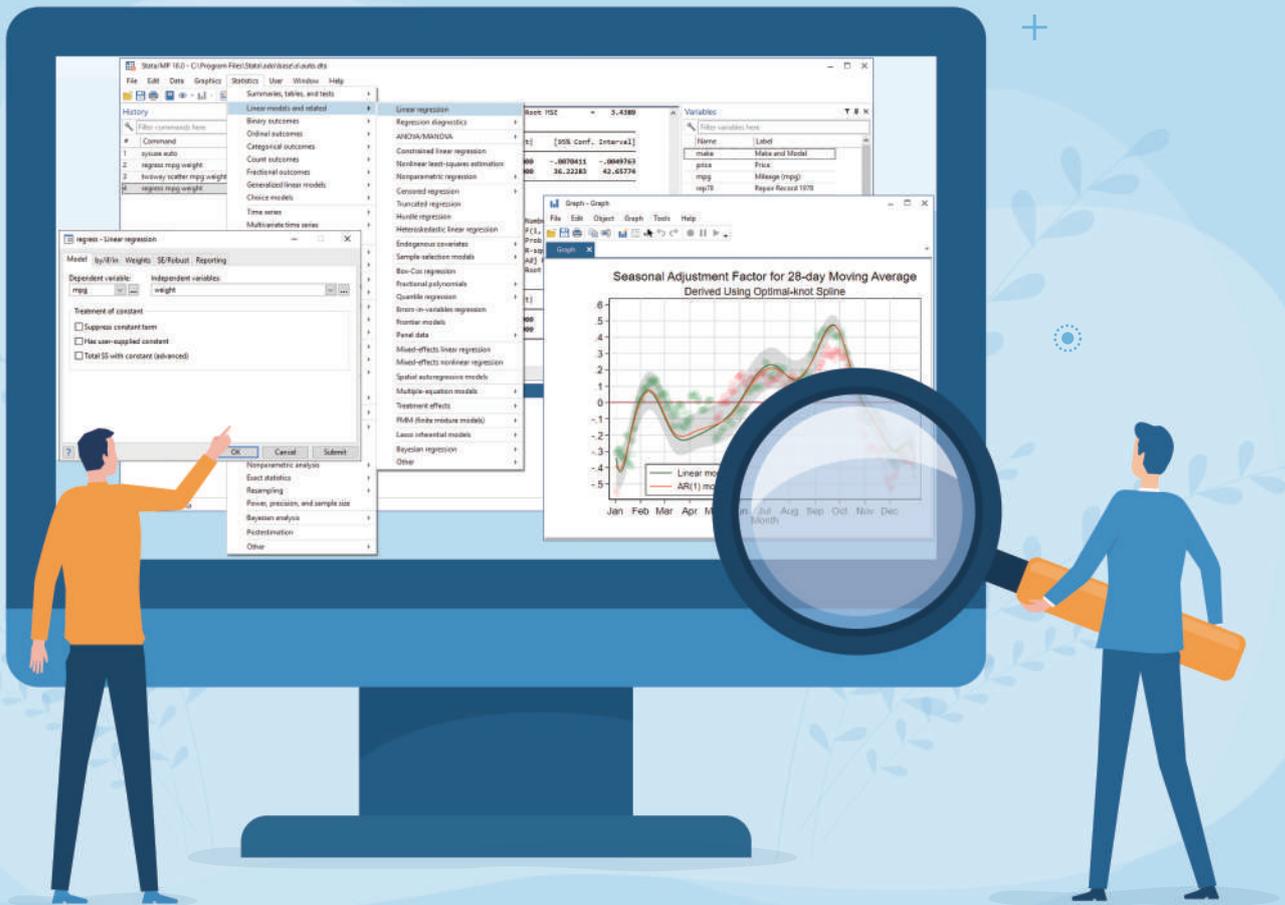


STATA®

活用ガイドブック



Stata 活用ガイドブック

株式会社ライトストーン

2023年12月

目次

第1章 基本操作とデータ入力.....	1
1.1 Stata のインターフェース.....	1
1.1.1 ログ機能.....	2
1.1.2 表示言語の切り替え.....	3
1.1.3 ウィンドウのフォントとサイズの変更.....	3
1.2 データ入力.....	4
1.2.1 データエディタ.....	4
1.2.2 変数名の設定.....	5
1.2.3 Excel・テキストファイルのインポート.....	6
1.2.4 データの保存.....	9
1.3 文字列変数と数値変数.....	12
1.3.1 変数の属性.....	12
1.3.2 Unicode.....	14
1.4 変数のラベル.....	14
1.4.1 変数ラベル.....	14
1.4.2 ラベルの確認.....	17
1.4.3 文字データを数値データに変換する.....	18
第2章 グラフの作成.....	20
2.1 ヒストグラム.....	20
2.1.1 分布曲線を追加する.....	22
2.1.2 カテゴリごとにヒストグラムを作成する.....	23
2.2 散布図.....	24
2.2.1 散布図をカテゴリごとに作成.....	26
2.3 折れ線グラフ.....	28
2.4 種類の異なるグラフを一つに統合する.....	29
2.5 グラフの共通設定.....	31
2.5.1 y軸・x軸オプション.....	32
2.5.2 グラフタイトルのオプション.....	32
2.5.3 凡例オプション.....	33
2.5.4 グラフの全般オプション.....	33
2.6 異なるグラフを並べて表示する.....	33
2.7 グラフエディタ.....	34
2.8 グラフの保存.....	35
第3章 基本的な統計操作.....	37

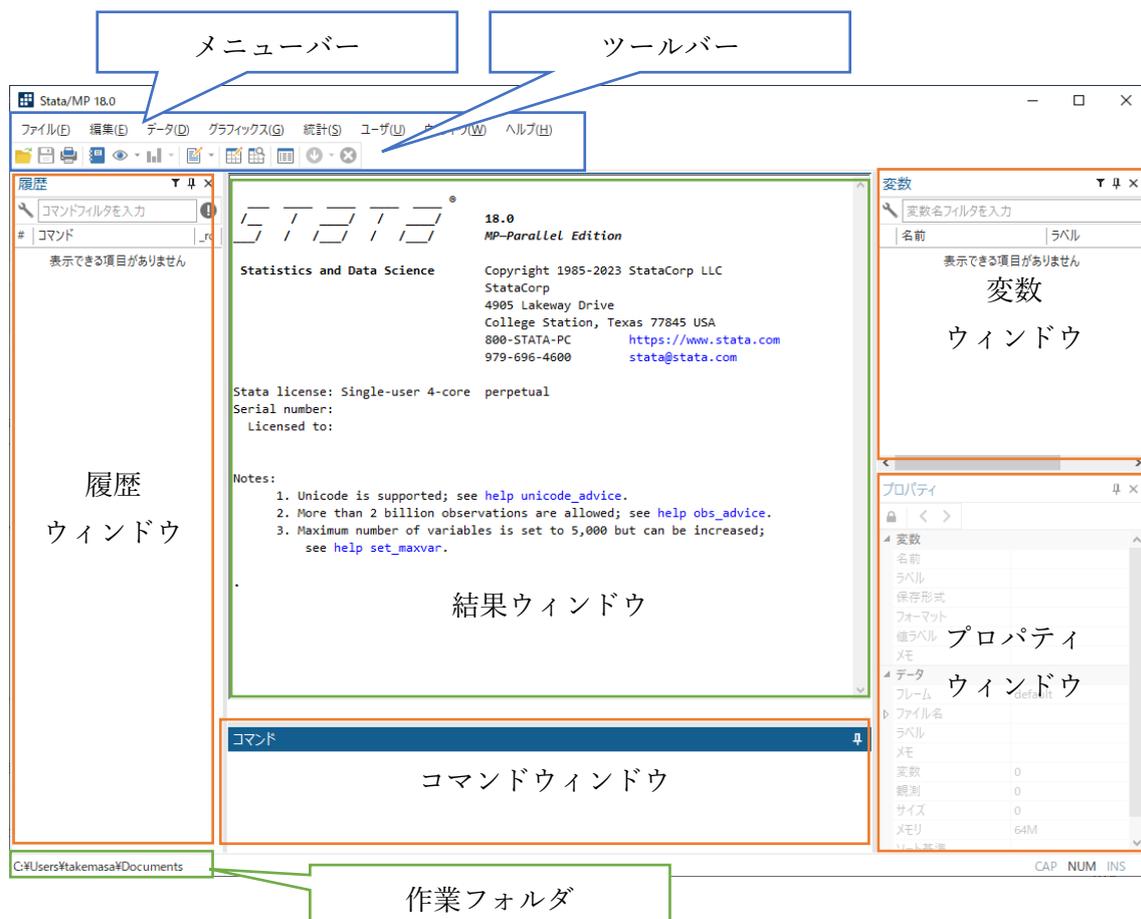
3.1	要約統計量	37
3.2.1	正規分布	38
3.3	度数分布表	39
3.3.1	一元配置表	39
3.3.2	二元配置（クロス集計）表	41
3.4	仮説検定	44
3.4.1	Stata のデータ構成と変形	44
3.4.2	仮説検定の実行	48
3.5	PDF マニュアルの活用	51
第 4 章	線形回帰分析	54
4.1	線形回帰モデル	54
4.2	因子変数と交互作用項	56
4.2.1	Stata 独自の因子変数	56
4.2.2	交互作用	57
4.3	限界効果	58
4.3.1	因子変数と交差項の限界効果	58
4.3.2	限界効果のグラフ化	59
4.4	推定後の診断	60
4.4.1	F 検定	60
4.4.2	多重共線性	61
4.4.3	標準的仮定	62
4.5	予測	63
4.5.1	フィット値	63
4.5.2	残差の計算	64
4.6	推定結果の表	65
第 5 章	Do ファイルと ado ファイル	69
5.1	Do ファイルの作成	69
5.1.1	コメントの付与	71
5.1.2	演算子	72
5.1.3	ブックマークとナビゲーションコントロール	72
5.2	Do ファイルの実行	73
5.3	Do ファイルの運用方法	73
5.4	ado の探し方と試し方	74
5.5	ado ファイルの検索とインストール	74
5.5.1	ado ファイルの実行	77
5.6	ado ファイルの更新	78

第1章 基本操作とデータ入力

本章では Stata の操作に関する基本的な事項について説明します。ここでは、解析するデータの準備や統計解析を取り巻く機能を中心にその操作方法を説明します。

1.1 Stata のインターフェース

まずは、Stata の操作画面について説明します。



Stata のウィンドウは次の5つで構成されています。

- 履歴ウィンドウ：実行したコマンドの履歴が表示されます。
- 結果ウィンドウ：実行したコマンドの結果が表示されます。
- 変数ウィンドウ：使用できる変数の一覧が表示されます。
- プロパティウィンドウ：変数で選択した変数の詳細が表示されます。
- コマンドウィンドウ：コマンドを入力するウィンドウです。

さらに、コマンドウィンドウの左下には、現在の作業フォルダの設定が表示されます。作業

第1章 基本操作とデータ入力

フォルダは解析に使用するデータや、Stata から出力したファイルが保存される場所です。作業フォルダを変更するにはコマンドウィンドウで、下記のように、`cd` コマンドの後にフォルダパスを指定するか、または、メインメニューで**ファイル > 作業フォルダの変更**、と操作してフォルダを指定します。

```
.cd c:¥data
```

1.1.1 ログ機能

Stata の操作結果は結果ウィンドウと履歴ウィンドウに表示されますが、Stata を終了させると消えてしまいます。Stata のセッション内容を記録するには、ログ機能を利用します。

1.1.1.1 ログの開始

メニューの**ファイル > ログ > 開始**、と操作してログファイルの出力先とファイル名を指定して、記録を開始します。通常、ログファイルは Stata 独自の `.smcl` 形式で作成されます。コマンド操作では、次のように行います。

```
. log using "c:¥data¥log_sample.smcl", replace
```

1.1.1.2 ログの終了

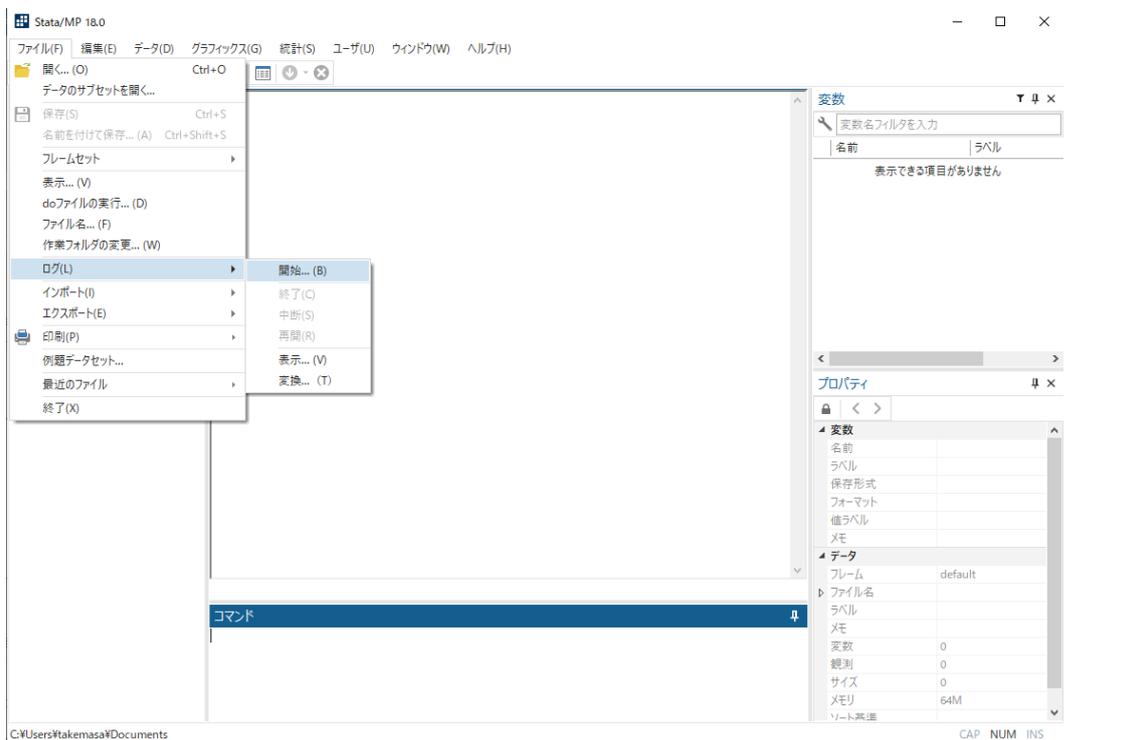
ログを終了するには、**ファイル > ログ > 終了**、と操作します。コマンドは次の通りです。

```
. log close
```

