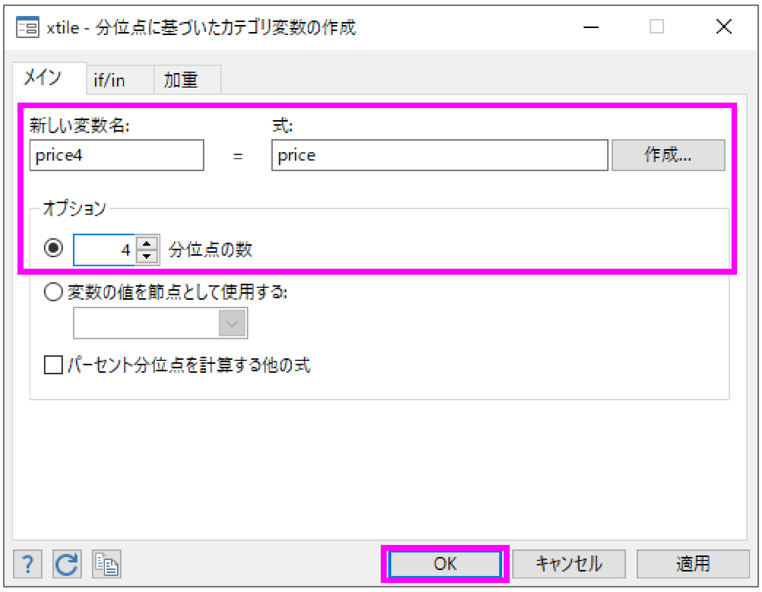
- code と price を文字列変数から数値変数に変換します。メニューから「データ ＞ データの作成または変更 ＞ その他の変数変換コマンド ＞ 文字列変数から数値変数に変換」を選択し、次のように設定します。



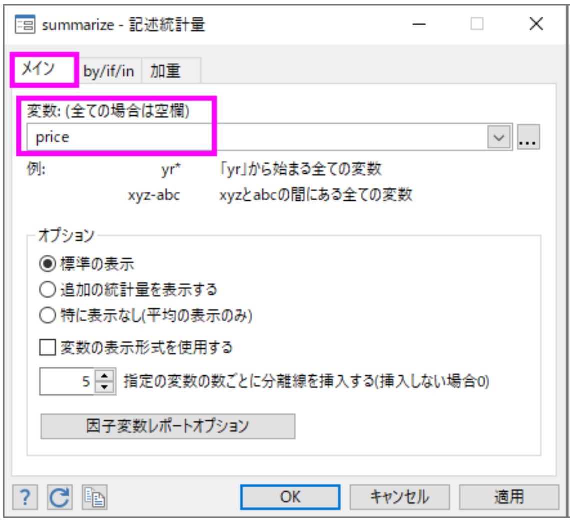
- 今回は島嶼部のデータを除きます。島嶼部の標準地行政コードは 13361 以降のため、code が 13361 以上のものを削除します。メニューから「データ ＞ データの作成または変更 ＞ 観測値の削除」を選択し、次のように設定します。

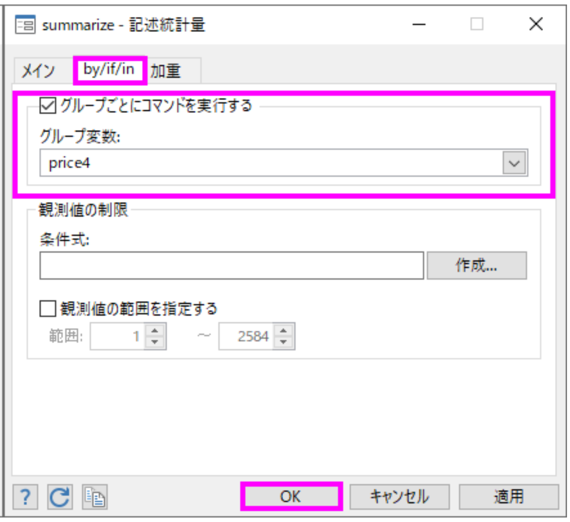


- 地価データを地図上に表示する際、価格の高低によって色分けして表示すると分かりやすくなります。カラーマップ用に、`price` の四分位数を表す新規変数 `price4` を作成します。メニューから「統計 > 要約と記述統計量 > 分位点情報の変数を作成」を選択し、次のように設定します。



- 作成した `price4` の内容を確認します。メニューから「データ > データの内容表示 > 記述統計量」を選択し、「メイン」タブと「by/if/in」タブで次のように設定します。

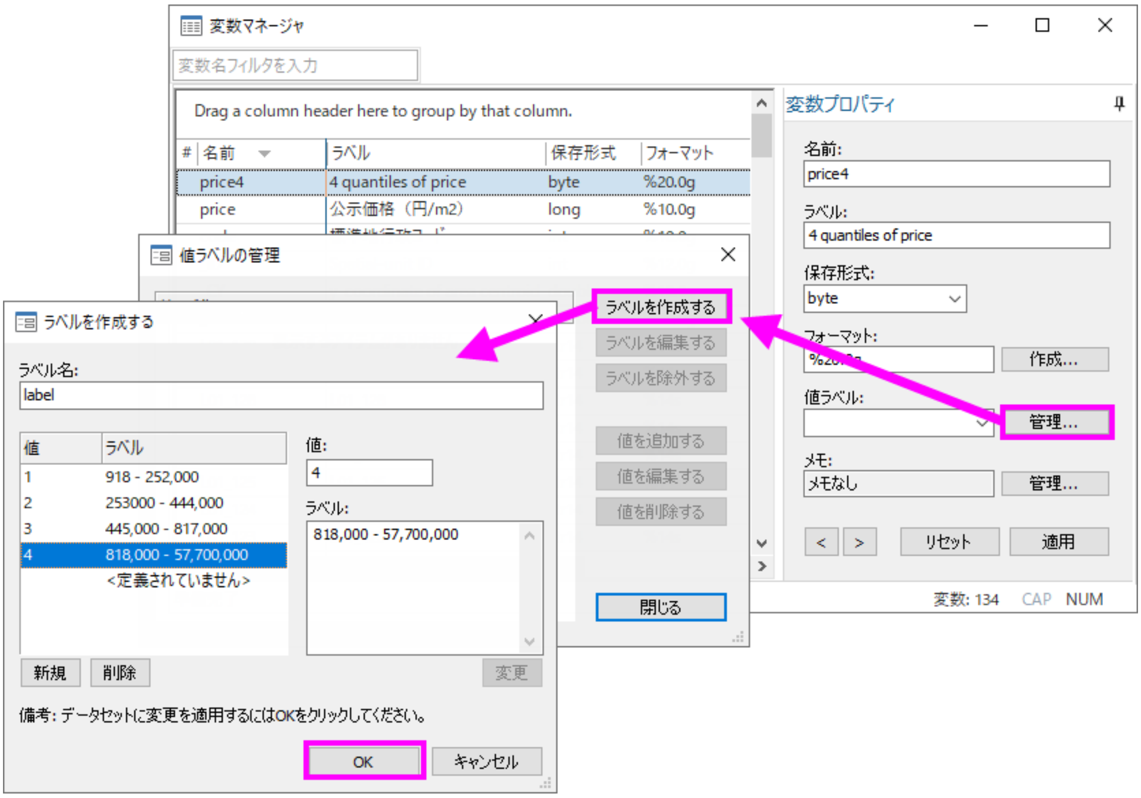




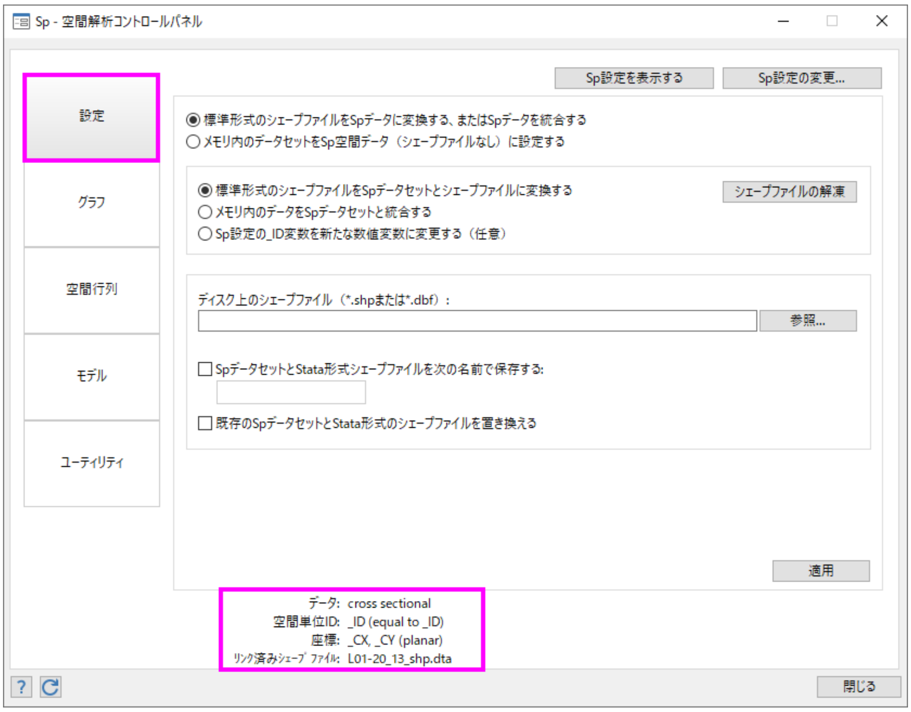
結果は次のようになります。`price4 = 1`～`4` の最小値と最高値がわかります。

```
-> price4 = 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       price |        648    157297.7    57053.52        918     252000

-> price4 = 2

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       price |        648    343248.5    55242.96     253000     444000

-> price4 = 3

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       price |        642    603686.9      105790     445000     817000

-> price4 = 4

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       price |        646     3573573     6183768     818000   5.77e+07
```

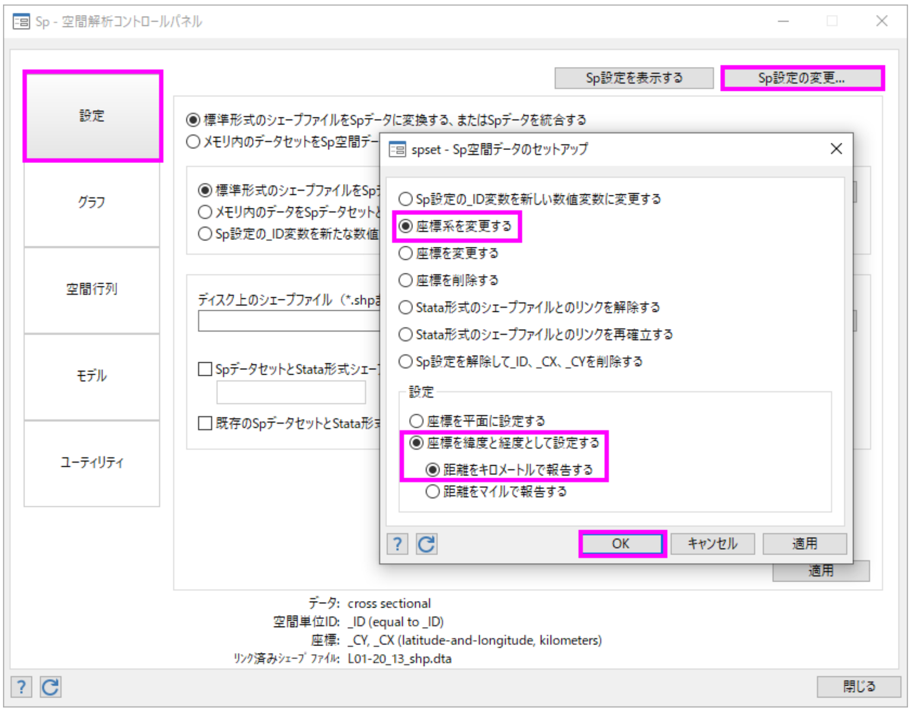
- price4 の値ラベルを作成します。ニューから「データ > 変数マネージャ」を選択し、次のように設定します。この値ラベルは後程カラーマップの凡例として使用されます。



- 空間データを扱うにあたって、どの変数を空間の ID とするか設定する必要があります。Stata 形式のシェープファイルを作成した際に自動で設定されますが、後から変更することも可能です。現在の設定を確認するには、メニューから「**統計 > 空間自己回帰モデル**」を選択して「空間解析コントロールパネル」を開きます。



- デフォルトでは、座標は上記の通り平面(planar)単位に設定されます。元のシェープファイルは緯度経度で記録されているため、座標単位の設定を変更します。

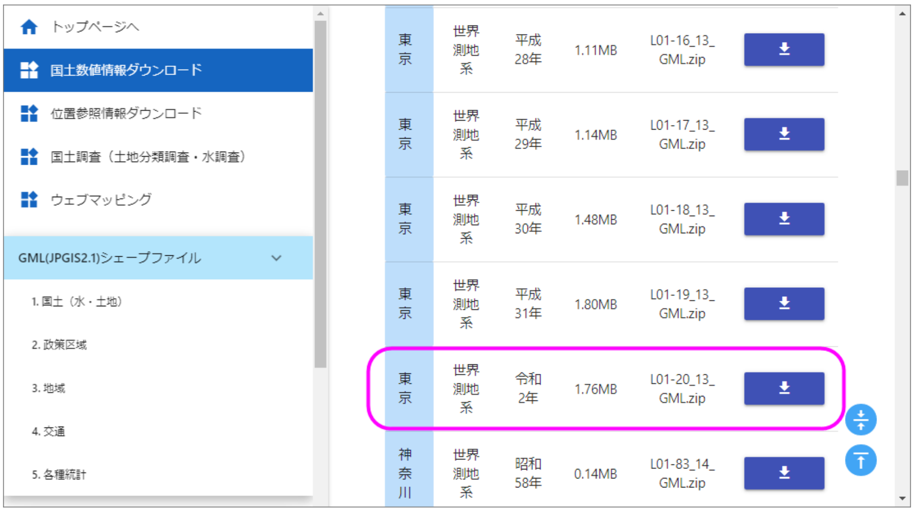


● メニューから「**ファイル ＞ 保存**」を選択してデータを保存します。以上で地価データの準備は完了です。

2.　地図データの準備

Step 1: 東京都の行政区域データのシェープファイルをダウンロードする

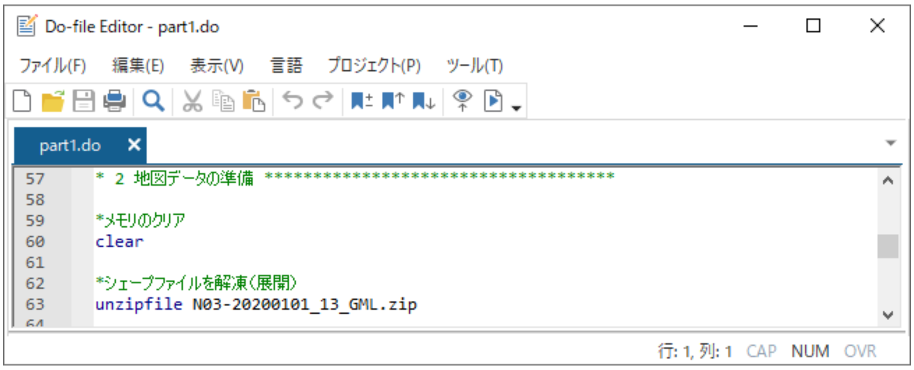
● こちらのページから令和 2 年の東京都の行政区域データをダウンロードします。これは都道府県名、支庁・振興局名、郡・政令都市名、市区町村名、行政区域コード等を GIS データとして整備したものです。
https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N03-v2_4.html