1.1.1.3 ログの確認

ログに記録した内容を確認するには、**ファイル > ログ > 表示**、を選択します。保存されているログファイルを指定し、OK をクリックすると、内容が表示されます。

1.1 Stata のインターフェース



1.1.2 表示言語の切り替え

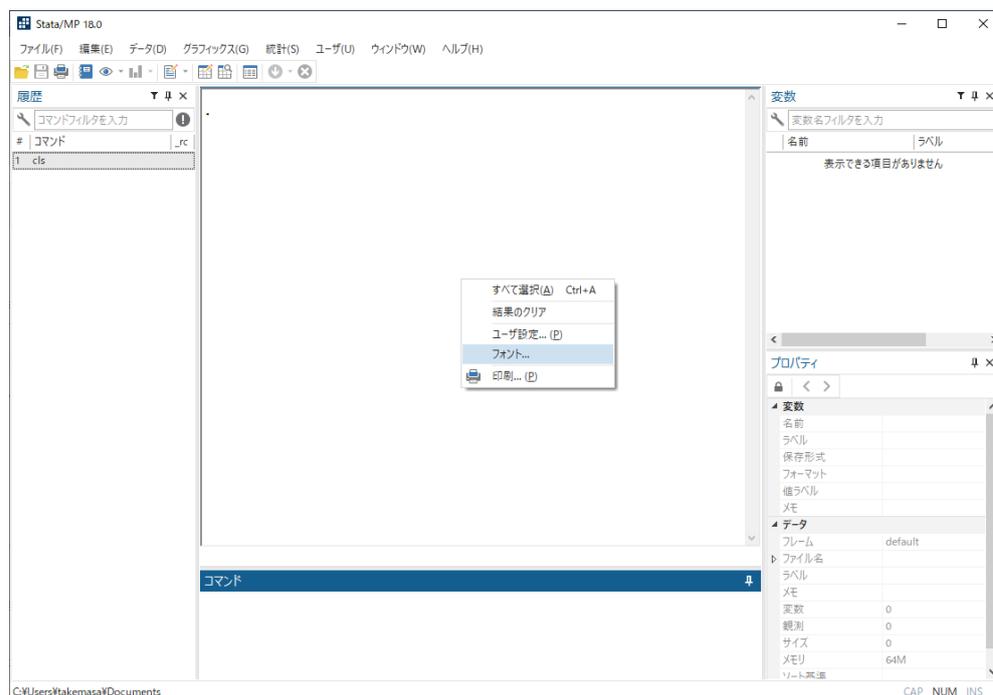
Stata14 以降では、メニュー表示を 5 言語¹から選択することができます。切り替えは、メニューの **編集 > ユーザ設定 > ユーザインターフェース言語**、を選んで、言語を選択します。選択後、Stata を再起動して、設定を適用させます。

1.1.3 ウィンドウのフォントとサイズの変更

コマンドウィンドウと結果ウィンドウに表示されるフォントとフォントサイズは、それぞれのウィンドウを右クリックして、ショートカットメニューの **フォント**、を選択することで変更できます。

¹ Stata16 では 6 言語から選択できます。デフォルト設定は、OS の言語に従います。

第1章 基本操作とデータ入力



1.2 データ入力

1.2.1 データエディタ

Stata が使用するデータセットはデータエディタに書き込まれています。データエディタはツールバーのデータエディタボタンで表示します。データエディタには編集モードとブラウザモードがあります。編集モードでは、スプレッドシートを直接編集して、データを入力したり、書き換えたりすることができます。ブラウザモードでは、スプレッドシートを閲覧するのみで、編集はできません。

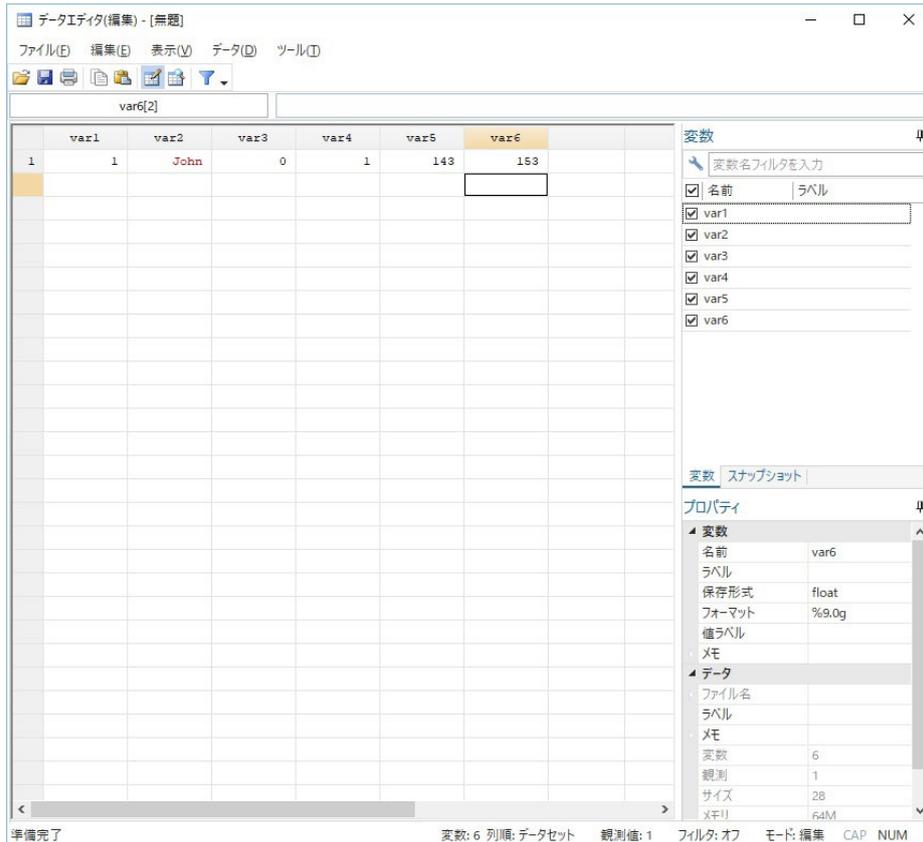


データを直接入力するには、「編集モード」ボタンをクリックして、データエディタを起動します。何らかのデータが既にロードされている場合、コマンドウィンドウに `clear` と入力して、メモリ上のデータを削除します。データエディタアイコンをクリックして、データエディタを起動します。データエディタのセルをクリックして、次の表のようにデータを直接入力してみましょう。

1.2 データ入力

1	John	0	1	143	153
2	Mary	0	2	157	127
3	Ken	0	3	147	176

データの入力を進めると、データエディタでは次のように表示されます。²

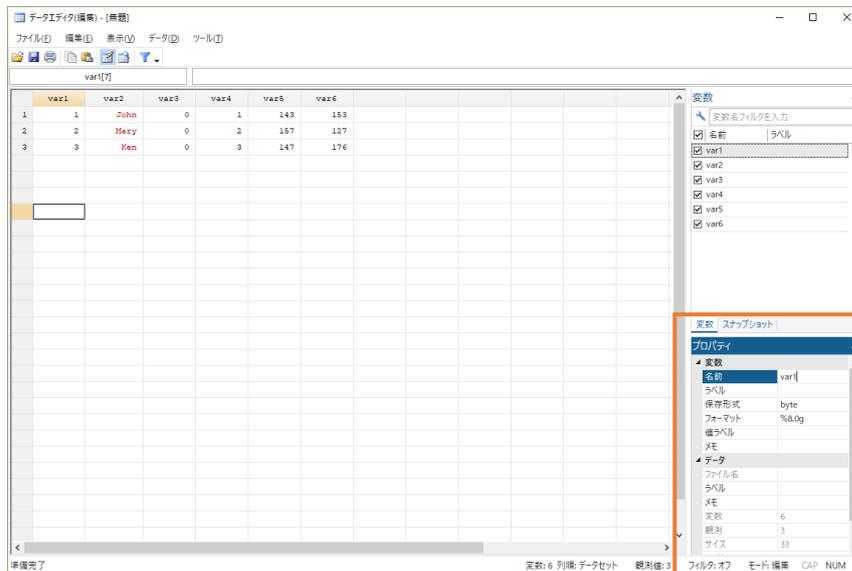


1.2.2 変数名の設定

この時点では変数名が `var1` から `var6` となっていますので、変数名を変更します。データエディタの変数ウィンドウで目的の変数を選択し、プロパティウィンドウで名前フィールドを変更します。変数名は最大 32 文字まで入力できます。

² データエディタ上では、文字列データは赤色で表示されます。

第1章 基本操作とデータ入力



1.2.3 Excel・テキストファイルのインポート

メニュー操作を用いて、Stata のインポート機能とコピー&ペーストでのファイルのインポートを行います。サンプルの Excel ファイルは、下記の URL からダウンロードできます。
<https://www.lightstone.co.jp/stata/tutorial.html>

1.2.3.1 メニュー操作での Excel ファイルのインポート

1. メニューバーで **ファイル > インポート > Excel シート形式(*.xls, *.xlsx)**、と操作します。
2. 「Excel ファイルのインポート」ダイアログ上で **sampledata.xls** を選択します。
3. Excel ファイルの内容に合わせてオプションを設定します。
 - シート: インポート元のシート名を指定します。
 - セル範囲: インポートするデータの範囲を指定します。(通常は A1 セルから始まります)
 - 第一行を変数名としてインポートする: Excel ファイルの一行目に変数名が入力されている場合、Stata はそれらを変数名としてインポートします。

設定が完了すると、次のような画面になります。プレビューウィンドウでインポートした時のイメージが確認できます。

1.2 データ入力

Excelのインポート - Excelファイルのインポート

Excelファイル:
C:\Stata\data\sampldata.xlsx

シート: Sheet1 A1:G97

セル範囲: A1:G97

第1行を変数名としてインポートする

全データを文字列としてインポートする

変数の大/小文字: 変更しない

プレビュー: (全97行のうち2-51行を表示する)

	city	yearmonth	rainfall	summerday	temp_mean	snow	fourseasons
2	新潟	2006/06	161	0	3	1	4
3	東京	2006/06	26	0	5.8	0	4
4	根室	2006/06	28.5	0	-4.2	1	4
5	那覇	2006/06	92.5	1	18.4	0	4
6	新潟	2006/07	206	0	1.7	1	4
7	東京	2006/07	48.5	0	4.7	1	4
8	根室	2006/07	16		-2.5	1	4

OK キャンセル

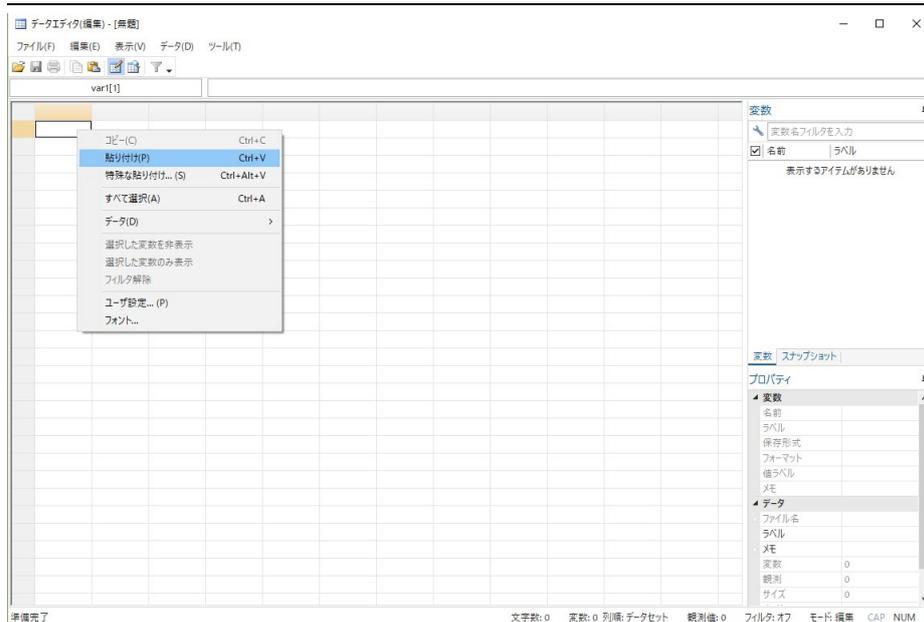
1.2.3.2 コピー&ペーストでの Excel ファイルのインポート

Excel ファイル上でセルをコピー&ペーストで Stata に入力できます。

1. Excel でファイルを開きます。
2. Excel 上で、データの入力されている範囲を選択して、コピーします。
3. Stata のデータエディタを開き、var1[1]のセルが選択された状態で、貼り付けます。

この際に、「クリップボードの 1 行目には、有効な変数名の値が含まれます」と表示されるので、「変数名」を選択します。これで、コピーした範囲の 1 行目は変数名となります(上記メニュー操作の「第 1 行を変数名としてインポートする」と同様に機能します)。

第1章 基本操作とデータ入力



1.2.3.3 テキストファイルのインポート

テキストファイルや csv ファイルをインポートするには次のように操作します。

1. メニューバーで**ファイル > インポート > テキストデータ(デリミタ、csv 等)**、と操作します。
2. 「デリミタテキストデータをインポートする」ダイアログ上で csv ファイル、またはテキストファイルを選択します。
3. ファイルの内容に合わせてオプションを設定します。
 - デリミタ: データの区切り文字を指定します。³
 - テキストのエンコード: ファイルのエンコードを指定します。

³ インポートダイアログで、デリミタの設定を変更することで、タブ区切りやスペース区切りのテキストファイルをインポートできます。また、区切り記号は、ユーザ独自に設定することも可能です。この場合は、「カスタム」を選択して、編集ボックスに区切り記号を入力します。

1.2 データ入力

import delimited - デリミタテキストデータをインポートする

インポートするファイル:
C:\Stata\data\sampladata.csv

第1行を実数名として使用する: 自動

変数の大文字/小文字: 小文字

浮動小数点の精度: 既定

テキストのエンコード: 日本語(Shift JIS)