ダウンロードしたファイル N03-20200101_13_GML.zip を現在の作業フォルダにコピーします。

Step 2: シェープファイルを Stata 形式に変換する

● 地価データと同様に、シェープファイルの解凍（展開）から空間設定までの作業を行います。ここではコマンド操作のみ記載します。do ファイル **part1.do** 内の『地図データの準備』をご確認ください。

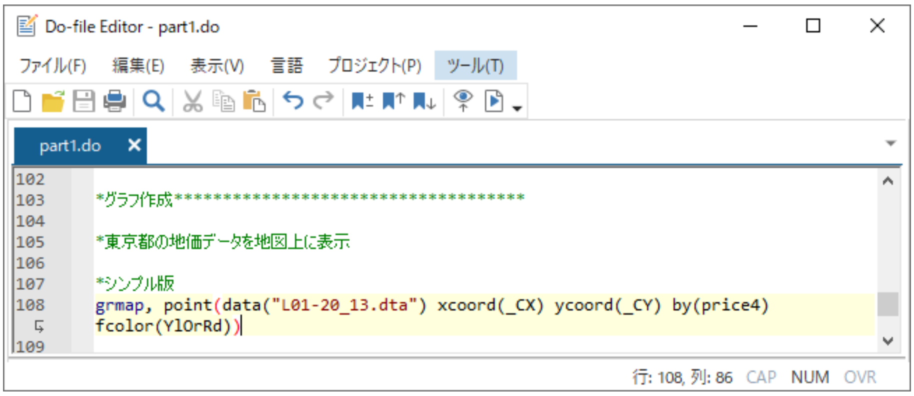


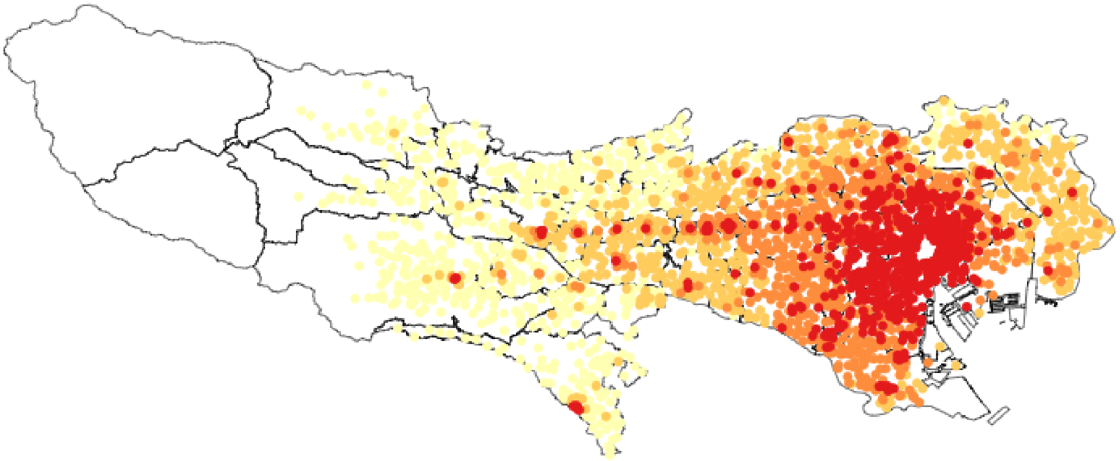
3.　グラフ作成

- 東京都の地価データを地図上に表示します。空間データ用のグラフ作成コマンド `grmap` を使用します。do ファイル part1.do 内の『グラフ作成』をご確認ください。

シンプル版

- 現在 Stata で読み込んでいる地図データ(N03-20_13_200101.dta)上に、地価データ(L01-20_13.dta)をポイント表示し、`price4` の値によって色分けしたカラーマップで表現します。

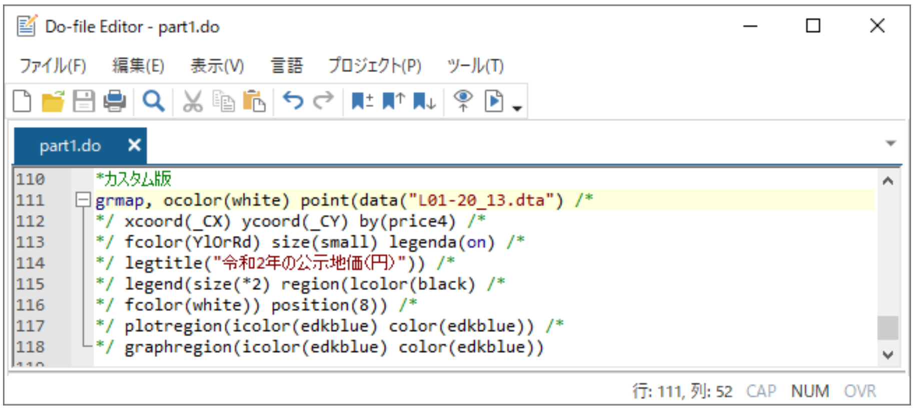


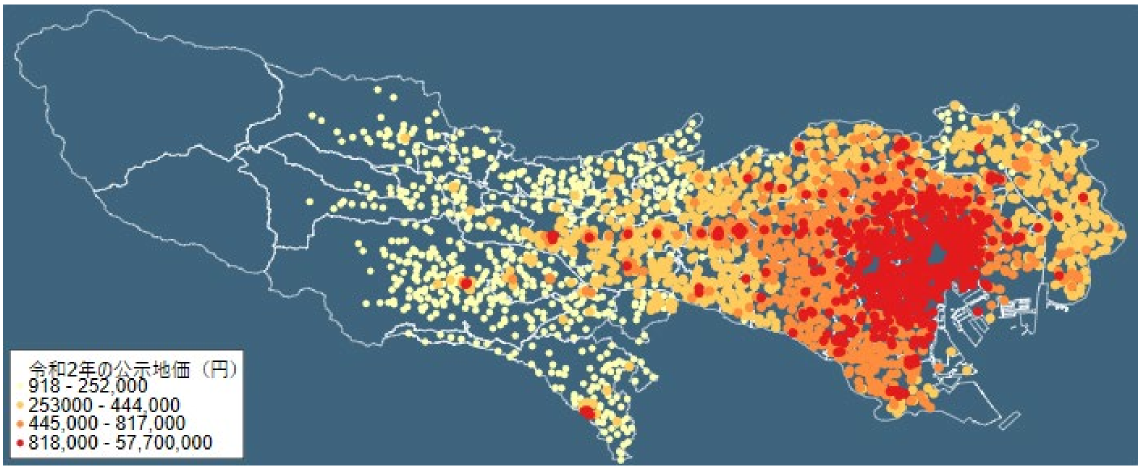


LightStone® 株式会社ライトストーン

カスタム版

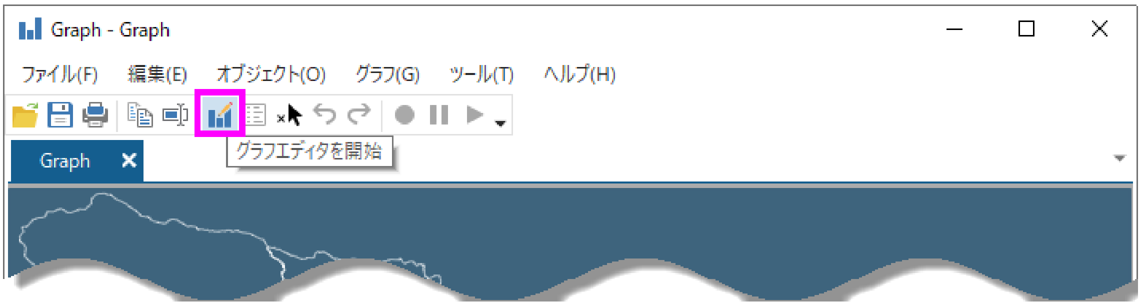
● 地図の枠線の色を変更、凡例と背景色を追加します。





✧ グラフエディタ機能

グラフウィンドウからグラフエディタを起動して、作成済みのグラフを編集できます。グラフ上の編集したい箇所（背景やプロットなど）をクリックすると編集メニューが表示されます。



LightStone® 株式会社ライトストーン

第二部　空間自己回帰モデル - 失業率と大学卒業率の関係をモデル化

はじめに

- 経済や医学的な状況は、隣接する地区では相互に強い影響を及ぼします。空間自己回帰モデルは他のモデル推定とは異なり、地理的な位置情報付きのデータに対してモデルをフィットします。
- 地図情報としては一般的に「シェープファイル」がよく利用されます。シェープファイルは図形情報と属性情報を持った地図データファイル群で、政府機関などのウェブサイトで配布されています。
- シェープファイルを用いた空間自己回帰モデル分析は次の手順で行います。

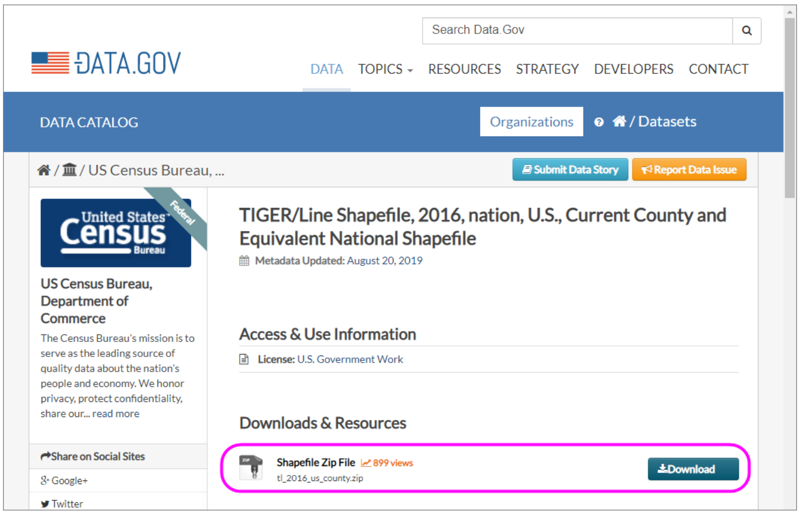
Step 1	シェープファイルをダウンロードする
Step 2	シェープファイルを Stata 形式に変換する
Step 3	データセットと Stata 形式のシェープファイルをマージする
Step 4	空間自己回帰モデル分析を行う

- この例題では、アメリカ合衆国テキサス州の失業率と大学卒業率の関係をモデル化します。メニュー操作を基本としていますが、一部コマンド操作のものもあります。do ファイル part2.do に全ての操作のコマンドをまとめています。

データの準備

Step 1: アメリカ合衆国のシェープファイルをダウンロードする

- こちらのページからシェープファイルをダウンロードします。
https://catalog.data.gov/dataset/tiger-line-shapefile-2016-nation-u-s-current-county-and-equivalent-national-shapefile

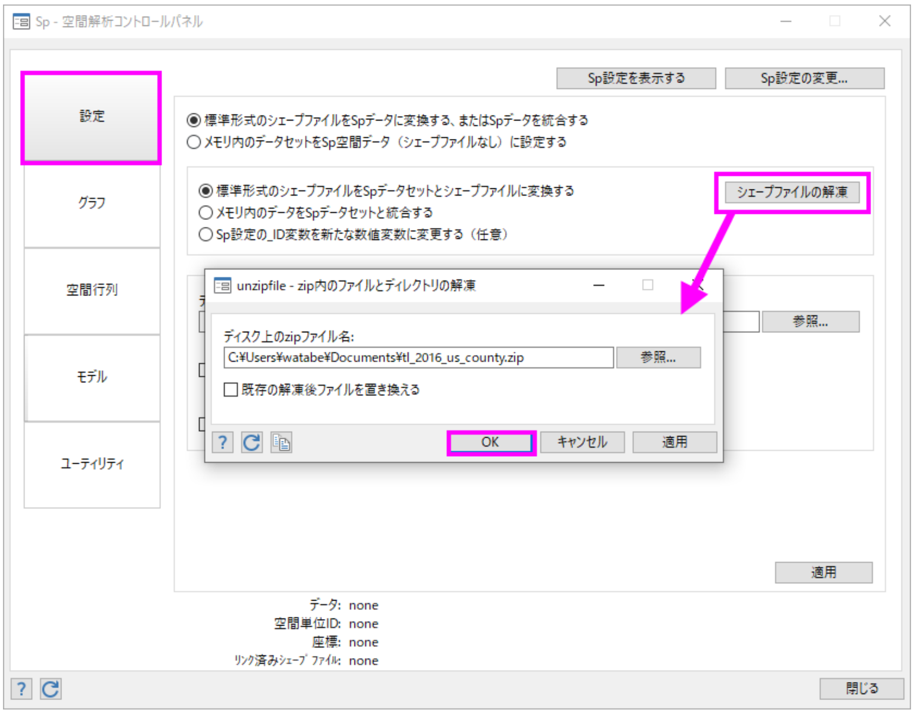


LightStone® 株式会社ライトストーン

ダウンロードしたファイル tl_2016_us_county.zip を現在の作業フォルダにコピーします。

Step 2: シェープファイルを Stata 形式に変換する

- ダウンロードしたシェープファイルを解凍（展開）します。Stata を起動し、メニューから「**統計 > 空間自己回帰モデル**」を選択します。「空間解析コントロールパネル」で次のように設定します。

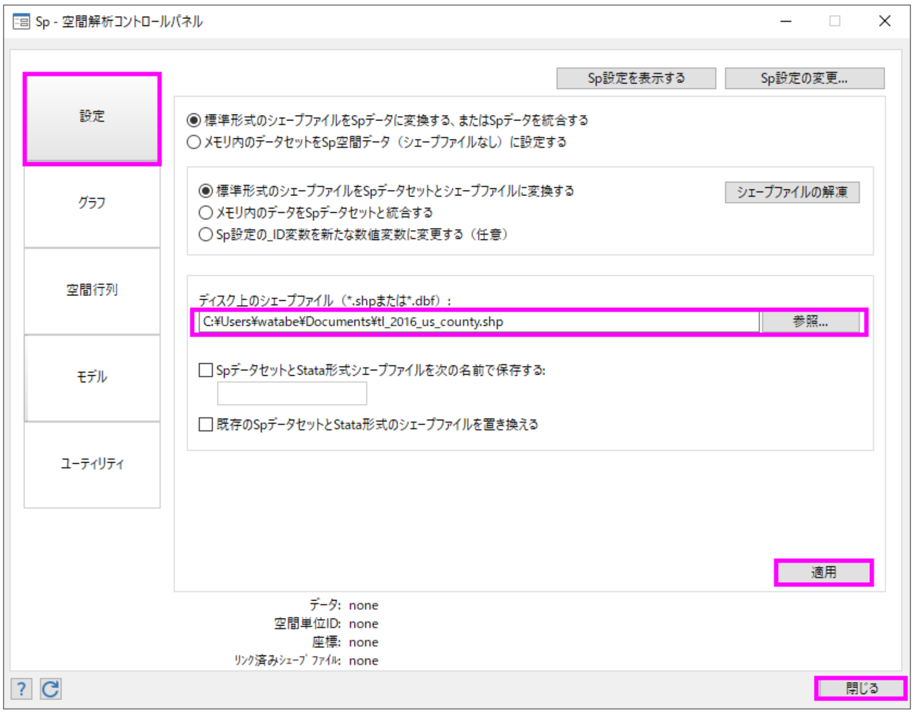


結果画面に解凍結果が表示されます。

```
  inflating: tl_2016_us_county.cpg
 inflating:  tl_2016_us_county.dbf
 inflating: tl_2016_us_county.prj
 inflating: tl_2016_us_county.shp
 inflating: tl_2016_us_county.shp.ea.iso.xml
 inflating: tl_2016_us_county.shp.iso.xml
 inflating: tl_2016_us_county.shp.xml
 inflating: tl_2016_us_county.shx

successfully unzipped C:\Users\watabe\Documents\tl_2016_us_county.zip to current directory
total processed:  8
        skipped:  0
      extracted:  8
```


LightStone® 株式会社ライトストーン

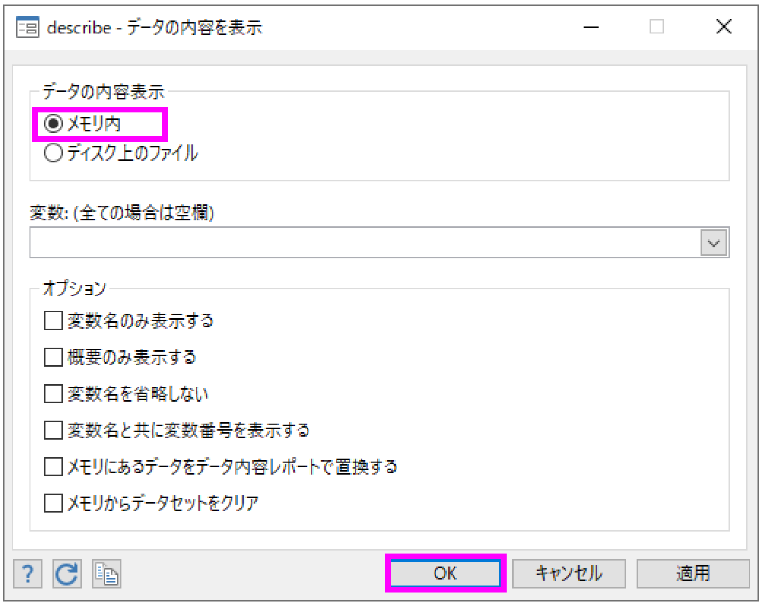
- シェープファイルを Stata のデータ形式に変換します。次のように設定して、「空間解析コントロールパネル」を閉じます。




```
(importing .shp file)
(importing .dbf file)
(creating _ID spatial-unit id)
(creating _CX coordinate)
(creating _CY coordinate)

file tl_2016_us_county_shp.dta created
file tl_2016_us_county.dta     created
```

- メニューから「**ファイル ＞ 開く**」を選択して、tl_2016_us_county.dta を開きます。

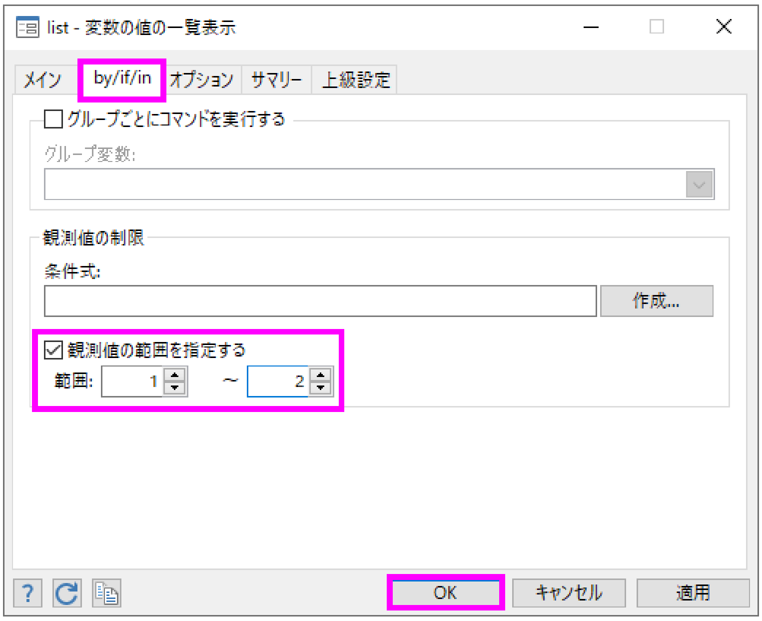
● データの内容を確認します。メニューから「データ > データの内容表示 > メモリ/ファイル内のデータの内容表示」を選択して、次のように設定します。




```
Contains data from tl_2016_us_county.dta
  obs:         3,233
 vars:            20                          29 Oct 2020 13:40
--------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
--------------------------------------------------------------------------------
_ID             int     %12.0g                Spatial-unit ID
_CX             double  %10.0g                x-coordinate of area centroid
_CY             double  %10.0g                y-coordinate of area centroid
STATEFP         str2    %9s                   STATEFP
COUNTYFP        str3    %9s                   COUNTYFP
COUNTYNS        str8    %9s                   COUNTYNS
GEOID           str5    %9s                   GEOID
NAME            str21   %21s                  NAME
NAMELSAD        str33   %33s                  NAMELSAD
LSAD            str2    %9s                   LSAD
CLASSFP         str2    %9s                   CLASSFP
MTFCC           str5    %9s                   MTFCC
CSAFP           str3    %9s                   CSAFP
CBSAFP          str5    %9s                   CBSAFP
METDIVFP        str5    %9s                   METDIVFP
FUNCSTAT        str1    %9s                   FUNCSTAT
ALAND           double  %14.0f                ALAND
AWATER          double  %14.0f                AWATER
INTPTLAT        str11   %11s                  INTPTLAT
INTPTLON        str12   %12s                  INTPTLON
--------------------------------------------------------------------------------
Sorted by: _ID
```


LightStone® 株式会社ライトストーン

- データの 1, 2 行目の内容を確認します。メニューから「データ > データの内容表示 > データの一覧表示」を選択します。「by/if/in」タブで次のように設定します。



1.

_ID	_CX	_CY	STATEFP	COUNTYFP	COUNTYNS	GEOID
1	-96.7874	41.916403	31	039	00835841	31039

NAME	NAMELSAD	LSAD	CLASSFP	MTFCC	CSAFP	CBSAFP
Cuming	Cuming County	06	H1	G4020		

METDIVFP	FUNCSTAT	ALAND	AWATER	INTPTLAT
	A	1477895811	10447360	+41.9158651

INTPTLON
-096.7885168

2.

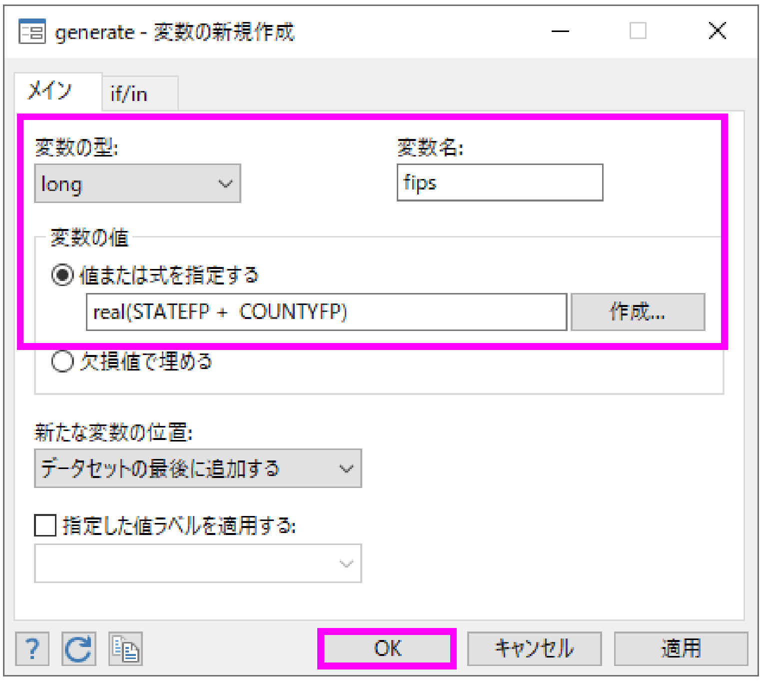
_ID	_CX	_CY	STATEFP	COUNTYFP	COUNTYNS	GEOID
2	-123.43347	46.291134	53	069	01513275	53069

NAME	NAMELSAD	LSAD	CLASSFP	MTFCC	CSAFP	CBSAFP
Wahkiakum	Wahkiakum County	06	H1	G4020		

METDIVFP	FUNCSTAT	ALAND	AWATER	INTPTLAT
	A	680956787	61588406	+46.2946377

INTPTLON
-123.4244583

- 変数 STATEFP は州を表す 2 桁のコード、変数 COUNTYFP は州内の群を表す 3 桁のコードです。これらを組み合わせて、個々の群を表す 5 桁のコードの数値型変数 **fips** を作成します。メニューから「**データ > データの作成または変更 > 新規変数の作成**」を選択し、次のように設定します。**real()**関数は文字列を数値に変換します。

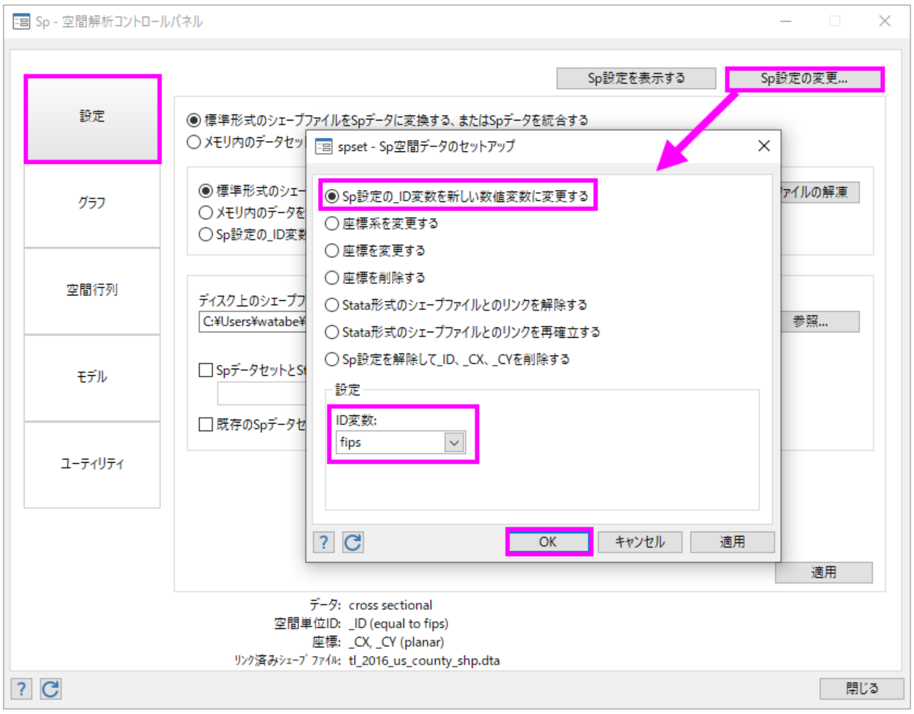


- 新しい変数 **fips** が観測値を一意に識別できているか確認します（コマンド操作のみ。/**/で囲まれたコメントは入力しません)。**assert** はそれに続く式が正しいかどうかを確認するコマンドで、正しくない場合は”**assertion is false**”のメッセージが表示されます。結果が正しければ何も表示しません。
  _N は Stata のシステム変数です（**help _variables** を参照)。

```
bysort fips: assert _N==1/*fips の各観測値に値が 1 つだけ存在することを確認*/

assert fips != ./* fips に欠損値がないことを確認*/
```