二重引用符: データ内

二重引用符の削除: 自動

デリミタ: 自動

連続するデリミタを1つとする

数値抽出のルール
 数値にのみ
ローカル: E

プレビュー:

#	city	yearmonth	rainfall	summerday	temp_mean	snow	fourseasons
2	新潟	2006/06	161	0	3.0	1	4
3	東京	2006/06	26	0	5.8	0	4
4	根室	2006/06	29	0	-4.2	1	4
5	那覇	2006/06	93	1	18.4	0	4
6	新潟	2006/07	206	0	1.7	1	4
7	東京	2006/07	49	0	4.7	1	4
8	根室	2006/07	16	0	-2.5	1	4
9	那覇	2006/07	151	0	17.2	0	4
10	新潟	2006/08	95	0	3.3	1	4
11	東京	2006/08	16	0	6.9	0	4
12	根室	2006/08	22	0	-3.1	1	4
13	那覇	2006/08	84	0	17.1	0	4
14	新潟	2006/09	109	0	1.4	1	4
15	東京	2006/09	20	0	5.4	1	4
16	根室	2006/09	32	0	-4.1	1	4
17	那覇	2006/09	84	0	16.9	0	4
18	新潟	2006/10	69	0	6.1	1	1
19	東京	2006/10	86	0	8.5	0	1
20	根室	2006/10	25	0	-1.4	1	1

列のデータ型を変更するには、選択した列を右クリックして適切な型を選択します。
注意: 文字列データを数値としてインポートすると、データが失われる可能性があります。

? [refresh] [help]

1.2.4 データの保存

データエディタ上のデータセットは Stata 形式または、Excel 形式、csv 形式などで保存することができます。

1.2.4.1 Stata 形式でデータファイルを保存

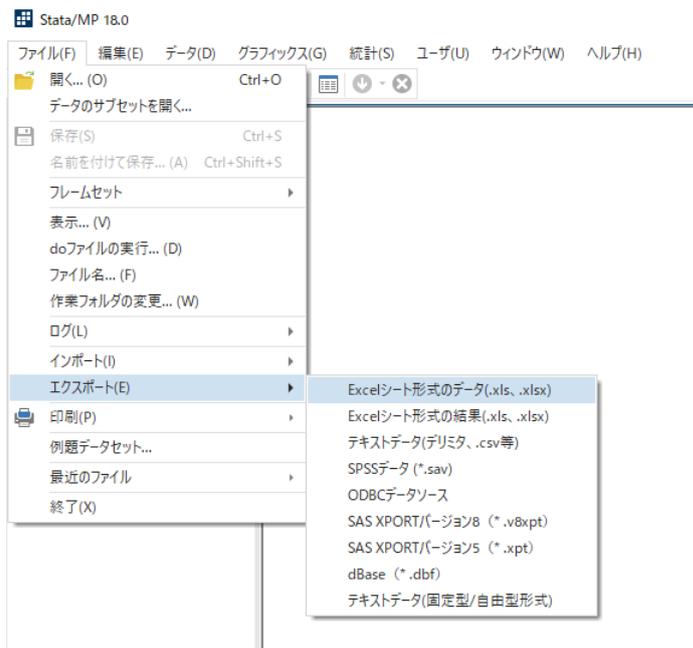
ツールバーの保存ボタンをクリックまたは、ファイル > 名前を付けて保存、と操作して、データエディタにあるデータセットを Stata 形式(*.dta)で保存します。⁴

1.2.4.2 その他の形式でファイルを保存

dta 形式以外で保存する場合は、ファイル > エクスポート、と選択して、保存形式を決定します。

⁴ Stata14 以降で作成されたデータファイルは、Stata13 および 12/11 では読み込むことができません。保存ダイアログで、“ファイルの種類”ドロップダウンリストで変更します。

第1章 基本操作とデータ入力



a. データセットをエクセルファイルとして保存

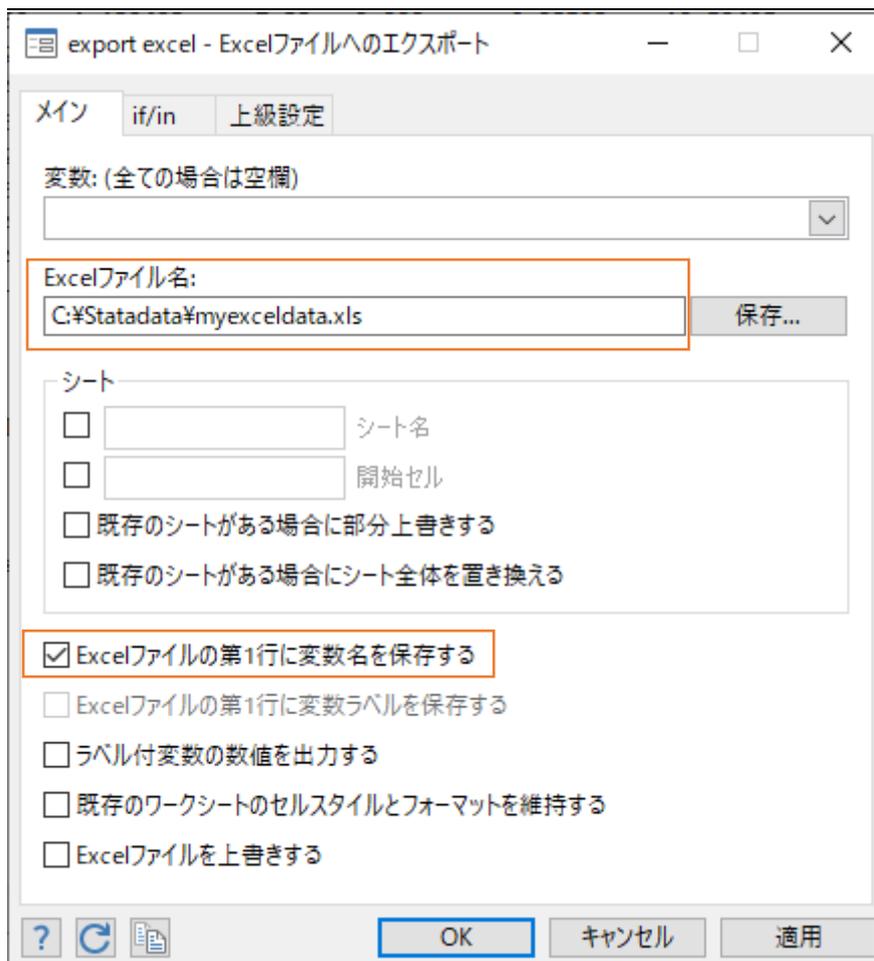
「Excel シート形式のデータ」を選択すると、「Excel ファイルへのエクスポート」ダイアログが開きます。

1. 「変数」ドロップダウンリストで、保存する変数を選択します。
2. 「Excel ファイル名」編集ボックスで、ファイル名を指定します。現在の作業フォルダ以外に保存する場合は、隣の「保存」ボタンをクリックして、保存場所を指定します。
3. 保存オプションを設定します。
 - 「Excel ファイルの第一行に変数名を保存する」: Excel ファイルの最初の行に変数名を記入します。
 - 「Excel ファイルを上書きする」: 既存のファイルを上書きする場合は、このチェックを入れます。

または、`export excel` コマンドを使用します。ダブルクォーテーション(” ”)で保存するファイル名を指定します。コマンド操作では次のように入力します。

． `export excel "mydata.xls"`

1.2 データ入力



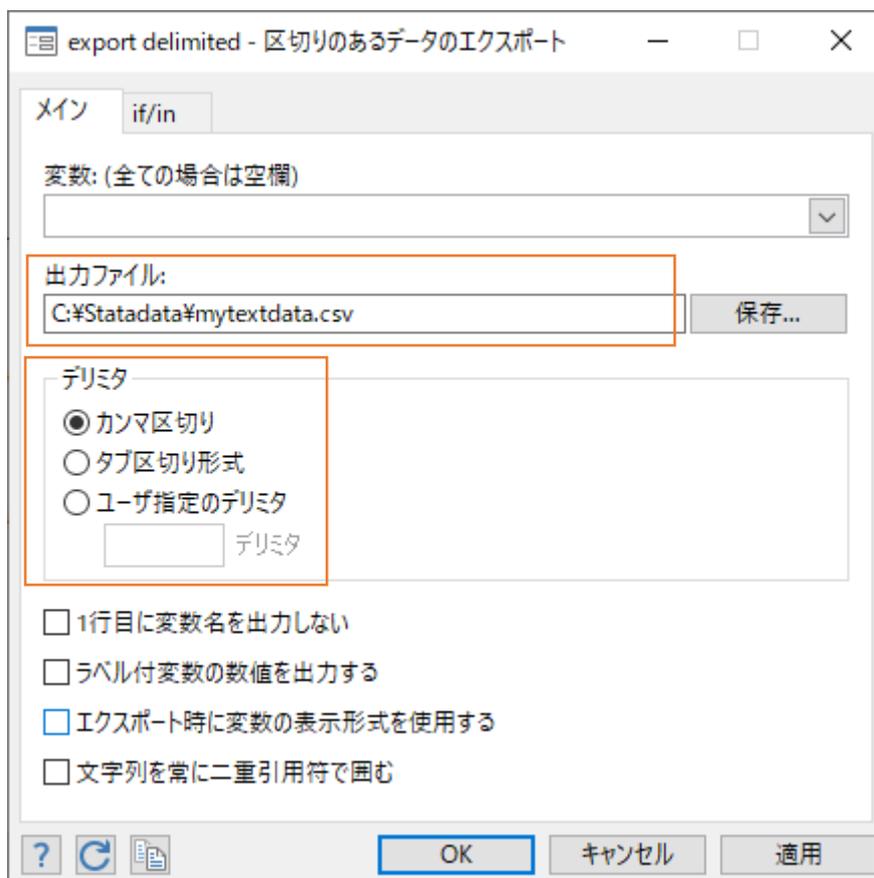
b. データセットを csv ファイルとして保存する

1. 「テキストデータ」を選択すると、ダイアログが開きます。
2. 「変数」ドロップダウンリストで、保存する変数を選択します。
3. 「出力ファイル」編集ボックスで、ファイル名を指定します。現在の作業フォルダ以外に保存する場合は、隣の「保存」ボタンをクリックして、保存場所を指定します。
4. 「デリミタ」ボックスで区切り記号を設定します。

コマンド操作では `export delimited using` を使用します。 `export excel` と同様にファイル名を指定します。

. `export delimited using "mydata.csv"`

第1章 基本操作とデータ入力



1.3 文字列変数と数値変数

1.3.1 変数の属性

Stata の変数は、まず、数値データと文字列データの2つに大別されます。数値データはさらに5種類の属性(データ型)を持ちます。これらは、プロパティウィンドウの「保存形式」の欄に表示されます。

数値データ

保存形式	最小値	最大値	1 観測値の使用 メモリ量(バイト)
byte	-127	100	1
int	-32,767	32740	2
long	-2,147,483,647	2,147,483,620	4
float	-1.70141173319 $\times 10^{38}$	1.70141173319 $\times 10^{38}$	4

1.3 文字列変数と数値変数

double	-8.9884656743 $\times 10^{307}$	8.9884656743 $\times 10^{307}$	8
--------	------------------------------------	-----------------------------------	---

byte、int、long は整数データを、float と double は浮動小数データを、それぞれ格納します。float タイプでは 7 桁の精度でデータを格納します。float は 1234567 という値を正確に保存できますが、7 桁を超える 123456789 という値は 123456792 と変形され、正確に保存できません。

つまり、7 桁以内の数値であれば float タイプ、9 桁以内の数値を正確に保存する場合は long タイプ、それ以上の桁数を扱う場合は、double タイプが適切です。

文字列データ

保存形式	最大文字数	1 観測値の使用メモリ量(バイト)
str1	1	1
str2	2	2
...
...
...
str2045	2,045	2,045
strL	2,000,000,000	2,000,000,000

文字列データを表す保存形式 str# は格納する文字数によって定義されます。例えば、str4 では male という単語を保存できますが、female は保存できません。

これらのデータセットの情報は、describe コマンドまたは、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を表示、を選択します。

第1章 基本操作とデータ入力

```
. describe

Contains data
  obs:      48
  vars:      5
  size:     1,008
```

variable name	storage type	display format	value label	variable label
month_year	str6	%9s		
city	str6	%9s		
rainfall	float	%9.0g		
freq_over25	byte	%8.0g		
citycode	long	%8.0g	citycode	

```
Sorted by:
  Note: Dataset has changed since last saved.
```

1.3.2 Unicode

Stata14 以降では、それ以前の Stata と異なり、拡張 ASCII 文字を UTF-8 でエンコードします。拡張 ASCII 文字を含むアスキーファイルをインポートする際には、適切なエンコード形式を選択する必要があります。Stata13 以前で作成したファイルは、14 以降では文字化けしてしまう可能性があります。UTF-8 でエンコードした新たなファイルに変換する必要があります。詳細は以下のライトストーンの web ページをご覧ください。

https://www.lightstone.co.jp/stata/files/Stata14_unicode.pdf

1.4 変数のラベル

データに付加できるラベルには、変数ラベル、値ラベル、データセットラベルの3種類があります。

1.4.1 変数ラベル

変数名だけではデータの内容が正確に認識できないこともあります。変数ラベルは、変数に関する情報を説明・補足します。

1.4.1.1 変数ラベルの設定

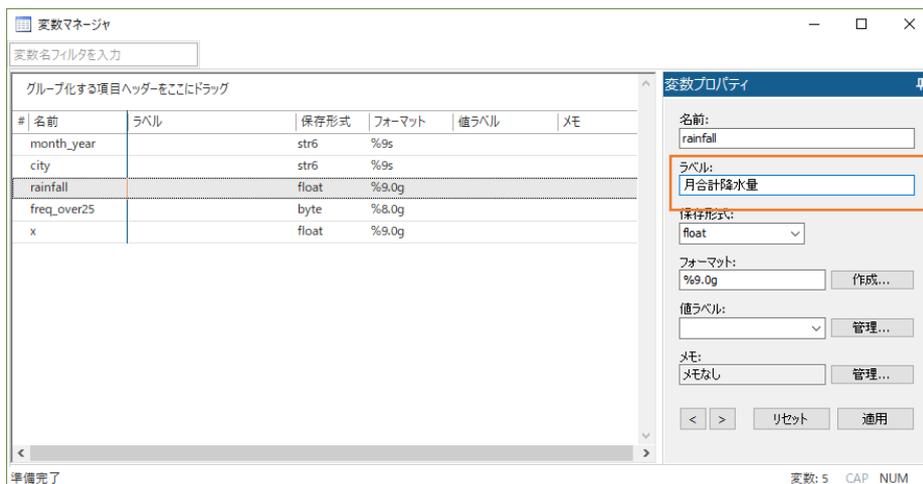
ツールバーで変数マネージャを開き、変数ラベルを設定します。変数マネージャの左側のテーブルで変数を選択、右側の変数プロパティで「ラベル」フィールドを編集します。