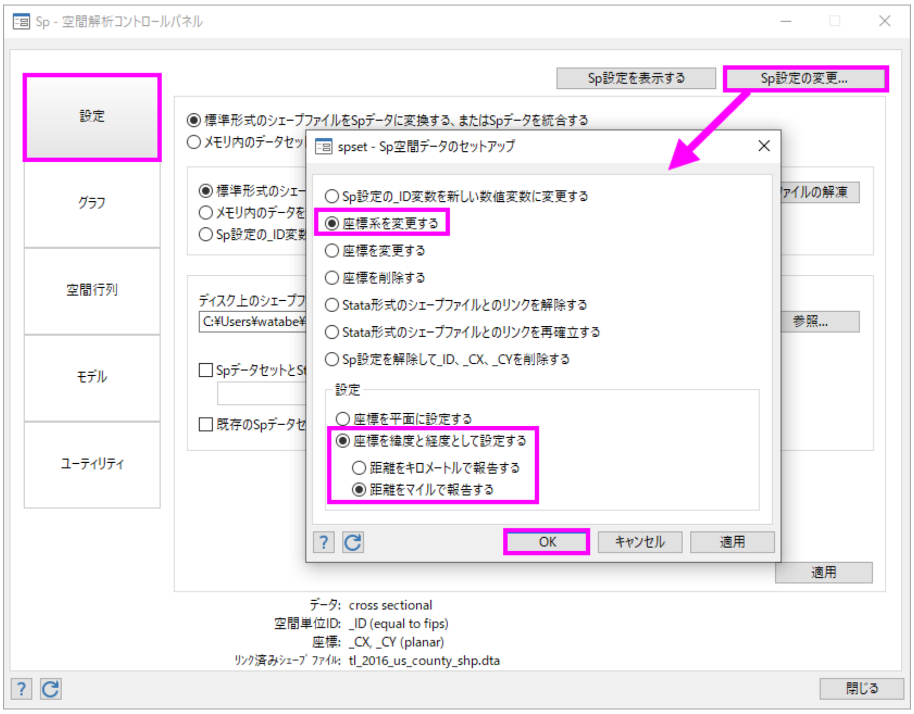

LightStone® 株式会社ライトストーン

- 空間データを扱うにあたって、どの変数を空間の ID とするか設定する必要があります。Stata 形式のシェープファイルを作成した際に自動で設定されますが、後から変更することも可能です。メニューから「**統計 > 空間自己回帰モデル**」を選択して「空間解析コントロールパネル」を開き、先程作成した **fips** を指定します。




```
(_shp.dta file saved)
(data in memory saved)
Sp dataset tl_2016_us_county.dta
              data:  cross sectional
   spatial-unit id:  _ID (equal to fips)
       coordinates:  _CX, _CY (planar)
  linked shapefile:  tl_2016_us_county_shp.dta
```

- デフォルトでは、座標は平面(planar)単位に設定されます。元のシェープファイルは緯度経度で記録されているため、座標単位の設定を変更します。

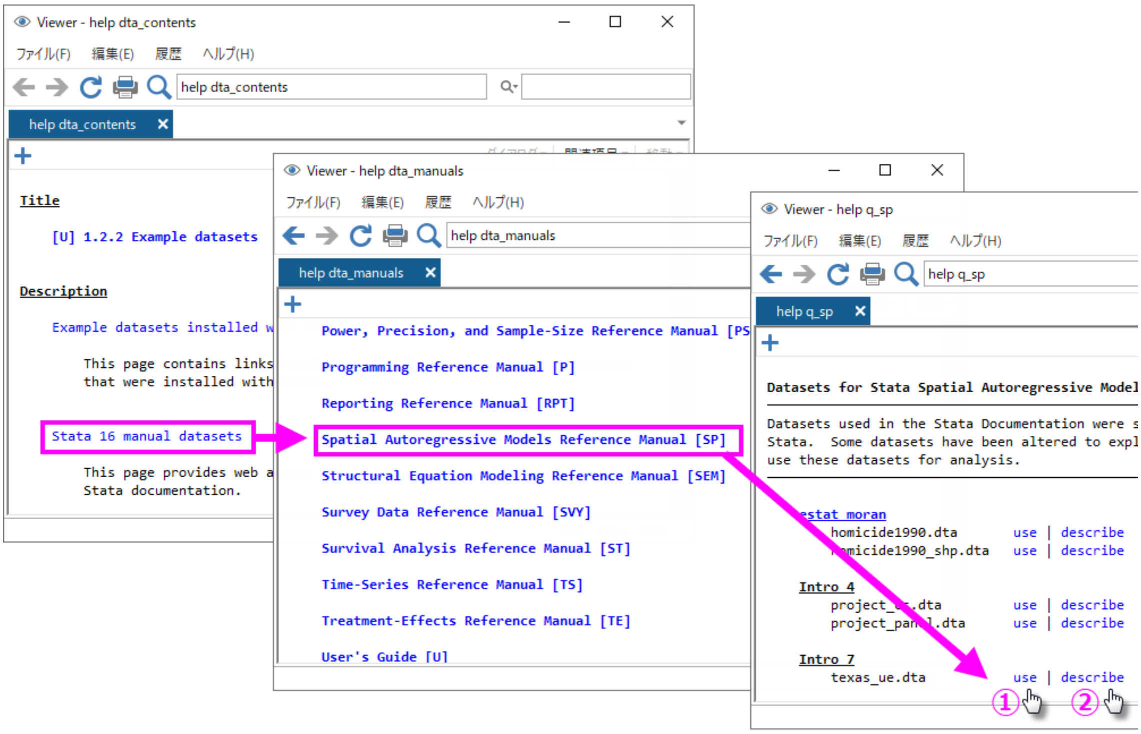



```
Sp dataset tl_2016_us_county.dta
              data:   cross sectional
   spatial-unit id:   _ID (equal to fips)
       coordinates:   _CY, _CX (latitude-and-longitude, miles)
  linked shapefile:   tl_2016_us_county_shp.dta
```

- メニューから「ファイル > 保存」を選択してデータを保存します。以上で地図データの準備は完了です。

Step 3: データセットと Stata 形式のシェープファイルをマージする

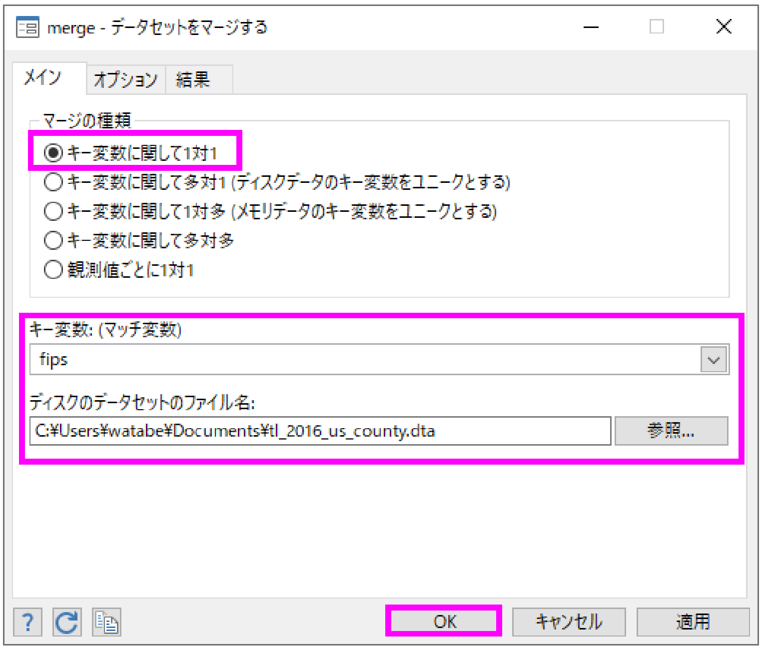
● テキサス州の失業率と大学卒業率のデータを使用します。メニューから「**ファイル > 例題データセット**」を選択して、次のように操作します。



データが開き、その内容が表示されます。個々の群を表す 5 桁のコード **fips** が含まれています。

```
Contains data
  obs:           254                          10 Feb 2019 12:36
 vars:             4
------------------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------------------------
fips            float   %9.0g                 FIPS
college         float   %9.0g                 Percent college degree
income          long    %12.0g                Median household income
unemployment    float   %9.0g                 Unemployment rate
------------------------------------------------------------------------------------------------
Sorted by: fips
```

- Stata 形式のシェープファイル tl_2016_us_county.dta と共通の変数 **fips** を用いて、2 つのファイルをマージします。メニューから「**データ ＞ データセットの統合 ＞2 データセットのマージ**」を選択して、次のように設定します。

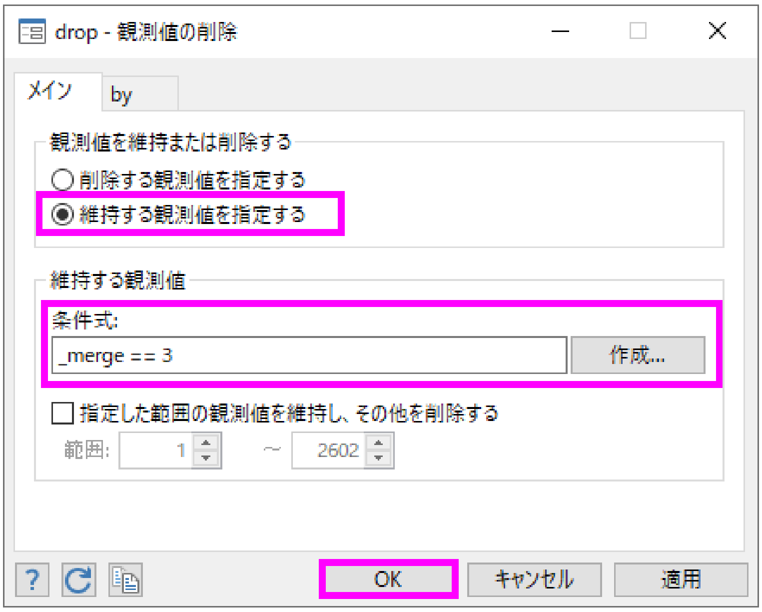



```
(note: variable fips was float, now double to accommodate using data's values)

    Result                           # of obs.
    -----------------------------------------
    not matched                         2,979
        from master                         0  (_merge==1)
        from using                      2,979  (_merge==2)

    matched                               254  (_merge==3)
    -----------------------------------------
```

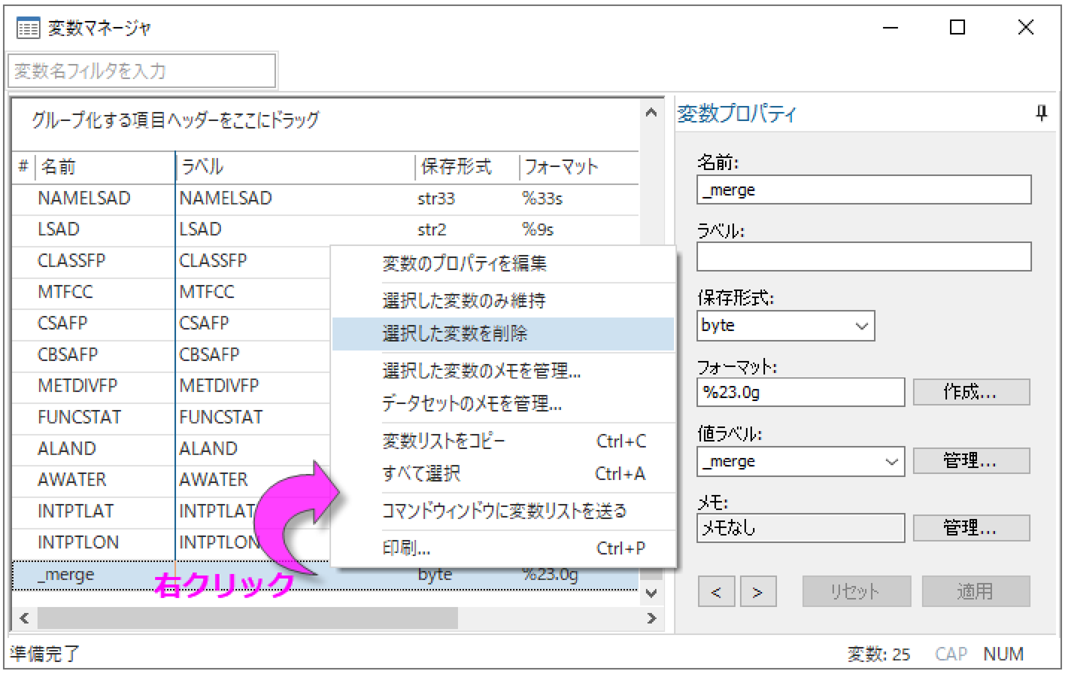
変数 **_merge** が作成されます。テキサス州のデータとマッチしたものは **_merge==3** と識別されています。

LightStone® 株式会社ライトストーン

- マッチしなかった観測値を削除します。メニューから「**データ ＞ データの作成または変更 ＞ 観測値の削除**」を選択して、次のように設定します。2,979 の観測値が削除されます。



- 今後、変数 `_merge` は使用しないため削除します。メニューから「**データ ＞ 変数マネージャ**」を選択して、次のように操作します。

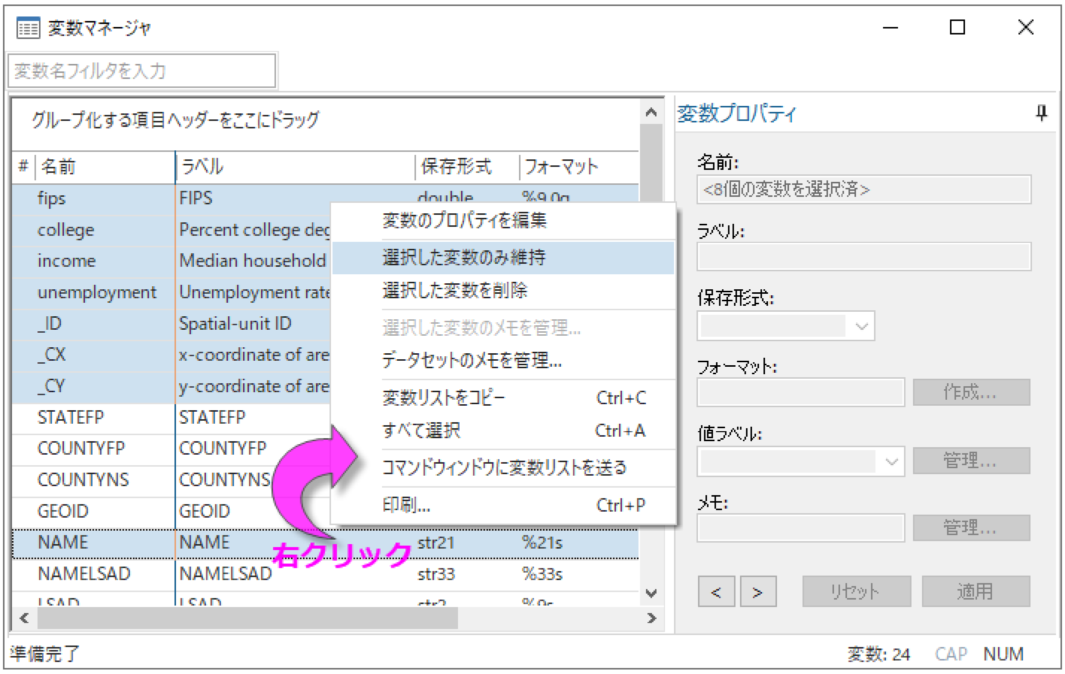


● データセットの内容を確認します。メニューから「データ ＞ データの内容表示 ＞ メモリ/ファイル内のデータの内容表示」を選択して、「メモリ内」のみ選び OK を押します。

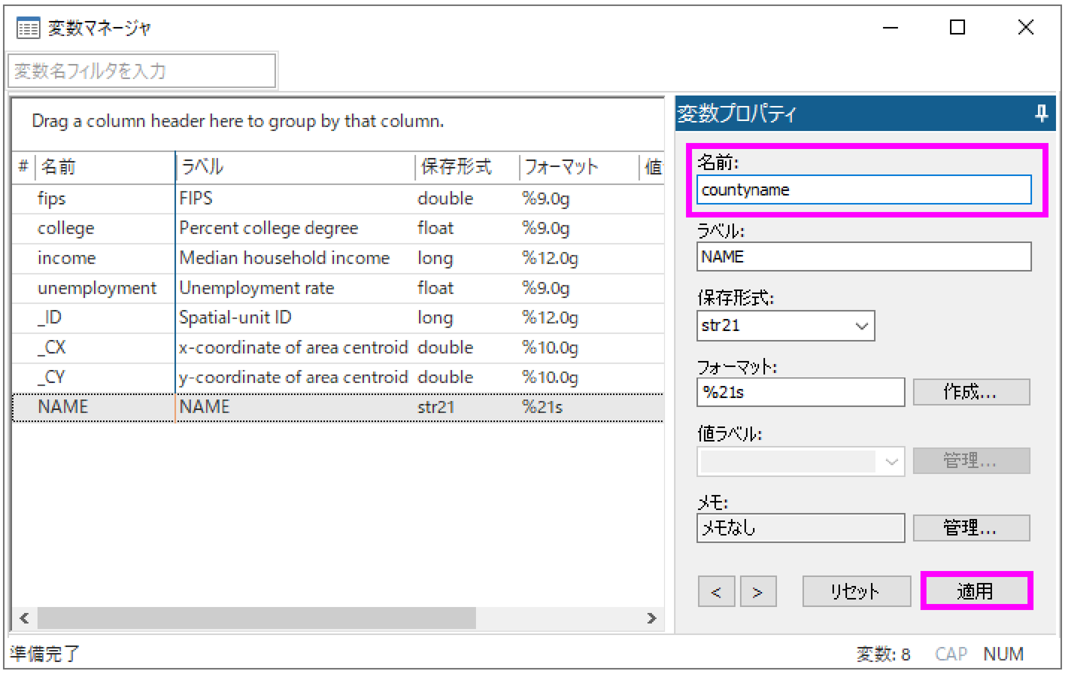
```
Contains data from texas_ue.dta
  obs:           254
 vars:            24                          29 Oct 2020 15:43
                                              (_dta has notes)
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
fips            double  %9.0g                 FIPS
college         float   %9.0g               * Percent college degree
income          long    %12.0g                Median household income
unemployment    float   %9.0g                 Unemployment rate
_ID             long    %12.0g                Spatial-unit ID
_CX             double  %10.0g                x-coordinate of area centroid
_CY             double  %10.0g                y-coordinate of area centroid
STATEFP         str2    %9s                   STATEFP
COUNTYFP        str3    %9s                   COUNTYFP
COUNTYNS        str8    %9s                   COUNTYNS
GEOID           str5    %9s                   GEOID
NAME            str21   %21s                  NAME
NAMELSAD        str33   %33s                  NAMELSAD
LSAD            str2    %9s                   LSAD
CLASSFP         str2    %9s                   CLASSFP
MTFCC           str5    %9s                   MTFCC
CSAFP           str3    %9s                   CSAFP
CBSAFP          str5    %9s                   CBSAFP
METDIVFP        str5    %9s                   METDIVFP
FUNCSTAT        str1    %9s                   FUNCSTAT
ALAND           double  %14.0f                ALAND
AWATER          double  %14.0f                AWATER
INTPTLAT        str11   %11s                  INTPTLAT
INTPTLON        str12   %12s                  INTPTLON
                                            * indicated variables have notes
-------------------------------------------------------------------------------
Sorted by:
```

- シェープファイル tl_2016_us_county.dta 由来の不要な変数が多くあるため、これらを削除します。メニューから「**データ > 変数マネージャ**」を選択して、次の 8 つの変数のみ維持します。

`fips  college  income  unemployment  _ID  _CX  _CY  NAME`



- 変数 `NAME` の名前を `countyname` に変更します。「適用」をクリックして変数マネージャを閉じます。



LightStone® 株式会社ライトストーン

- データセットを保存します。例題データセットは上書き保存ができないため、メニューから「ファイル > 名前を付けて保存」を選択して、texas_ue.dta という名前で保存します。以上でデータの準備は完了です。

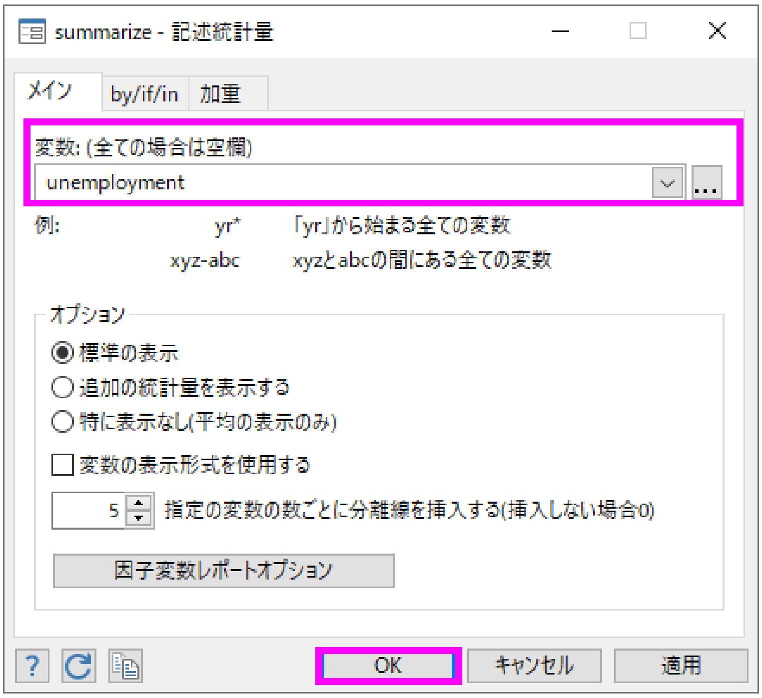
Step 4: テキサス州のデータの分析

- これまでの操作で、texas_ue.dta を空間自己回帰モデルで使用する準備が整いました。改めてデータセットの内容を確認します。メニューから「データ > データの内容表示 > メモリ/ファイル内のデータの内容表示」を選択して、「メモリ内」のみ選び OK を押します。

```
Contains data from texas_ue.dta
  obs:           254
 vars:             8                         29 Oct 2020 15:52
                                             (_dta has notes)
--------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
--------------------------------------------------------------------------------
fips            double  %9.0g                 FIPS
college         float   %9.0g               * Percent college degree
income          long    %12.0g                Median household income
unemployment    float   %9.0g                 Unemployment rate
_ID             long    %12.0g                Spatial-unit ID
_CX             double  %10.0g                x-coordinate of area centroid
_CY             double  %10.0g                y-coordinate of area centroid
countyname      str21   %21s                  NAME
                                            * indicated variables have notes
--------------------------------------------------------------------------------
Sorted by:
```


LightStone® 株式会社ライトストーン

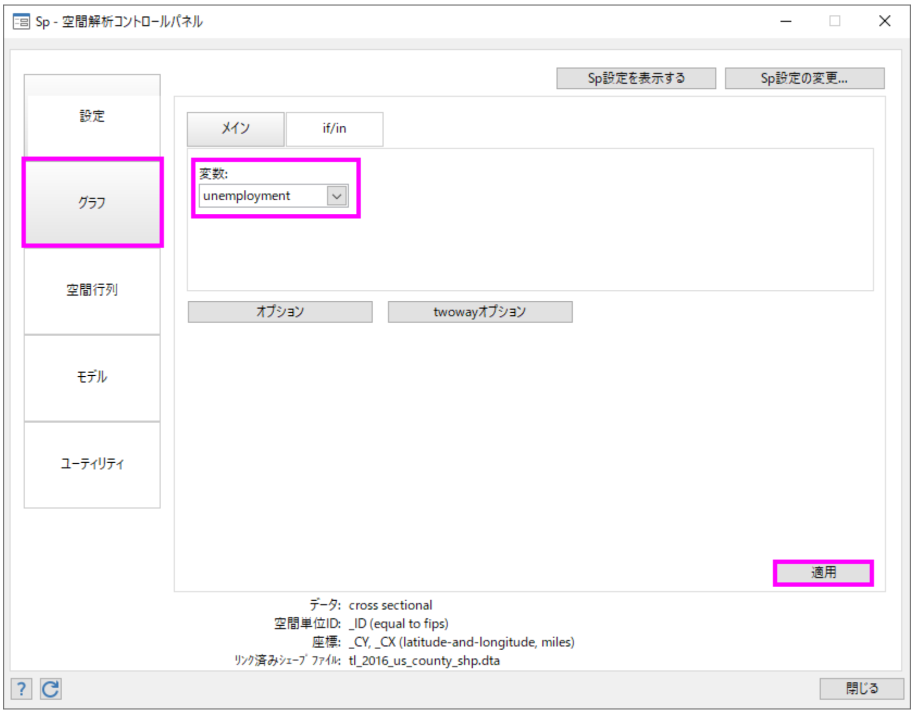
- 失業率の記述統計量を求めます。メニューから「統計 ＞ 要約/表/検定 ＞ 要約と記述統計量 ＞ 記述統計量」を選択して、次のように設定します。

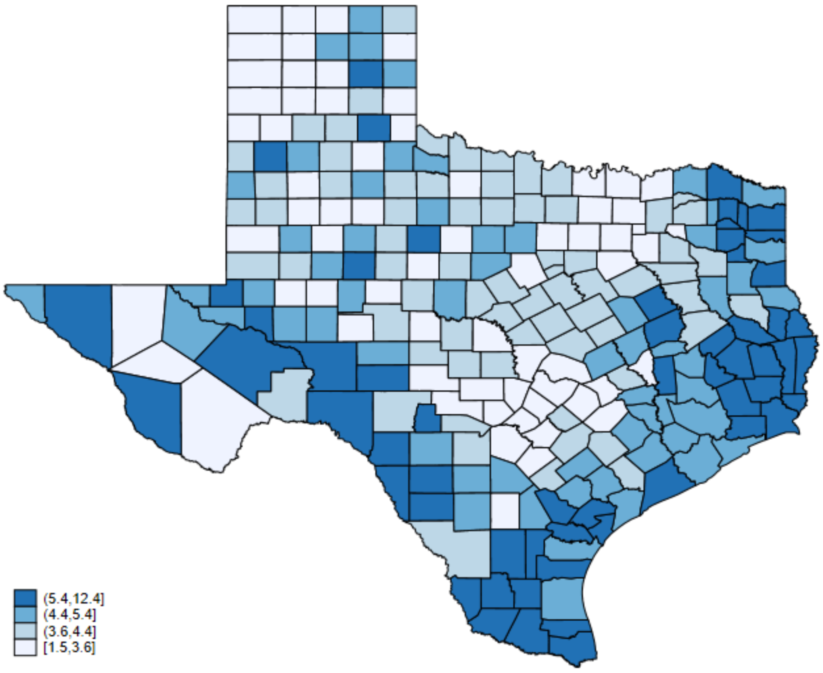


結果は次のようになります。失業率の範囲は 1.5%から 12.4%であることがわかります。

Variable	Obs	Mean	Std. Dev.	Min	Max
unemployment	254	4.731102	1.716514	1.5	12.4

● texas_ue.dta は空間データとして設定されており、シェープファイルを含むため、コロプレスマップを作成できます。メニューから「**統計 > 空間自己回帰モデル**」を選択して、「空間解析コントロールパネル」で次のように設定します。





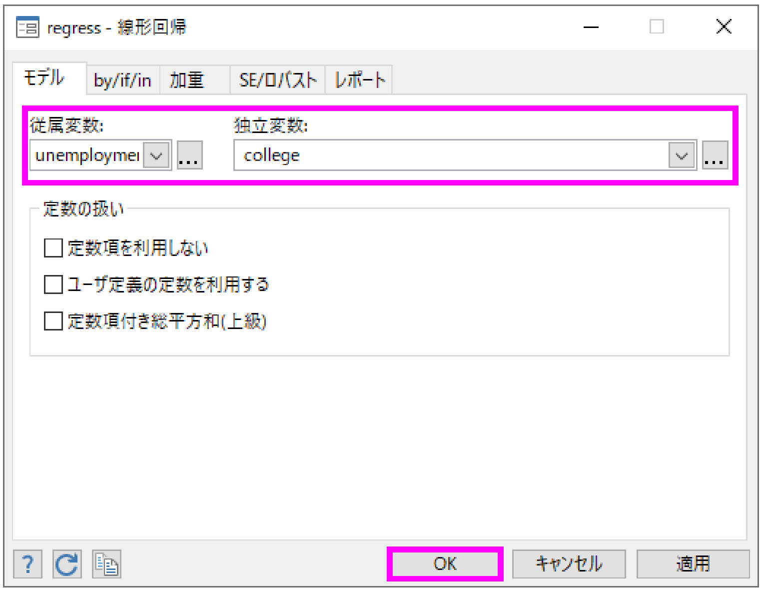
グラフから、失業率が高い群はクラスター化しているように見えます。これは、郡間に波及効果があることを示唆しています。