ここでは、次のようにラベルを入力して、変数の説明を追加します。

変数名	変数ラベル
city	観測地

1.4 変数のラベル

yearmonth	観測年月
rainfall	月合計降水量
summerday	夏日の観測数
temp_mean	月平均気温
snow	降雪の有無
fourseasons	四季

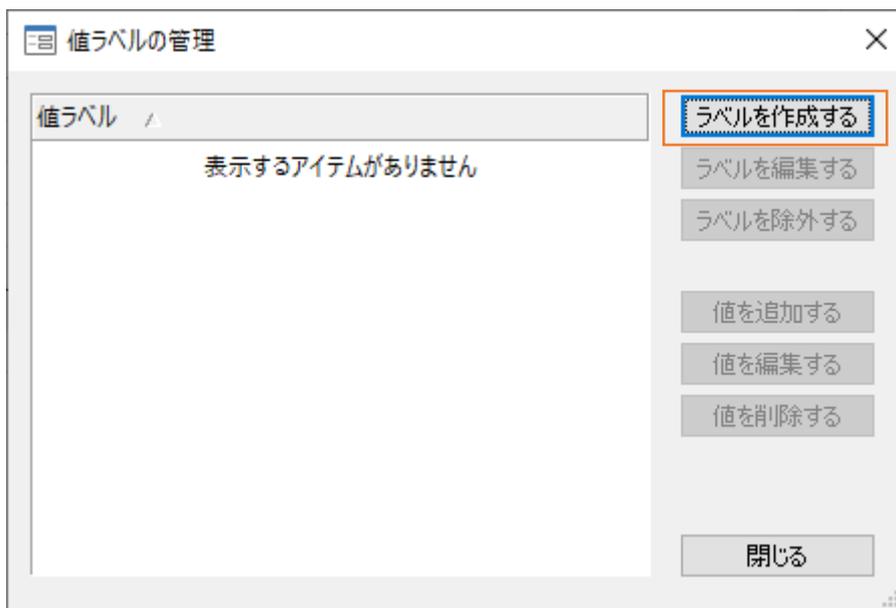


1.4.1.2 値ラベル

値ラベルは、カテゴリ変数の各値が何を意味しているか、わかりやすくするための対応表として機能します。Stataには数値コードを認識させつつ、表示上はラベル情報を表示させることができます。

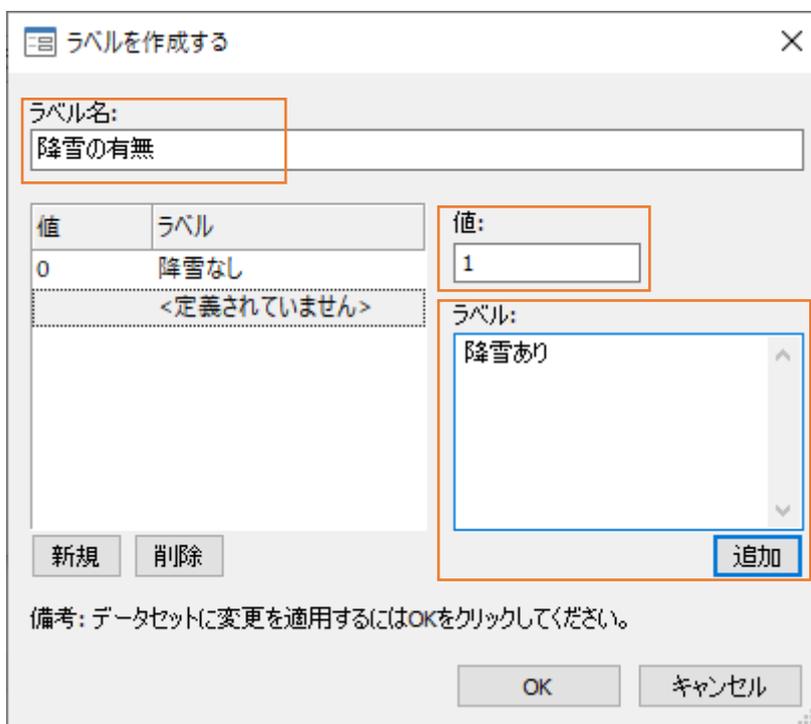
1. 変数マネージャでカテゴリ変数を選択、左側の変数プロパティウィンドウで「値ラベル」を選択します。
2. 「値ラベルの管理」ダイアログで、「ラベルを作成する」ボタンをクリックします。

第1章 基本操作とデータ入力



3. ラベルを作成するダイアログで、値ラベルの名前を設定し、コードに対応するラベルを設定し、追加ボタンをクリックします。
4. 値ラベルが設定できたら、変数マネージャに戻って、値ラベルのドロップダウンリストから、作成したラベルを選択します。

ここでは、変数 `snow` の値 0 を「降雪なし」、1 を「降雪あり」として対応させます。



1.4 変数のラベル

1.4.1.3 データセットラベル

データセットラベルは、データセット全体に対してラベルを付加します。データセットのメモとして機能します。

1. メニューバーで、**データ > データユーティリティ > ラベルユーティリティ > データセットのラベル**、と操作します。
2. 開いたダイアログで、ラベル情報を入力します。

1.4.2 ラベルの確認

`describe` コマンドでデータセットに記録されている各種情報を確認できます。メニューバーで、**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**、と操作します。`describe` ダイアログで、「変数」フィールドを空欄のまま「OK」をクリックします。`describe` コマンドによって、3つのラベル情報が出力されていることを確認します。

```
Contains data
  obs:          96
  vars:          7
```

variable name	storage type	display format	value label	variable label
city	str6	%9s		観測地
yearmonth	str7	%9s		観測年月
rainfall	double	%10.0g		月合計降水量
summerday	byte	%10.0g		夏日の観測数
temp_mean	double	%10.0g		月平均気温
snow	byte	%12.0g	降雪の有無	
fourseasons	byte	%10.0g		降雪の有無 四季

```
Sorted by:
  Note: Dataset has changed since last saved.
```

値ラベルの内容は `codebook` コマンドにより確認できます。メニューバーで、**データ > データの内容表示 > コードブックの表示**、と操作します。変数フィールドに、値ラベルを付加したカテゴリ変数を選んで OK をクリックします。コマンド操作では次のように入力します。

```
. codebook snow
```

第1章 基本操作とデータ入力

```
. codebook snow
```

```

snow                                     降雪の有無

      type: numeric (byte)
      label: 降雪の有無

      range: [0,1]                      units: 1
unique values: 2                        missing .: 0/96

      tabulation:  Freq.   Numeric  Label
                   75      0      降雪なし
                   21      1      降雪あり

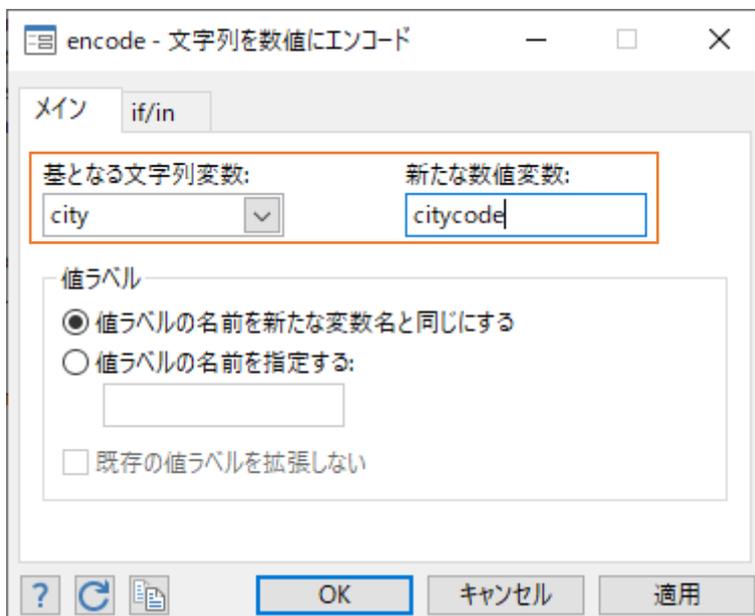
```

1.4.3 文字データを数値データに変換する

サンプルデータの変数 `city` は観測地を表す文字列変数ですが、文字列変数はそのままでは計算に使用することができません。Stata の `encode` コマンドでは、各文字列に数値を当てはめて、カテゴリカルな変数を作成することができます。

メインメニューの **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 > 文字列変数から値ラベルをエンコード**、を選択します。コマンドは次のようになります。

```
. encode city, generate(citycode)
```



データエディタを開き、変数 `city` と生成した変数 `citycode` を確認します。2 つの変数は離れた位置にあるため、変数 `city` を選択した状態で右クリックし、「選択した変数を固定する」をクリックします。

1.4 変数のラベル

データエディタ(ブラウズ) - [無題]

ファイル(F) 編集(E) 表示(V) データ(D) ツール(T)

1C 新潟

	city	yearmo...	rainfall	summerdav	temp_mean	snow	fourseaso...	citycode
1	新潟				3	降雪あり		4 新潟
2	東京				5.8	降雪なし		4 東京
3	根室				-4.2	降雪あり		4 根室
4	那覇				18.4	降雪なし		4 那覇
5	新潟				1.7	降雪あり		4 新潟
6	東京				4.7	降雪あり		4 東京
7	根室				-2.5	降雪あり		4 根室
8	那覇				17.2	降雪なし		4 那覇
9	新潟				3.8	降雪あり		4 新潟
10	東京				6.9	降雪なし		4 東京
11	根室				-3.1	降雪あり		4 根室
12	那覇				17.1	降雪なし		4 那覇
13	新潟				1.4	降雪あり		4 新潟
14	東京				5.4	降雪あり		4 東京
15	根室				-4.1	降雪あり		4 根室
16	那覇				16.9	降雪なし		4 那覇
17	新潟				6.1	降雪あり		1 新潟
18	東京	2006/10	80.5		8.5	降雪なし		1 東京
19	根室	2006/10	24.5	0	-1.4	降雪あり		1 根室
20	那覇	2006/10	96.5	1	18.8	降雪なし		1 那覇

変数 city が固定されるので右方向にスクロールして 2 つの変数を並べて表示することができます。

データエディタ(ブラウズ) - [無題]

ファイル(F) 編集(E) 表示(V) データ(D) ツール(T)

city[13] 新潟

	city	citycode
1	新潟	新潟
2	東京	東京
3	根室	根室
4	那覇	那覇
5	新潟	新潟
6	東京	東京
7	根室	根室
8	那覇	那覇
9	新潟	新潟
10	東京	東京
11	根室	根室
12	那覇	那覇
13	新潟	新潟
14	東京	東京
15	根室	根室
16	那覇	那覇
17	新潟	新潟
18	東京	東京
19	根室	根室
20	那覇	那覇