LightStone® 株式会社ライトストーン

通常の線形回帰が適切であるか検定する

- このデータセットは空間分析と非空間分析の両方に使用できます。
- 大学卒業率を説明変数として、失業率の線形回帰を実行します。
- 回帰を実行した後、モデルの残差が空間的に相関しているかどうかを判断します。

- 線形回帰を実行します。メニューから「**統計 > 線形モデル他 > 線形回帰**」を選択して、次のように設定します。



Source	SS	df	MS
Model	139.314746	1	139.314746
Residual	606.129539	252	2.40527595
Total	745.444285	253	2.9464201

Number of obs	=	254
F(1, 252)	=	57.92
Prob > F	=	0.0000
R-squared	=	0.1869
Adj R-squared	=	0.1837
Root MSE	=	1.5509

unemployment	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
college	-.1008791	.0132552	-7.61	0.000	-.1269842	-.0747741
_cons	6.542796	.2571722	25.44	0.000	6.036316	7.049277

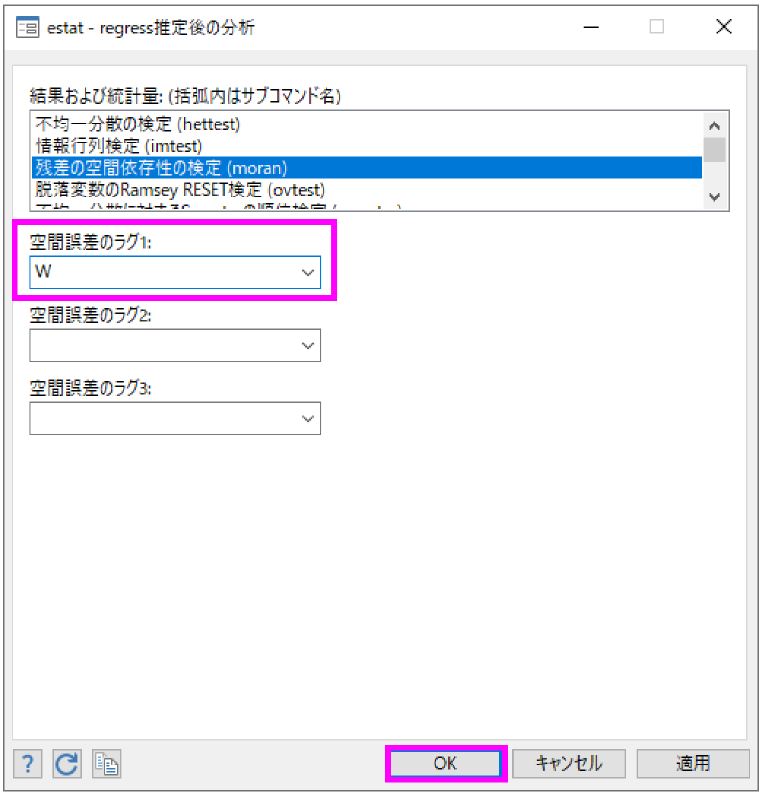
結果から、この過度に単純化されたモデルは、大学卒業率によって失業率が著しく減少することを示しています。

- 残差が空間的にクラスター化されているかどうかを調べます。Moran 検定を使用して、通常の線形回帰でフィットされたモデルの残差が近くの残差と相関しているかどうかを判断します。
- 検定を行うために「近く」を定義する必要があります。`spmatrix` コマンドで空間重み行列を定義します。

- 「近く」を「境界を共有する」と定義して、W という名前の隣接行列を作成します。メニューから「**統計 > 空間自己回帰モデル**」を選択して、「空間解析コントロールパネル」で次のように設定します。

LightStone® 株式会社ライトストーン

- Moran 検定を実行します。メニューから「**統計 > 線形モデル他 > 回帰診断 > モデル選択のための検定など**」を選択し、次のように設定します。



帰無仮説	モデルの残差が独立で一様分布している（i.i.d.）
対立仮説	モデルの残差が、W で定義される近くの残差と相関している

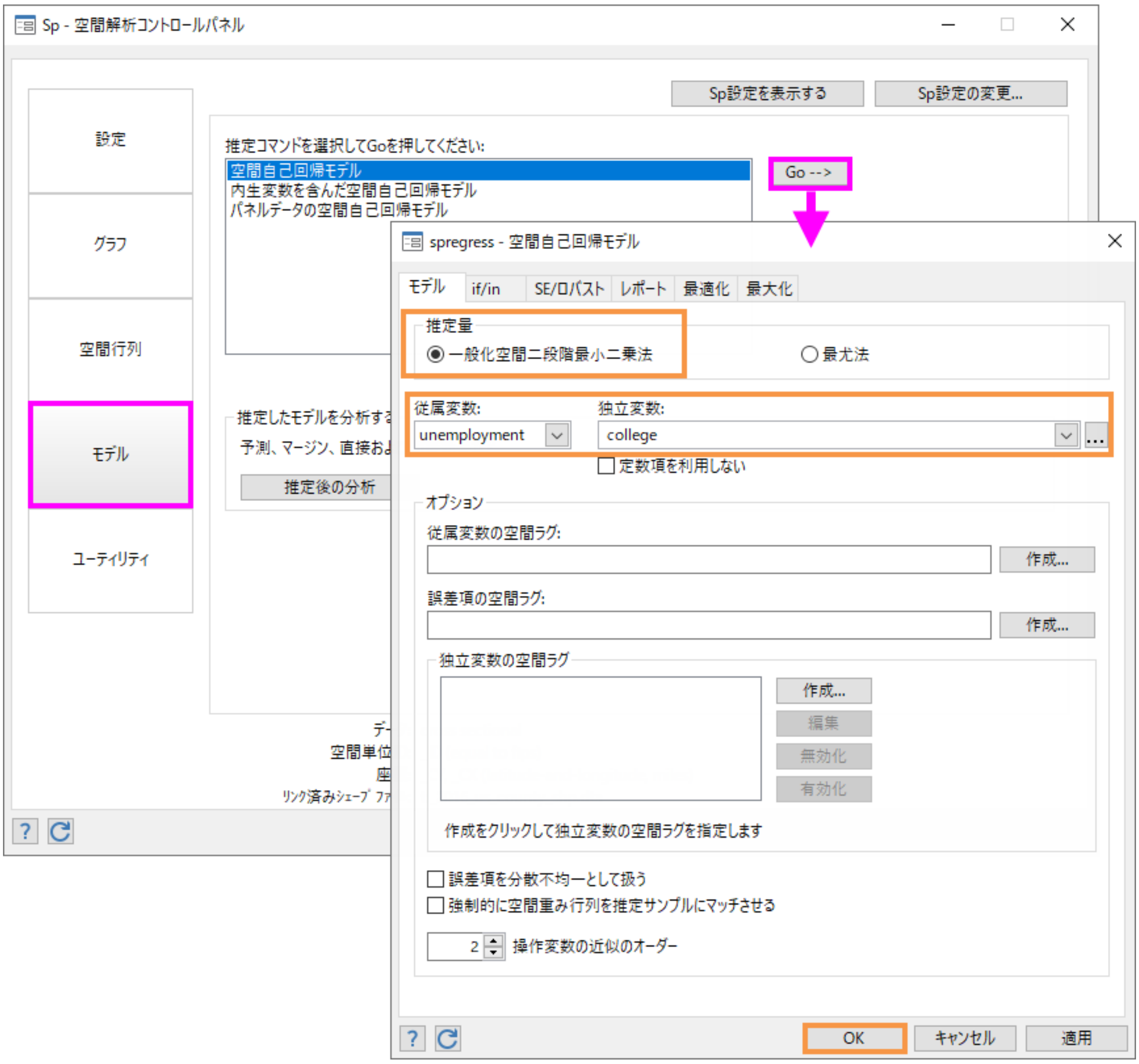
```
Moran test for spatial dependence
         Ho: error is i.i.d.
         Errorlags:  W

         chi2(1)      =    94.06
         Prob > chi2  =   0.0000
```

結果から、帰無仮説が棄却されます。

線形回帰の結果を空間自己回帰で再現する

- 空間自己回帰のコマンドは spregress です。spregress は、空間重み行列 W で定義される、観測値が独立でないモデルをフィットします。
- 先に行った通常の線形回帰（regress コマンド）では、各郡が独立であるという仮定のもとモデルをフィットしました。spregress でも同じモデルをフィットして同じ結果を得ることができます。

- spregress で一般化空間二段階最小二乗法を使用する場合、メニューから「統計 > 空間自己回帰モデル」を選択して、「空間解析コントロールパネル」で次のように設定します。



```
  (254 observations)
  (254 observations (places) used)

Spatial autoregressive model                    Number of obs     =        254
GS2SLS estimates                                Wald chi2(1)      =      58.38
                                                Prob > chi2       =     0.0000
                                                Pseudo R2         =     0.1869

------------------------------------------------------------------------------
unemployment |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     college |  -.1008791   .0132029    -7.64   0.000    -.1267563   -.075002
       _cons |   6.542796   .2561577    25.54   0.000     6.040736    7.044856
------------------------------------------------------------------------------
```

- フィット結果の係数は通常の線形回帰(**regress** コマンド)の結果と同じですが、標準誤差がわずかに異なります。また、**regress** では F 統計量が報告されているのに対し、**spregress** では Z 統計量とカイ二乗統計量が報告されています。**spregress** は **regress** と異なり、有限標本調整を行いません。これは、適切な状況で調整が行われることを想定していないためです。

LightStone® 株式会社ライトストーン

被説明変数の空間ラグでモデルをフィットする

- 失業率の空間ラグを追加して、空間自己回帰モデルをフィットします。モデルは次の通りです。

$$y_{ue} = \beta_0 + \beta_1 X_{cr} + \beta_2 W y_{ue} + \epsilon$$

y_{ue}	失業率（使用データセット中の変数 unemployment）
X_{cr}	大学卒業率（使用データセット中の変数 college）

- モデルには$\beta_2 W y_{ue}$項が含まれます。失業率は近くの群から波及すると想定します。このようなモデルには実際の理論があり、失業率の高い群の労働者は周囲の群で雇用を探すと期待されます。

- 空間自己回帰モデルには二通りのフィット方法があります。
  1. 一般化空間二段階最小二乗法
  2. 最尤法

- 誤差が正規分布しているとき、最尤法は一般化空間二段階最小二乗法よりも統計的に効率的です。一方、一般化空間二段階最小二乗法は、正規性の違反に対して堅牢です。最尤法と一般化空間二段階最小二乗法のどちらを選択するかは、正規性の仮定に依存します。
- 最尤法でも標準誤差の種類に「ロバスト」を選択（コマンド操作の場合は vce(robust) オプション）すれば、正規性の違反に対して堅牢になります。
- さらに、最尤法は一般化空間二段階最小二乗法に比べて実行に時間がかかり、観測数が増えるほど計算時間が長くなります。この例題では今後一般化空間二段階最小二乗法を使用します。

LightStone® 株式会社ライトストーン

● 被説明変数の空間ラグでモデルをフィットします。「空間解析コントロールパネル」で次のように設定します。

```
  (254 observat ions)
  (254 observations (places) used)
  (weighting matrix defines 254 places)

Spatial autoregressive model                        Number of obs     =        254
GS2SLS estimates                                    Wald chi2(2)      =      67.66
                                                    Prob > chi2       =     0.0000
                                                    Pseudo R2         =     0.1453
```

unemployment	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
unemployment						
college	-.0939834	.0131033	-7.17	0.000	-.1196653	-.0683015
_cons	5.607379	.5033813	11.14	0.000	4.620769	6.593988
W						
unemployment	.2007728	.0942205	2.13	0.033	.016104	.3854415

```
Wald test of spatial terms:          chi2(1) = 4.54       Prob > chi2 = 0.0331
```

- β_0とβ_1の係数は通常の線形回帰（regress コマンド）の結果と似ていますが、これは偶然です。通常、波及効果が大きい場合、他のパラメータが変化します。
- 一方、β_2（Wy_{ue}の係数）は有意であることがわかりますが、正確には推定されていません。β_2の信頼区間は［0.02, 0.39］となっています。
- 表の下に Wald 検定の結果が表示されています。ここでの帰無仮説は「モデルの空間ラグ項の係数がゼロである」です。結果から、帰無仮説が棄却されているため、モデルに空間ラグ項を追加した意味があると考えられます。

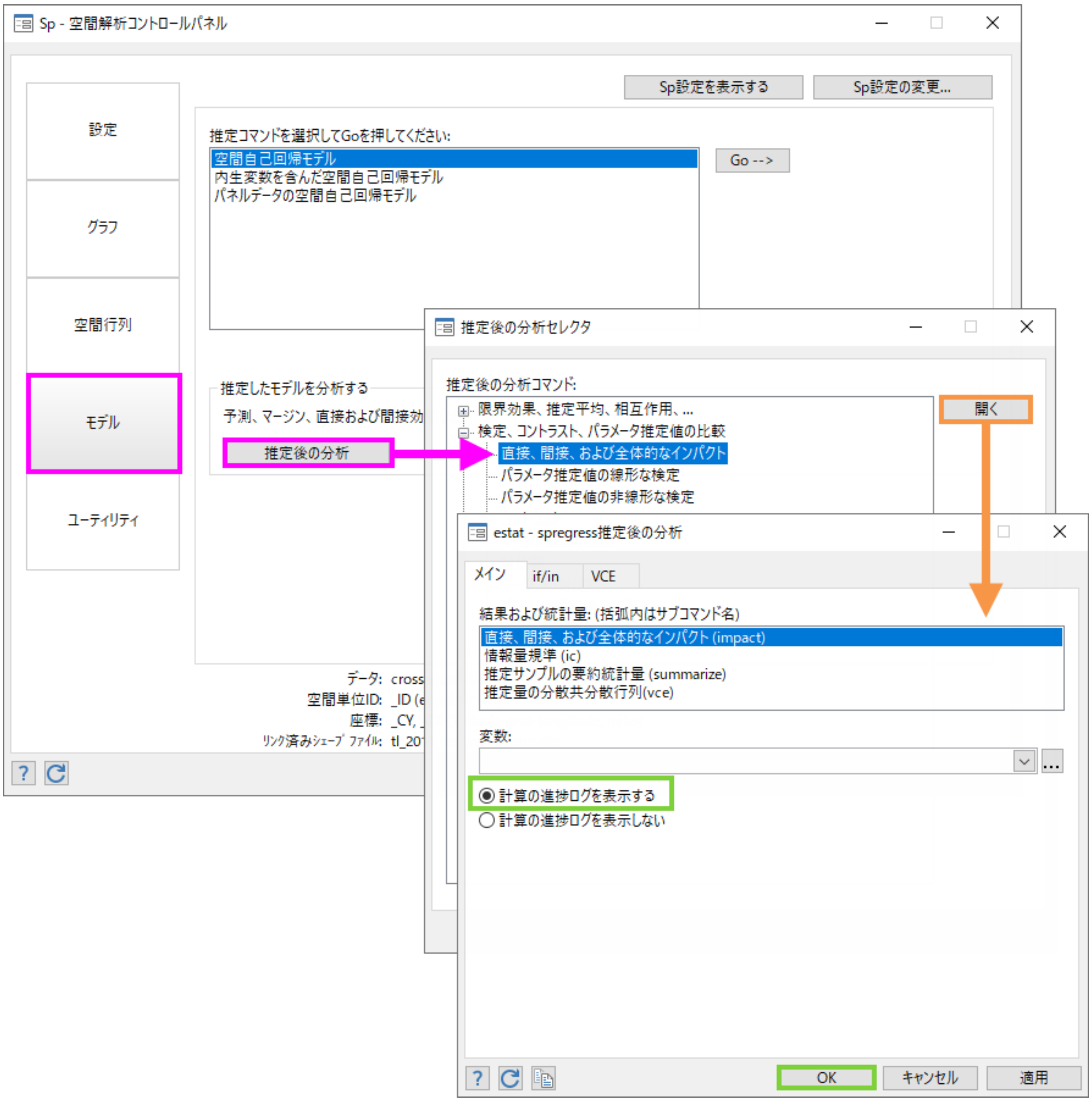
被説明変数の空間ラグを使用したモデルの解釈

- β_1を大学卒業率の直接効果、β_2を波及効果と考えることはできません。これらの効果の再帰的計算の要素です。

$$y_{ue} = \beta_0 + \beta_1 X_{cr} + \beta_2 W y_{ue} + \epsilon$$

- X_{cr}が増加すると、y_{ue}がβ_1減少し、y_{ue}の減少が波及して、$\beta_2 W$のy_{ue}がさらに減少し、その減少が波及して、y_{ue}がさらに減少します。

- 推定後の分析として、`estat impact` コマンドを使用して、平均的な効果を求めることができます。「空間解析コントロールパネル」で次のように設定します。



```
progress   :100%

Average impacts                              Number of obs     =        254

------------------------------------------------------------------------------
             |            Delta-Method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
direct       |
     college |  -.0945245   .0130576    -7.24   0.000     -.120117   -.0689321
-------------+----------------------------------------------------------------
indirect     |
     college |  -.0195459    .010691    -1.83   0.068       -.0405    .0014081
-------------+----------------------------------------------------------------
total        |
     college |  -.1140705   .0171995    -6.63   0.000    -.1477808   -.0803602
------------------------------------------------------------------------------
```

- この表は、大学卒業率が 1 パーセントポイント増加した場合の平均変化を示しています。微分した値が報告されていますが、結果を 1 単位の変更として解釈することができます。

直接効果（**direct**）	波及効果を無視した郡内の変化の効果。郡内の直接効果により失業率は 0.09 パーセントポイント減少します。
間接効果（**indirect**）	波及効果。大学卒業率が 1 パーセントポイント増加すると失業率が減少し、それが波及して失業率がさらに減少します。結果として失業率は 0.02 パーセントポイント減少します。
総効果（**total**）	直接効果と間接効果の合計。-0.09 + (-0.02) = -0.11

- 効果の解釈には、必ず **estat impact** を使用して直接効果、間接効果、および総効果を求めます。空間自己回帰の結果の係数をそのまま解釈することはできません。この検証のために、失業率を 100 倍した結果を見てみましょう。

Summary of **spregress** results
Regression of **unemployment** and **100*unemployment**
on **college** and **W*unemployment**

	unemployment	**100*unemployment**
college	-0.094	-9.4
W*unemployment	0.201	0.201

Notes: Column 1 from **spregress** output above.
Column 2 from:

```
generate ue100 = 100*unemployment
spregress ue100 college, gs2sls dvarlag(W)
```


LightStone® 株式会社ライトストーン

- 失業率を 100 倍（**100*unemployment**）した結果、**college** の係数も 100 倍されています。一方$\mathrm{Wy}_{\mathrm{ue}}$の係数$\beta_1$には変化がありません。
- 失業率を 100 倍したモデルでは間接効果と直接効果の比率が小さくなると思われるかもしれません。しかし **estat impact** を実行すると、元のモデルでの **estat impact** の結果をすべて 100 倍したものが報告され、間接効果と直接効果の比率は変化しないことがわかります。

```
. estat impact

progress   :100%

Average impacts                                 Number of obs     =        254
```

	dy/dx	Delta-Method Std. Err.	z	P>|z|	[95% Conf.	Interval]
direct						
college	-9.452455	1.30576	-7.24	0.000	-12.0117	-6.893213
indirect						
college	-1.954593	1.069105	-1.83	0.068	-4.05	.1408134
total						
college	-11.40705	1.719946	-6.63	0.000	-14.77808	-8.036016

LightStone® 株式会社ライトストーン

説明変数の空間ラグでモデルをフィットする

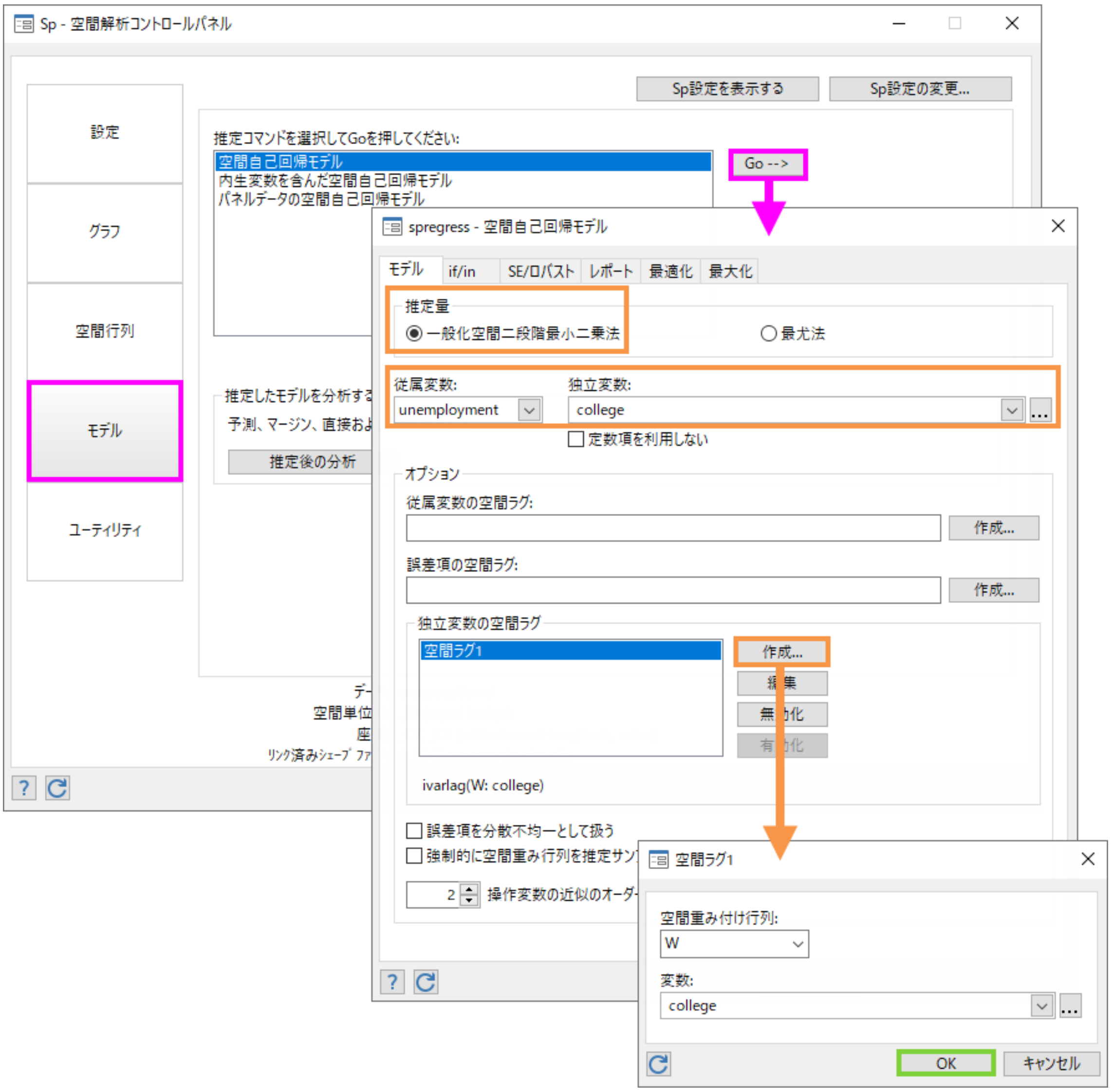
- これまで、被説明変数の空間ラグを使用したモデルについて解説しました。

$$\mathrm{y_{ue}} = \beta_0 + \beta_1 \mathrm{X_{cr}} + \beta_2 \mathrm{Wy_{ue}} + \epsilon$$

説明変数の空間ラグを使用してモデルをフィットすることも可能です。

$$\mathrm{y_{ue}} = \beta_0 + \beta_1 \mathrm{X_{cr}} + \beta_2 \mathrm{WX_{cr}} + \epsilon$$

- フィットを実行します。「空間解析コントロールパネル」で次のように設定します。

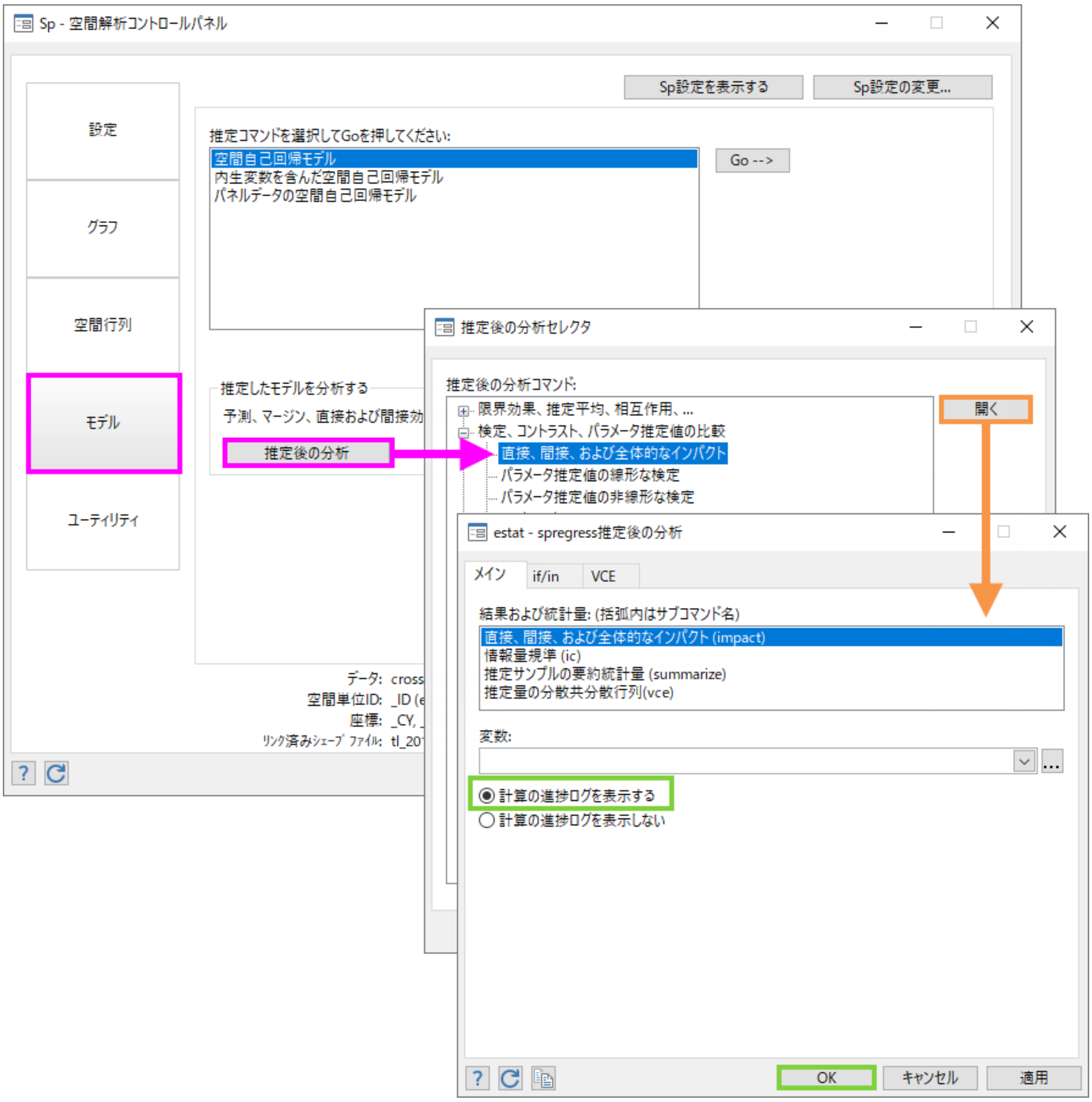


```
  (254 observat ions)
  (254 observations (places) used)
  (weighting matrix defines 254 places)

Spatial autoregressive model                     Number of obs     =        254
GS2SLS estimates                                 Wald chi2(2)      =      81.13
                                                 Prob > chi2       =     0.0000
                                                 Pseudo R2         =     0.2421

------------------------------------------------------------------------------
unemployment |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
unemployment |
     college |   -.077997   .0138127    -5.65   0.000    -.1050695   -.0509245
       _cons |   7.424453   .3212299    23.11   0.000     6.794854    8.054053
-------------+----------------------------------------------------------------
W            |
     college |  -.0823959   .0191586    -4.30   0.000    -.1199461   -.0448458
------------------------------------------------------------------------------
Wald test of spatial terms:          chi2(1) = 18.50      Prob > chi2 = 0.0000
```