準備完了 文字数: 6 変数: 8 列順: データセット 観測値: 96

右クリックして「ピン留めされた変数をクリアする」を選択すると解除されます。

第2章 グラフの作成

この章では、ロードしたデータセットからグラフを作成します。頻繁に使用されるヒストグラムと二元グラフを例として紹介します。

2.1 ヒストグラム

メニューバーで、**グラフィックス > ヒストグラム**、と操作して、「ヒストグラム」ダイアログを呼び出します。

1. 「データ」フィールドの「変数」テキストボックスでプロットする変数を選択します。

histogram - 連続とカテゴリ変数のヒストグラム

メイン if/in 加重 密度プロット プロット追加 y軸 x軸 タイトル 凡例 全般 by条件

データ

変数: summerday

連続データ
 離散データ

ビン

10 ビンの数
 ビンの幅
 最初のビンの下限値

y軸

密度
 割合
 度数
 パーセント

棒の高さのラベルを追加する

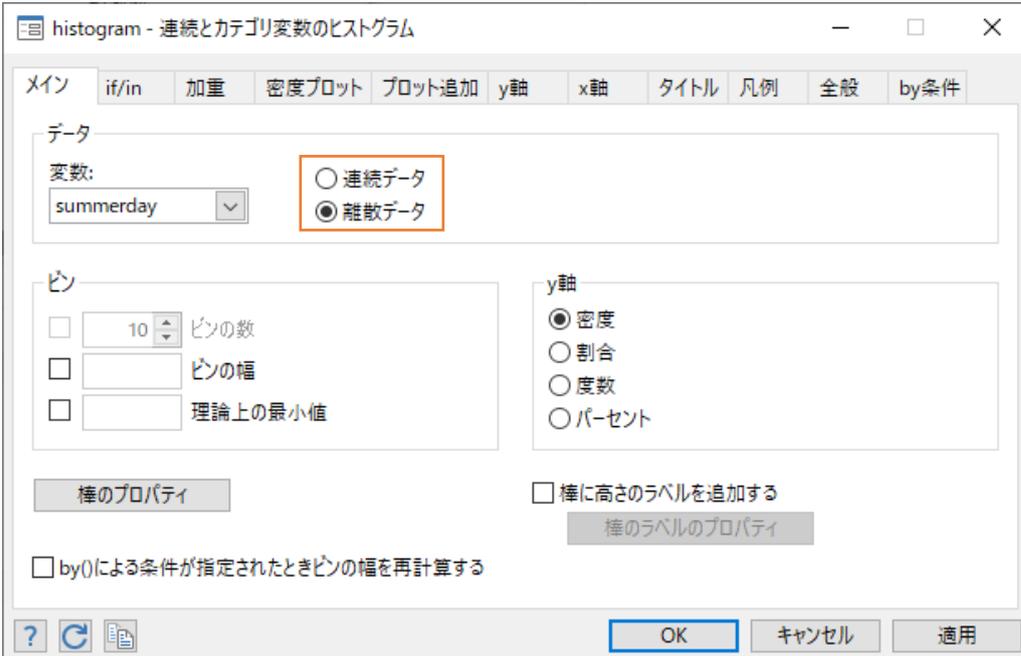
棒のラベルのプロパティ

by()による条件が指定されたときビンの幅を再計算する

? ↻ 📄 OK キャンセル 適用

2. 右隣のラジオボタンで、離散・連続データのいずれかを指定します。

2.1 ヒストグラム



histogram - 連続とカテゴリ変数のヒストグラム

メイン if/in 加重 密度プロット プロット追加 y軸 x軸 タイトル 凡例 全般 by条件

データ
変数: summerday
 連続データ
 離散データ

ピン
 10 ピンの数
 ピンの幅
 理論上の最小値

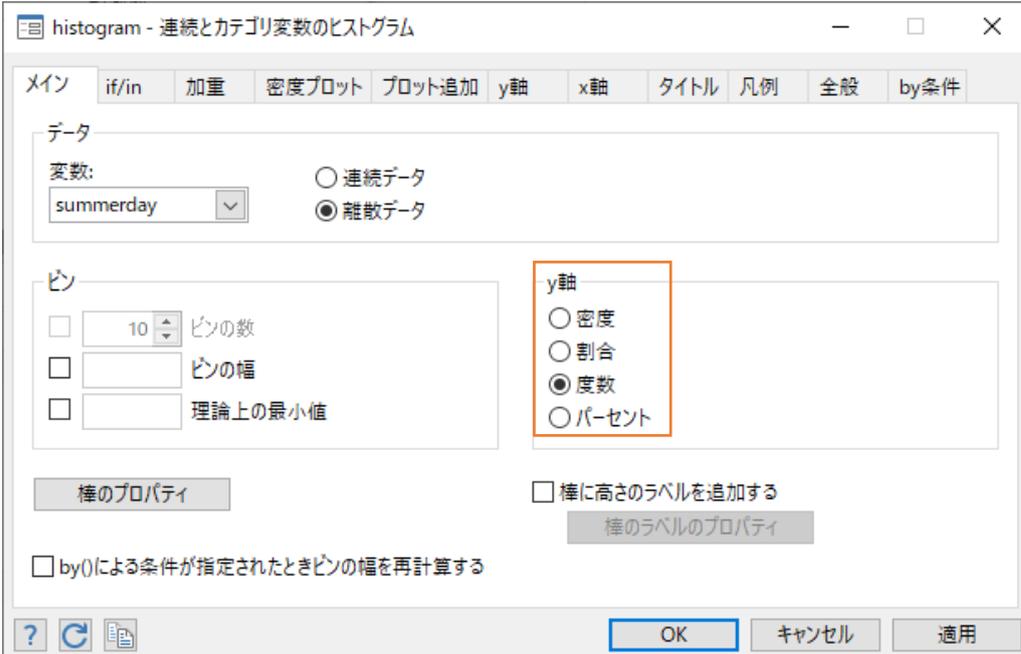
y軸
 密度
 割合
 度数
 パーセント

棒のプロパティ
 棒に高さのラベルを追加する
棒のラベルのプロパティ

by()による条件が指定されたときピンの幅を再計算する

? [OK] キャンセル 適用

3. 「y軸」フィールドで、「度数」を選択します。



histogram - 連続とカテゴリ変数のヒストグラム

メイン if/in 加重 密度プロット プロット追加 y軸 x軸 タイトル 凡例 全般 by条件

データ
変数: summerday
 連続データ
 離散データ

ピン
 10 ピンの数
 ピンの幅
 理論上の最小値

y軸
 密度
 割合
 度数
 パーセント

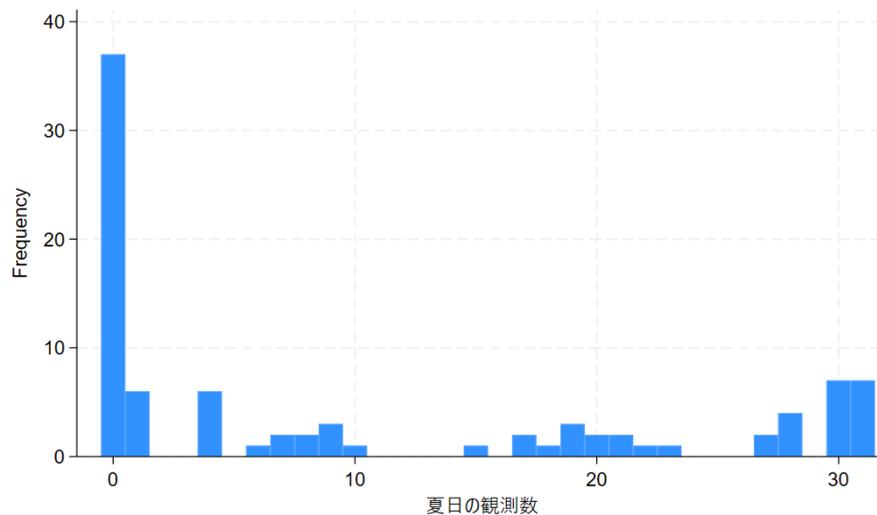
棒のプロパティ
 棒に高さのラベルを追加する
棒のラベルのプロパティ

by()による条件が指定されたときピンの幅を再計算する

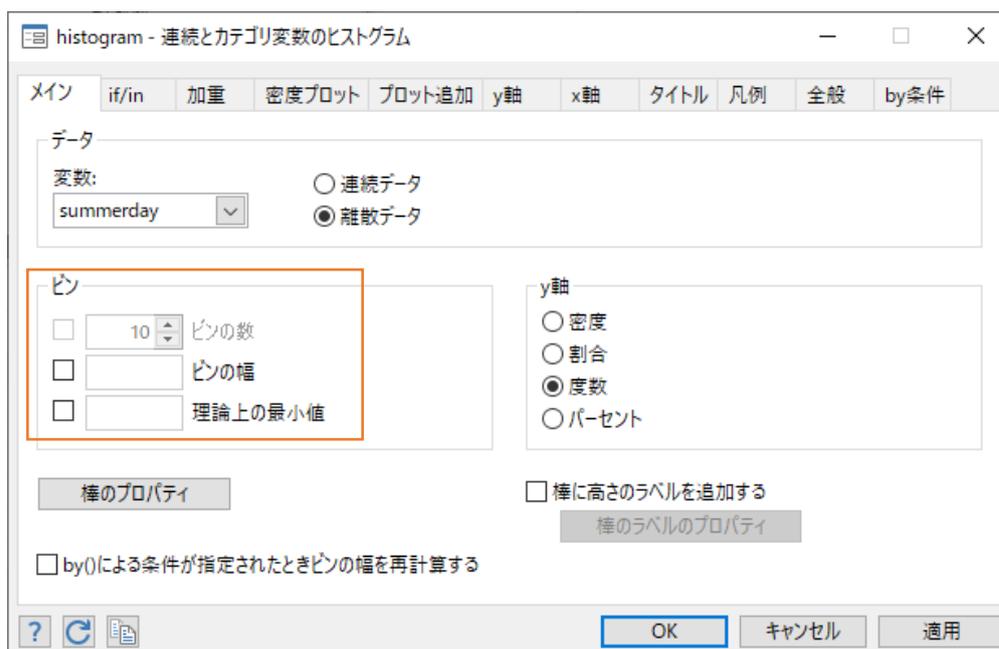
? [OK] キャンセル 適用

4. 「OK」ボタンをクリックして作成します。

第1章 基本操作とデータ入力



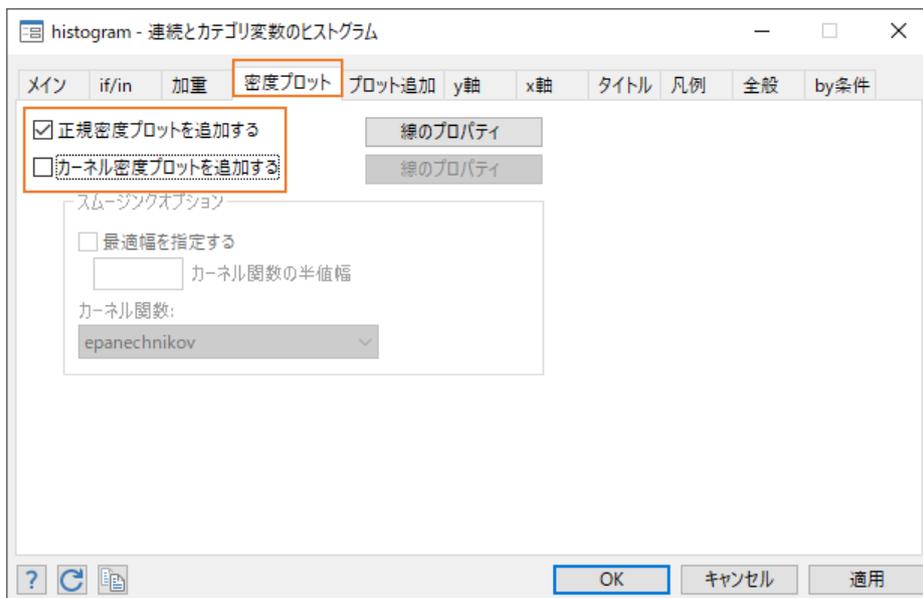
- 「ビン」フィールドでは、ヒストグラムのビン数（「連続データ」を選択した時のみ）、ビンの幅、ビンの下限値をそれぞれ指定してグラフを調整できます。



2.1.1 分布曲線を追加する

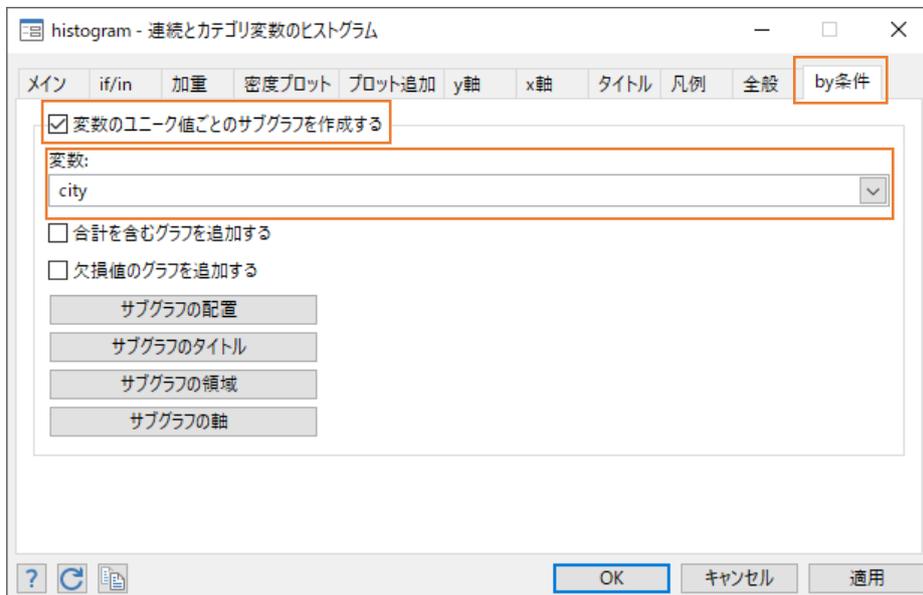
ヒストグラムに分布曲線を追加するには、「密度プロット」タブを開きます。追加する曲線の種類を選択してチェックを入れ、「OK」をクリックしてグラフを作成します。

2.1 ヒストグラム



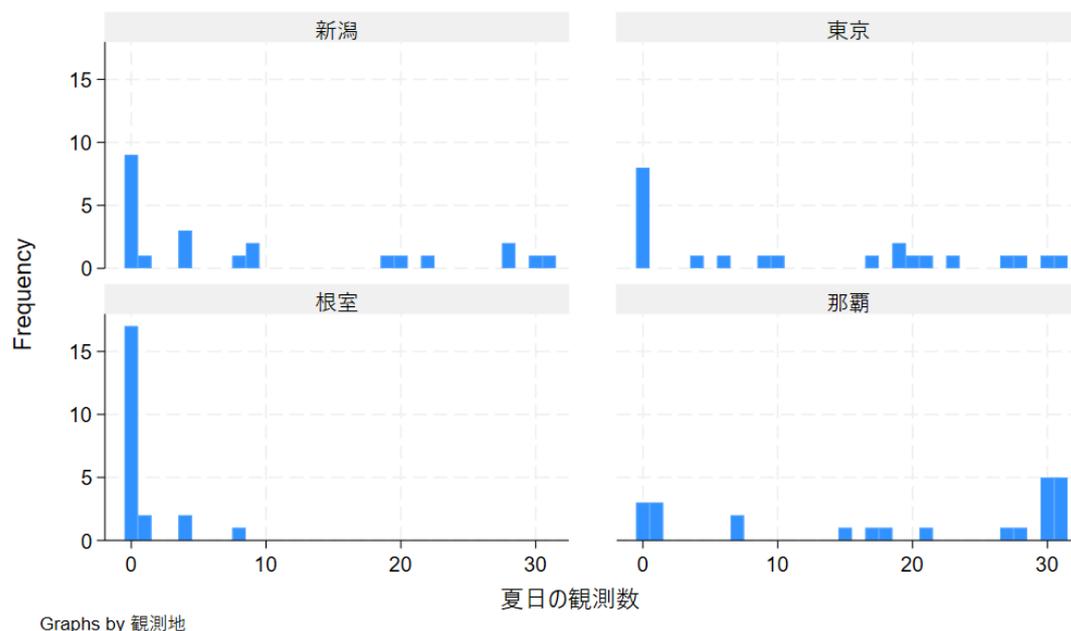
2.1.2 カテゴリごとにヒストグラムを作成する

カテゴリ変数別にヒストグラムを複数作成するには、ダイアログの「by 条件」タブを開きます。「変数のユニーク値ごとにサブグラフを作成する」にチェックして、ここでは、変数 **city** を指定して作成します。



カテゴリ変数ごとにヒストグラムを分割して作成されました。

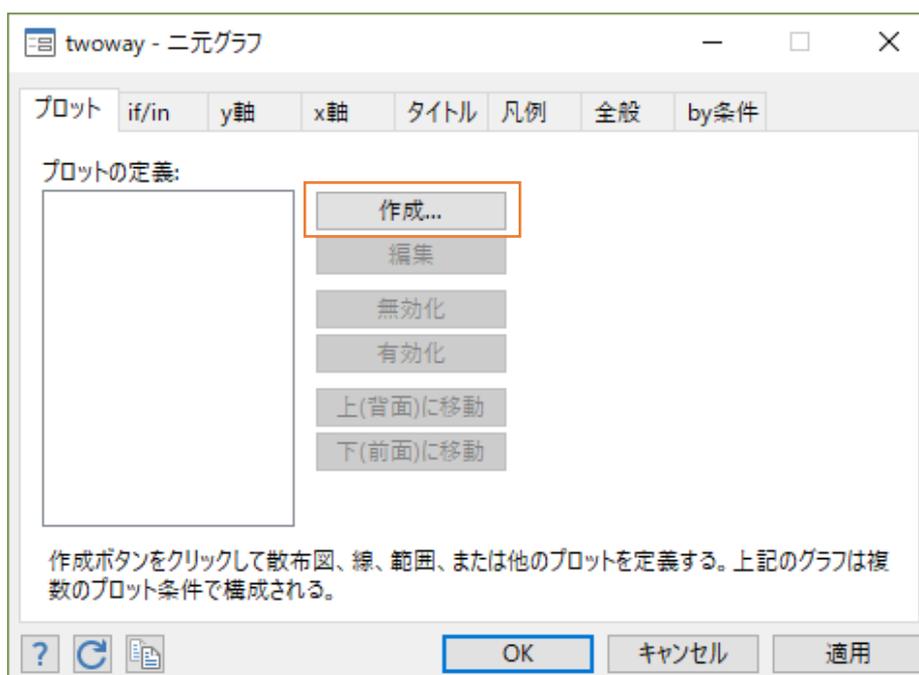
第1章 基本操作とデータ入力



2.2 散布図

散布図は2つの変数の関係を可視化するのに便利です。

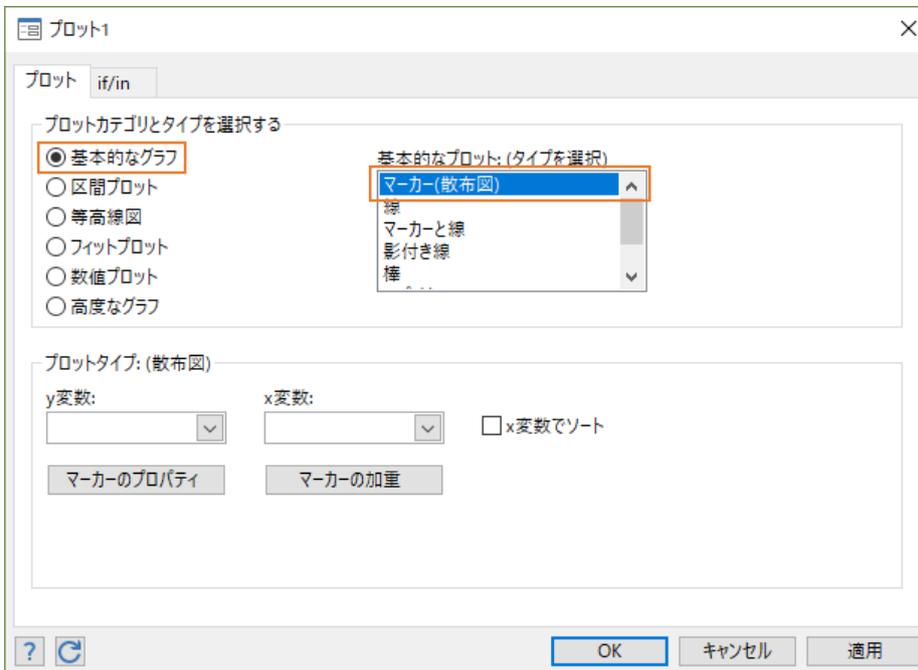
1. メニューバーで、**グラフィックス > 二元グラフ (散布図/折れ線など)**、と操作します。
2. 開いた「二元グラフ」ダイアログで、「作成」ボタンをクリックします。



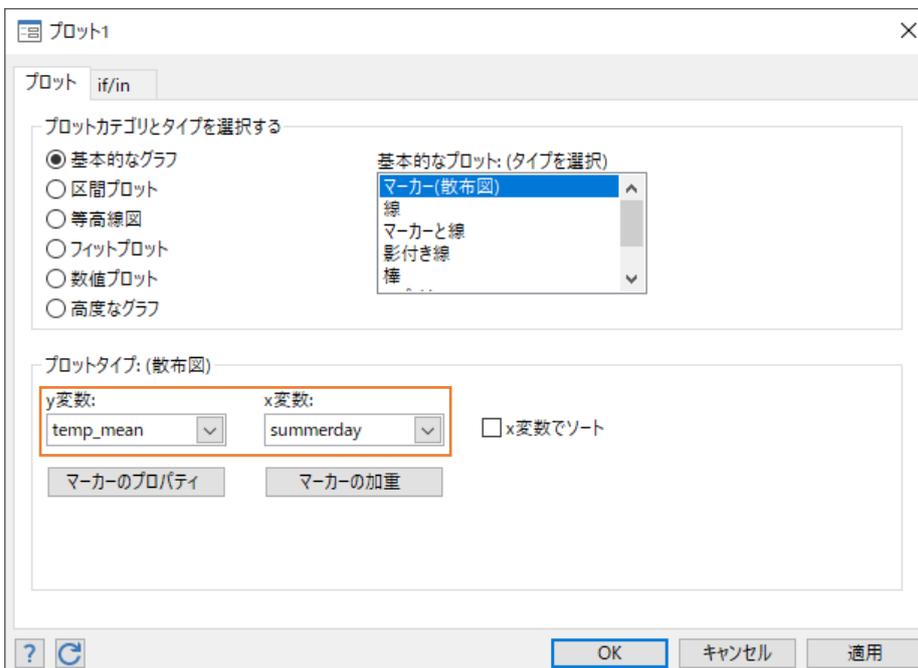
3. 「プロット 1」ダイアログが開きます。このダイアログで、「プロットカテゴリとタイ

2.2 散布図

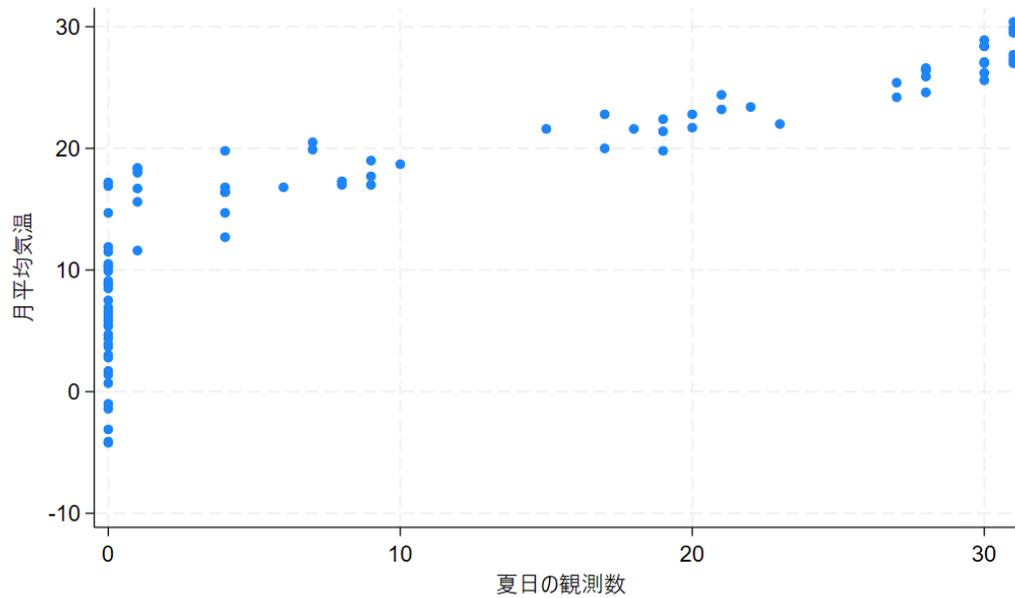
プを選択する」フィールドで「基本的なグラフ」を選択、「基本的なプロット」リストで「マーカー(散布図)」を選択します。



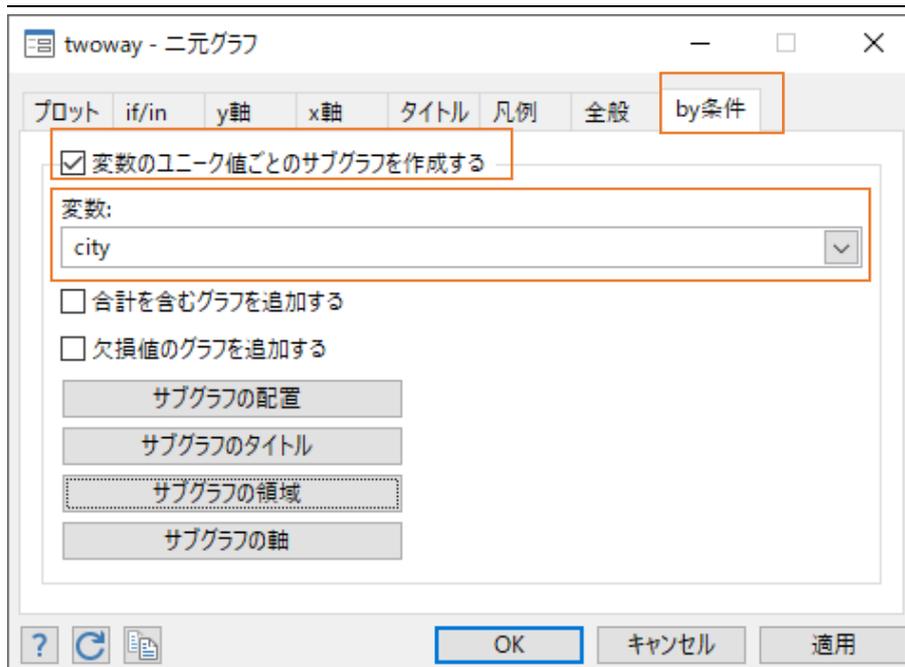
4. 「プロットタイプ: (散布図)」フィールドで「y 変数」と「x 変数」を設定します。



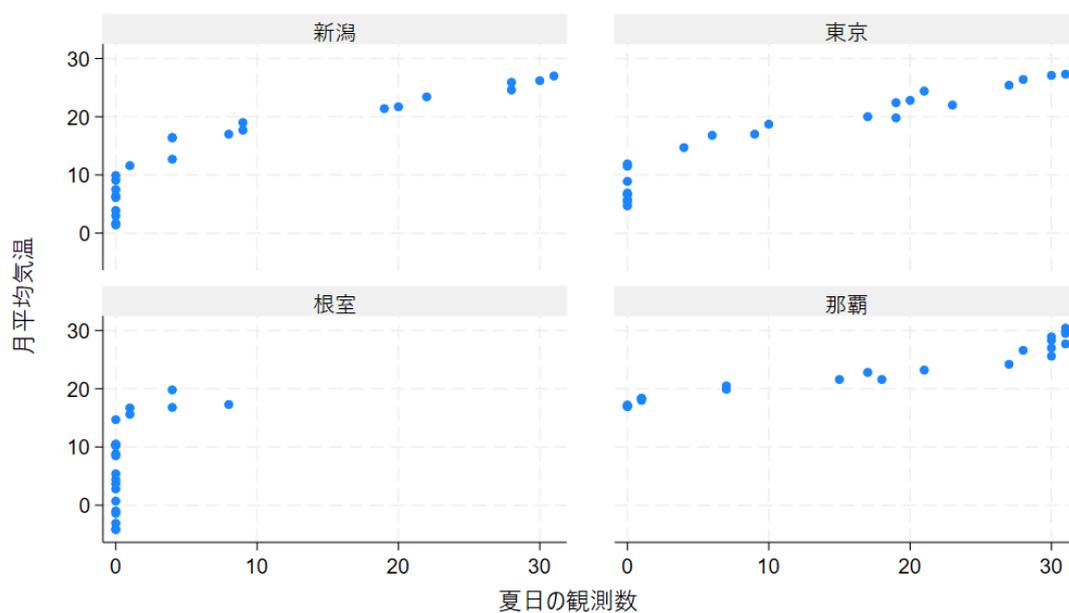
第1章 基本操作とデータ入力



2.2 散布図



カテゴリ変数ごとに散布図を分割して作成されます。



コマンド操作では、`by(カテゴリ変数)`オプションを追加します。全体では次のようになります。

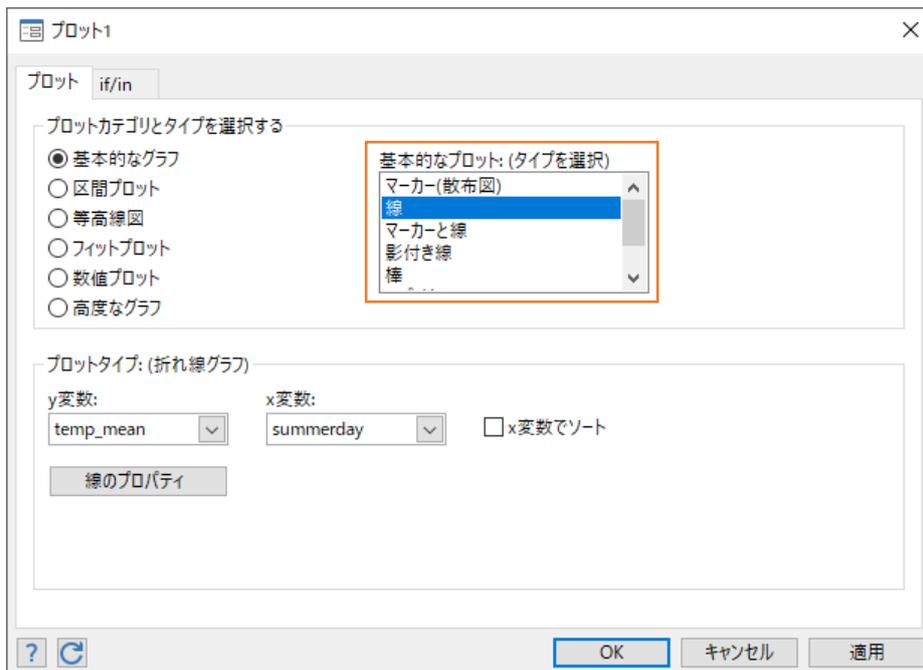
```
. twoway (scatter temp_mean summerday), by(city)
```