LightStone® 株式会社ライトストーン

説明変数の空間ラグを使用したモデルの解釈

- 被説明変数の空間ラグを使用したモデルと同様に、`estat impact` を使用して説明変数の直接効果・間接効果をみることができます。「空間解析コントロールパネル」で次のように設定します。



```
progress    :100%

Average impacts                              Number of obs     =        254

------------------------------------------------------------------------------
             |            Delta-Method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
direct       |
     college |   -.077997   .0138127    -5.65   0.000    -.1050695   -.0509245
-------------+----------------------------------------------------------------
indirect     |
     college |  -.0715273   .0166314    -4.30   0.000    -.1041243   -.0389303
-------------+----------------------------------------------------------------
total        |
     college |  -.1495243   .0170417    -8.77   0.000    -.1829255   -.1161231
------------------------------------------------------------------------------
```

- 上記の表は、大学卒業率が 1 パーセントポイント増加した場合の平均変化を示しています。

直接効果（direct）	波及効果を無視した郡内の変化の効果。郡内の直接効果により失業率は 0.078 パーセントポイント減少します。
間接効果（indirect）	郡全体の波及効果。失業率は 0.072 パーセントポイント減少します。
総効果（total）	直接効果と間接効果の合計。-0.078 + (-0.072) = -0.15

- 詳細
  - 大学卒業率の直接効果は$\beta_1 X_{cr}$
  - 大学卒業率の間接効果は$\beta_2 WX_{cr}$
  - すべての郡でX_{cr}が 1 増加したときの直接効果は、

$$\Delta y_{ue} = \beta_1(X_{cr} + 1) - \beta_1 X_{cr} = \beta_1 1$$

ここで、1 は 1 の N×1 ベクトル

  - 直接効果は、各郡でy_{ue}がβ_1ずつ増加すること
  - 間接効果も同様にして、

$$\Delta y_{ue} = \beta_2 W(X_{cr} + 1) - \beta_2 WX_{cr} = \beta_2 W1$$

- この結果は、郡 i でy_{ue}が $(\beta_2 W1)i$ だけ増加することを示しています。
- 各郡がそれぞれ隣接する群の影響を受けるため、郡ごとに異なる影響があります。
- 郡全体の平均効果は$\beta_2 W1$の平均です。

空間自己回帰誤差でモデルをフィットする

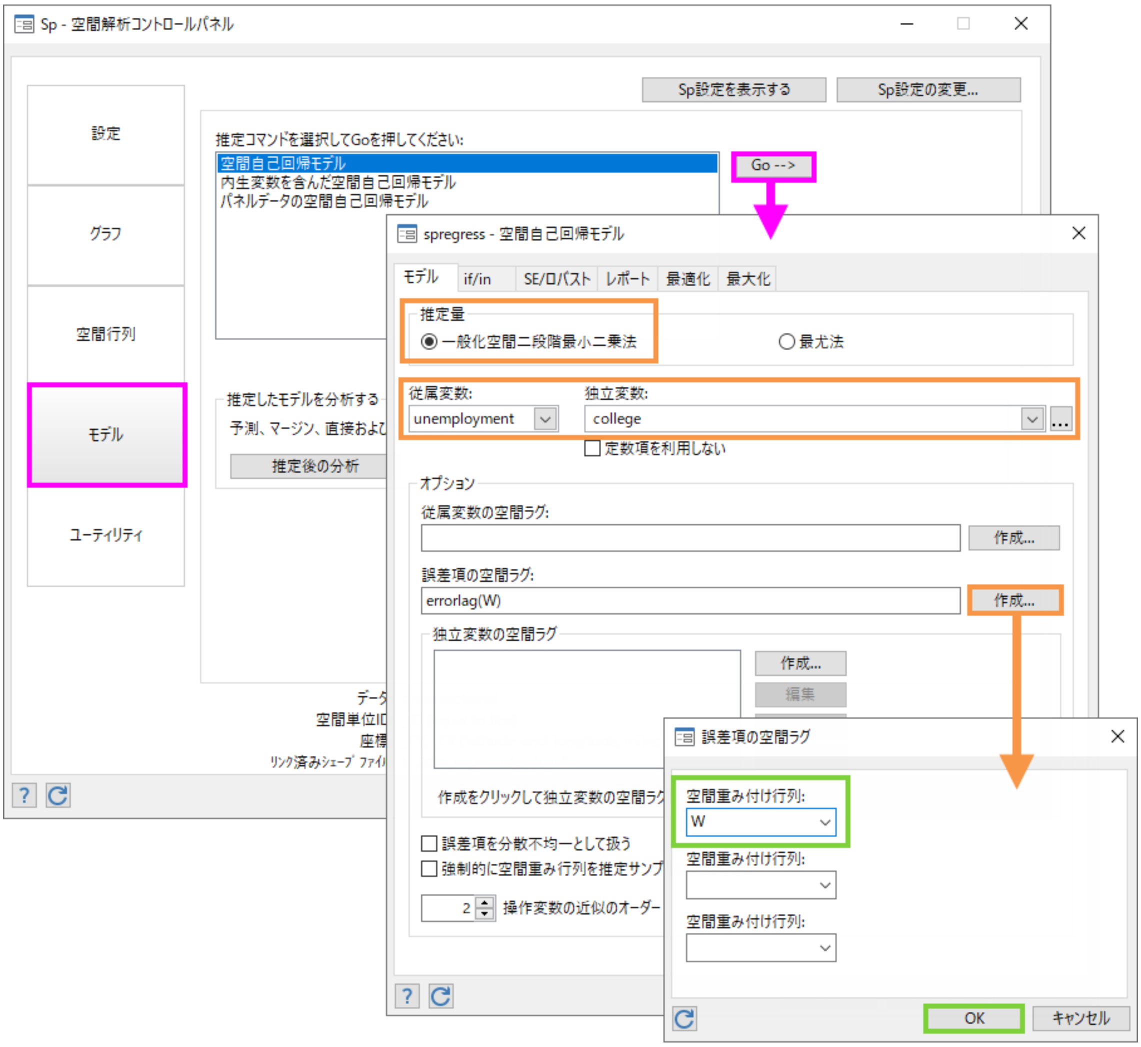
- これまで、被説明変数の空間ラグを使用したモデルと説明変数の空間ラグを使用したモデルについて解説しました。

$$y_{ue} = \beta_0 + \beta_1 X_{cr} + \beta_2 W y_{ue} + \epsilon$$
$$y_{ue} = \beta_0 + \beta_1 X_{cr} + \beta_2 W X_{cr} + \epsilon$$

誤差項の空間ラグを使用してモデルをフィットすることも可能です。

$$y_{ue} = \beta_0 + \beta_1 X_{cr} + (I - \rho W)^{-1} \epsilon$$

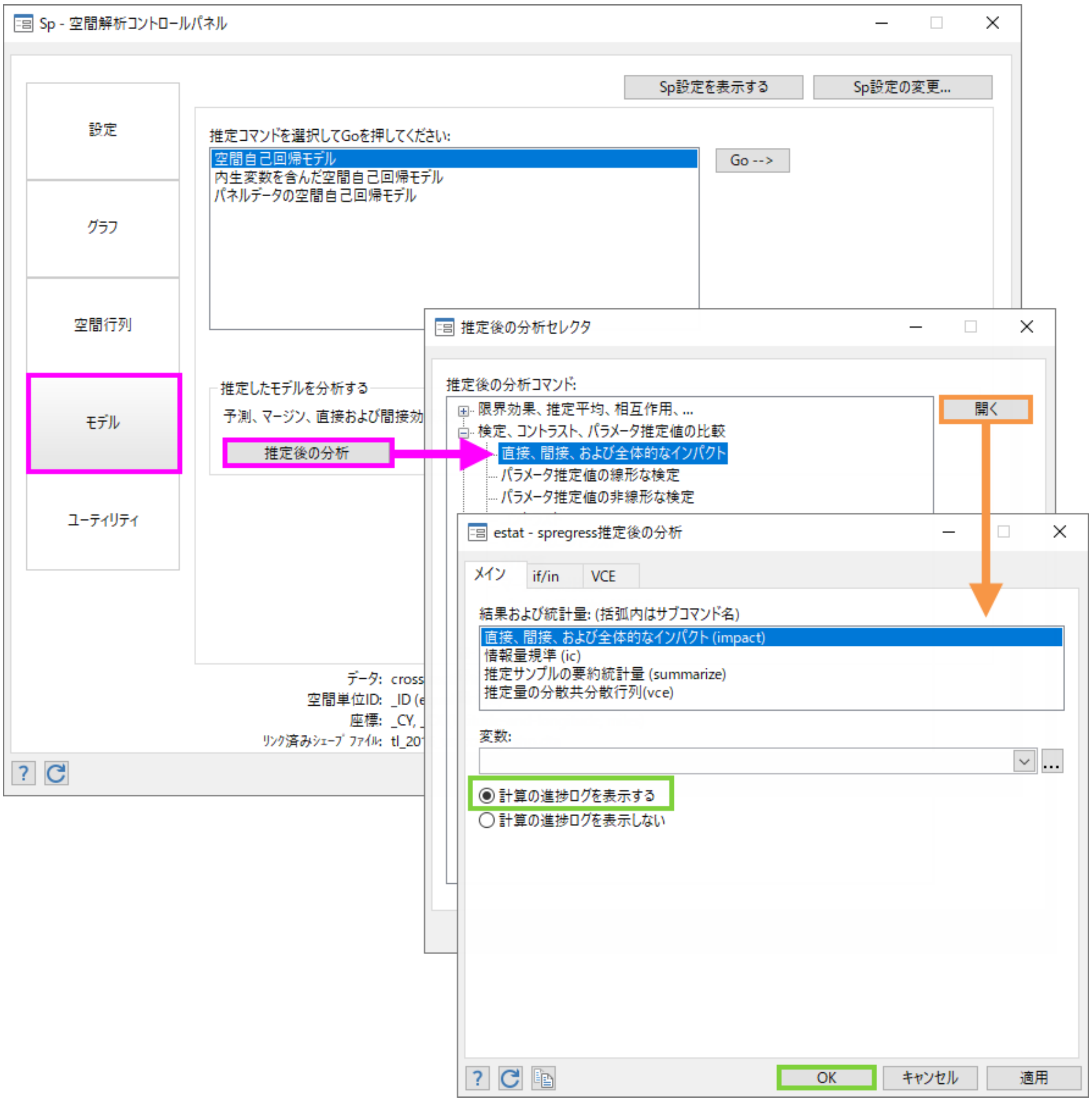
- フィットを実行します。「空間解析コントロールパネル」で次のように設定します。



```
  (254 observations)
  (254 observations (places) used)
  (weighting matrix defines 254 places)

Estimating rho using 2SLS residuals:

initial:       GMM criterion =  .71251706
alternative:   GMM criterion =  .04381608
rescale:       GMM criterion =  .02453154
Iteration 0:   GMM criterion =  .02453154
Iteration 1:   GMM criterion =  .00420723
Iteration 2:   GMM criterion =   .0002217
Iteration 3:   GMM criterion =  .00021298
Iteration 4:   GMM criterion =  .00021298

Estimating rho using GS2SLS residuals:

Iteration 0:   GMM criterion =  .00566696
Iteration 1:   GMM criterion =  .00486118
Iteration 2:   GMM criterion =  .00486066
Iteration 3:   GMM criterion =  .00486066

Spatial autoregressive model                     Number of obs     =        254
GS2SLS estimates                                 Wald chi2(1)      =      37.76
                                                 Prob > chi2       =     0.0000
                                                 Pseudo R2         =     0.1869

------------------------------------------------------------------------------
unemployment |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
unemployment |
     college |  -.0759125   .0123532    -6.15   0.000    -.1001243   -.0517008
       _cons |   6.292997   .2968272    21.20   0.000     5.711227    6.874768
-------------+----------------------------------------------------------------
W            |
e.unemploym~t|   .7697395   .0690499    11.15   0.000     .6344043    .9050748
------------------------------------------------------------------------------
Wald test of spatial terms:          chi2(1) = 124.27     Prob > chi2 = 0.0000
```

- 空間自己相関パラメータρの推定値は、表の一番下の行に示されており、大きくて有意であると推定されています。$\hat{\rho} = 0.77$
- ρは自己相関パラメータと呼ばれ、相関係数といくつかの特性を共有しています。自己相関パラメータの理論上の範囲は［-1, 1］で、$\rho = 0$は自己相関が 0 であることを意味します。

● `estat impact` ではρは報告されません。「空間解析コントロールパネル」で次のように設定して、結果を確認します。



```
progress   :100%

Average impacts                              Number of obs     =        254

-----------------------------------------------------------------------------
             |              Delta-Method
             |      dy/dx    Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+---------------------------------------------------------------
direct       |
     college |  -.0759125    .0123532    -6.15   0.000    -.1001243   -.0517008
-------------+---------------------------------------------------------------
indirect     |
     college |          0  (omitted)
-------------+---------------------------------------------------------------
total        |
     college |  -.0759125    .0123532    -6.15   0.000    -.1001243   -.0517008
-----------------------------------------------------------------------------
```

- 被説明変数または説明変数のラグがない場合に **estat impact** を実行すると、このような結果になります。
- 波及効果はありません。空間相関誤差は、共変量に波及効果を及ぼしません。

3 種類すべての空間ラグ項を含めることが可能

- 1 つのモデルに複数種類の空間ラグ項を含めることも可能です。
- **estat impact** を使用して、被説明変数のラグ、説明変数のラグ、あるいはその両方の共変量の効果を推定します。
- 空間相関誤差項を含める場合は、推定されたρの大きさと有意性を確認します。