第1章 基本操作とデータ入力

2.3 折れ線グラフ

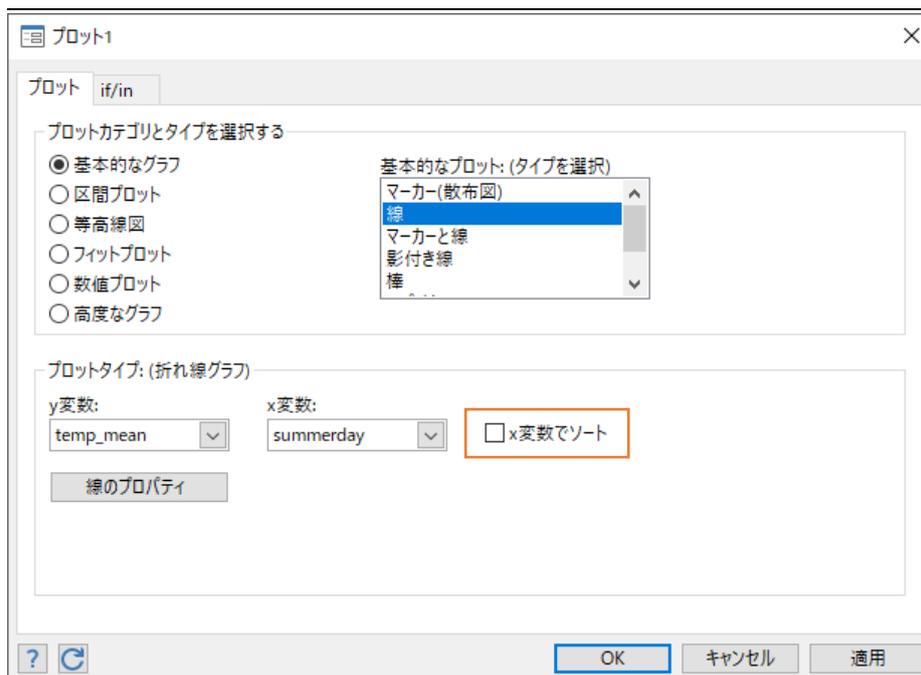
セクション 2.2 では散布図を作成しましたが、プロットタイプに「線」を選ぶと、折れ線グラフを作成できます。

1. 二元グラフダイアログから「作成」をクリックしてプロットダイアログを開き、プロットタイプに「線」を選びます。

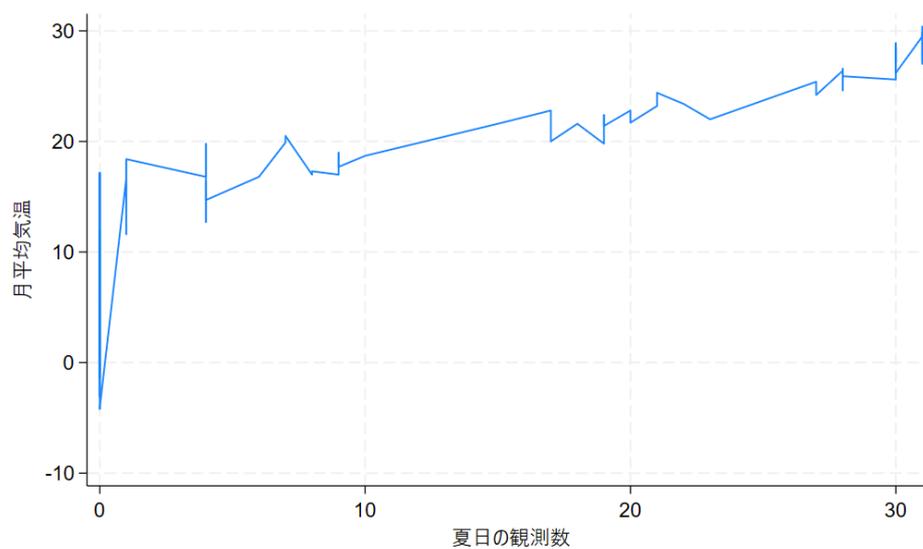


2. 「y 軸」と「x 軸」の変数を指定して、最後に「x 変数でソート」にチェックを入れます。

2.4 種類の異なるグラフを一つに統合する



つぎのような折れ線グラフが作成されます。



コマンド操作では、次のように `sort` オプションを使用します。

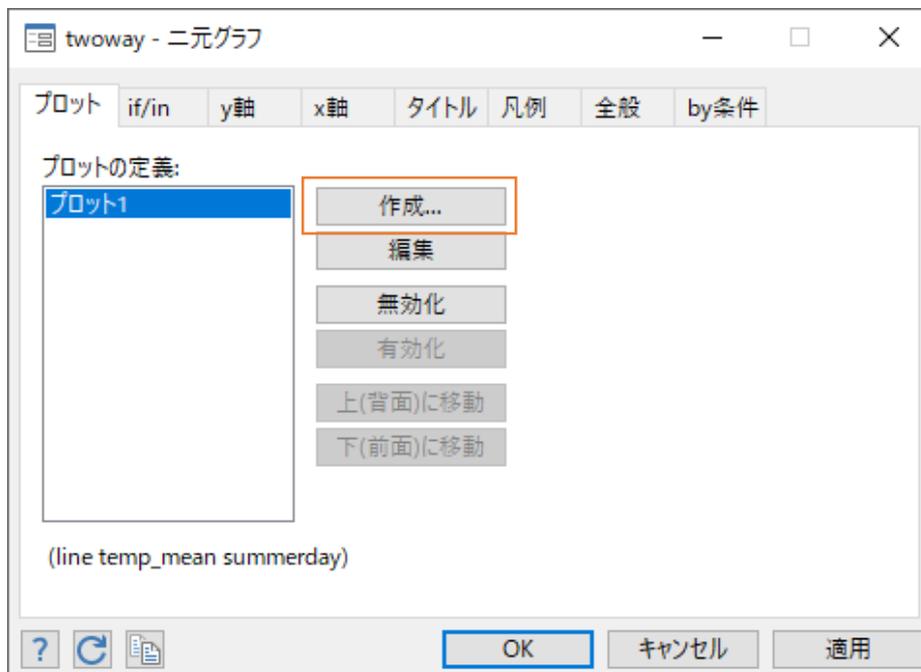
```
. twoway (line temp_mean summerday, sort)
```

2.4 種類の異なるグラフを一つに統合する

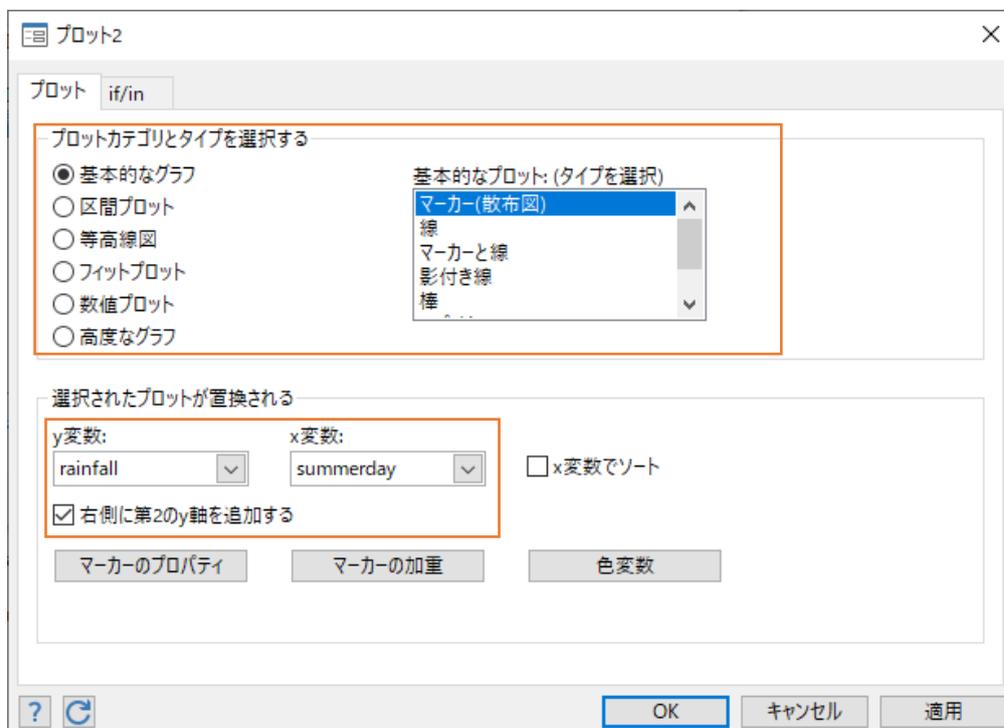
2元グラフでは、複数のグラフを一度に作成して、重ねて1つのグラフにできます。

第1章 基本操作とデータ入力

1. 「二元グラフ」ダイアログで、「作成」をクリックします。

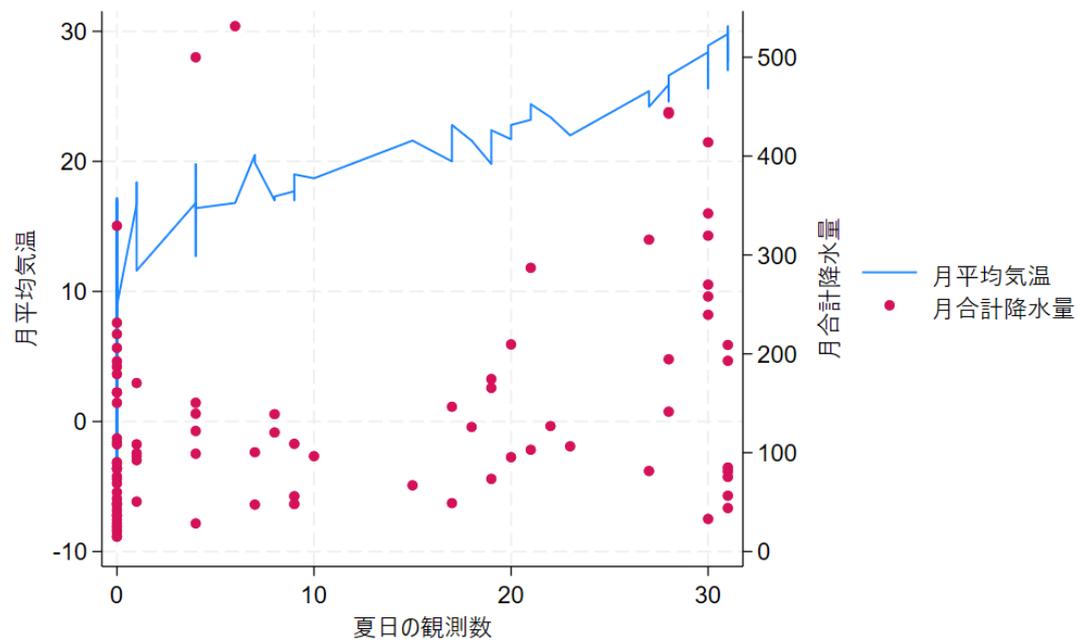


2. プロット 2 ダイアログで、前のセクションのように、プロットタイプと変数を選択します。さらに、プロット 1 と別の y 軸を使用するため、「右側に第 2 の y 軸を追加する」にチェックを入れます。「OK」をクリックします。



2.5 グラフの共通設定

3. 「二元グラフ」ダイアログに戻り、OK をクリックしてグラフを作成します。



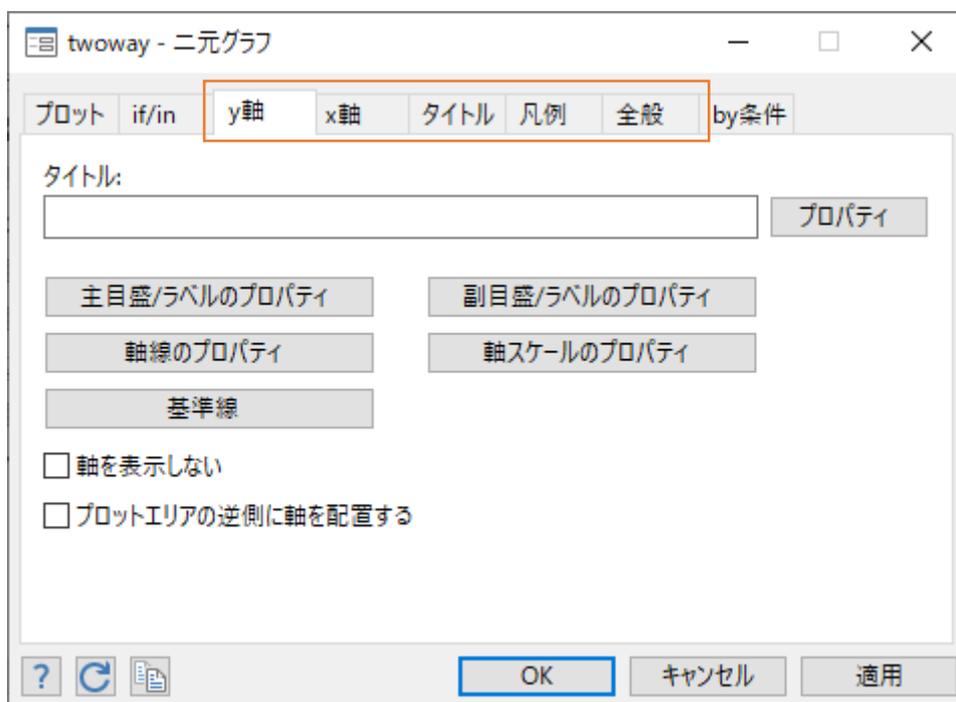
コマンド操作では、次のようになります。

```
. twoway (line temp_mean summerday, sort) (scatter rainfall summerday,
yaxis(2))
```

2.5 グラフの共通設定

「二元グラフ」ダイアログでは、グラフタイトルや軸タイトル、軸目盛をオプションで設定します。

第1章 基本操作とデータ入力



2.5.1 y軸・x軸オプション

「二元グラフ」ダイアログの「y軸」と「x軸」タブでは、次の設定を行うことができます。

- タイトル：編集ボックスでyまたはx軸のタイトルを入力します。
- 主目盛/ラベルのプロパティ：軸の目盛の範囲・増分と、目盛・ラベル・グリッド線の書式を設定します。
- 副目盛/ラベルのプロパティ：軸の副目盛のスケールと、副目盛・ラベル・グリッド線の書式を設定します。
- 軸線のプロパティ：軸の書式を設定します。
- 軸スケールのプロパティ：軸のスケールを設定します。
- 基準線：基準線の書式を設定します。
- 軸を表示しない：軸線を非表示にします。
- プロットエリアの逆側に軸を配置する：軸線を左右または上下逆に表示します。

2.5.2 グラフタイトルのオプション

「タイトル」タブでは、グラフに表示されるタイトル等の情報を設定できます。

- タイトル：グラフのタイトルと書式を設定します。
- サブタイトル：グラフのサブタイトルと書式を設定します。
- キャプション：グラフへのコメントなどを設定します。
- 備考：グラフのメモなどを設定します。

2.6 異なるグラフを並べて表示する

2.5.3 凡例オプション

「凡例」タブでは、グラフの凡例の表示と配置の設定を行います。

- 凡例の表示：グラフに凡例を表示/非表示を選択します。
- 規定のキーを一時的に変更する：カテゴリ変数を使用する際のラベル表示を一時的に変更します。
- 配置/表示：凡例ボックス内のラベルの配置・書式・表示位置を設定します。
- 配置：凡例ボックスを表示する位置を設定します。

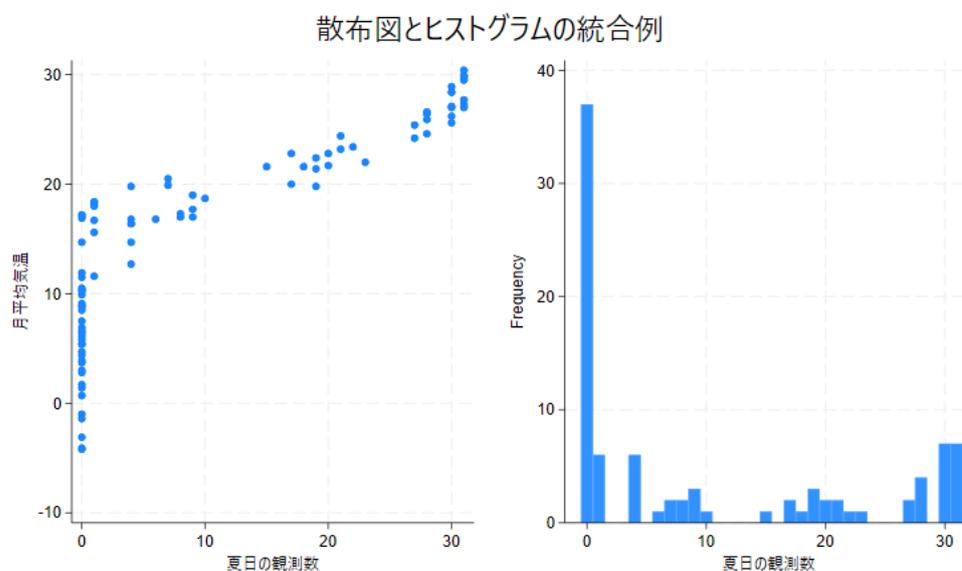
2.5.4 グラフの全般オプション

「全般」タブでは、グラフの全体に関する設定を行います。

- スキーム：軸・タイトル・背景などの各設定を、規定のスキームに当てはめ、一括変更します。
- グラフ名：グラフの名前を入力します。この名前を設定すると、グラフを閉じても再び呼び出すことができます。
- グラフサイズ：グラフの幅と高さを設定します。単位はインチです。

2.6 異なるグラフを並べて表示する

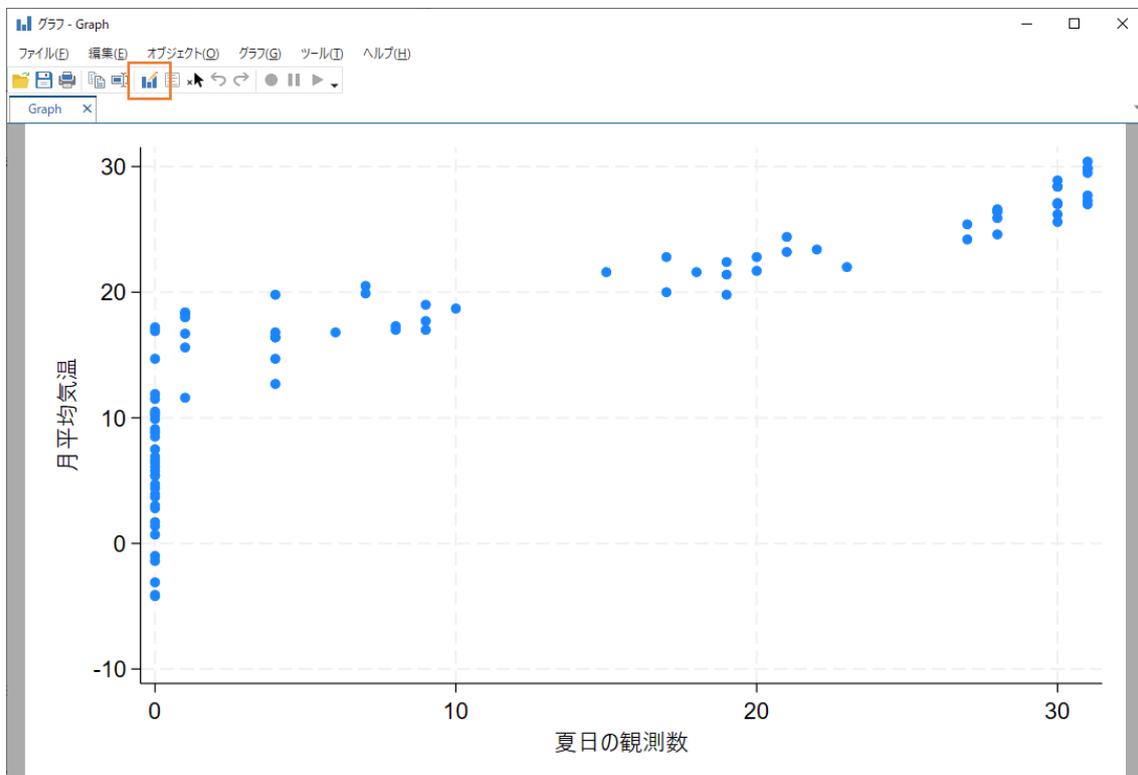
メモリに保存したグラフファイルは呼び出して、再表示することができます。また、複数のグラフを呼び出して、並べて表示できます。**グラフィックス > グラフを統合**、を選択します。「複数のグラフを統合」ダイアログの「統合するグラフ」ボックスでディスク上に保存されているグラフを選択します。OK をクリックして統合します。



第1章 基本操作とデータ入力

2.7 グラフエディタ

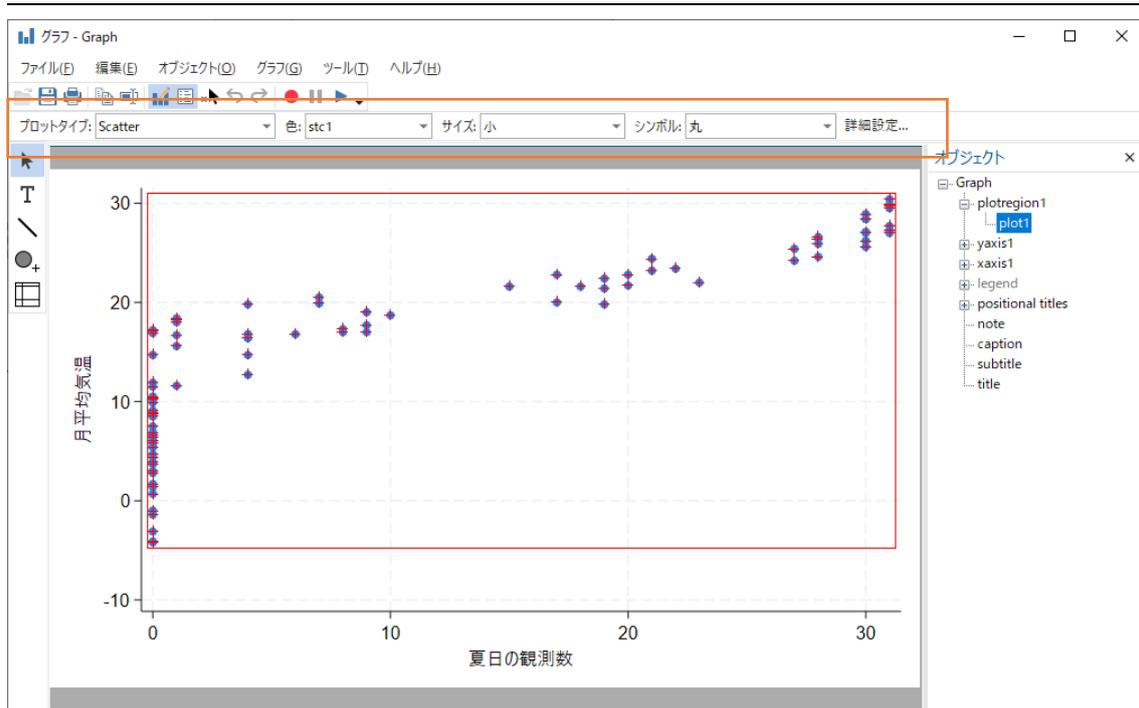
グラフウィンドウからグラフエディタを起動すると、作成済みのグラフを編集できます。



グラフエディタでは、グラフを構成するタイトルやマーカー、目盛をカーソルで選択して書式を変更できます。変更できる主な項目は次の通りです。

- マーカー：色、サイズ、シンボル
- タイトル・サブタイトル・軸タイトル：色、サイズ、余白、表示テキスト
- 軸ラベル：サイズ、表示角度、グリッド線の有無
- 凡例ボックス：サイズ、色、ボックスの色

2.8 グラフの保存



2.8 グラフの保存

Stata で作成したグラフは、グラフウィンドウを閉じると、削除されてしまいます。保存はグラフウィンドウで、**ファイル > 保存**、を選択します。デフォルトでは、グラフを Stata 独自の .gph 形式でグラフを保存しますが、保存ダイアログでは PNG 形式や PDF 形式で保存することもできます。



第3章 基本的な統計操作

本章では、基本的な統計機能、及びそれらに関連したデータ管理機能について、その用法を紹介します。

3.1 要約統計量

ある標本（データセット）が与えられたとき、その平均値や中央値、標準偏差などのデータの分布に関する情報を求めたいことが良くあります。このような場合に使用される代表的なコマンドが `summarize` です。

1. 変数の要約統計量は**統計 > 要約/表/検定 > 記述統計量**、を選択します。
2. 「記述統計量」ダイアログの「変数」ドロップダウンリストで、統計量を求める変数を選択します。
3. オプションを選び、OK をクリックします。

`summarize` コマンドは標準では、観測数、平均値、標準偏差、最小値、最大値を表示します。

```
. summarize temp_mean
```

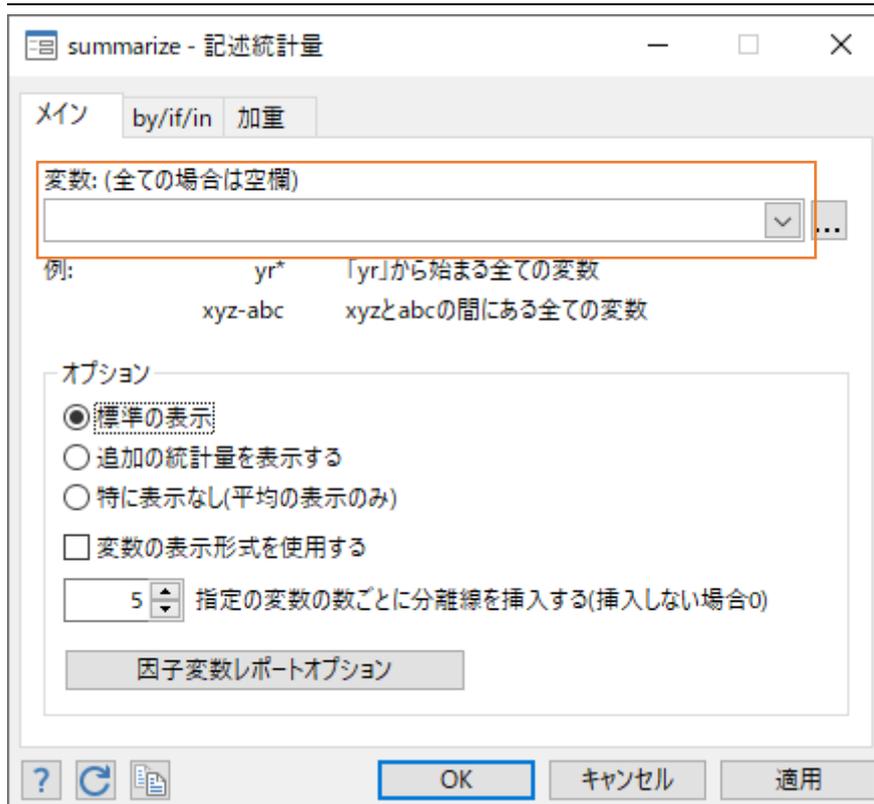
Variable	Obs	Mean	Std. Dev.	Min	Max
temp_mean	96	15.11354	9.480835	-4.2	30.4

より詳細な統計情報を入手するには、オプションの「追加の統計量を表示する」を選択します。これは、分散、歪度、尖度、パーセンタイルを表示します。

```
. summarize temp_mean, detail
```

temp_mean					
Percentiles		Smallest			
1%	-4.2	-4.2			
5%	-1.4	-4.1			
10%	1.7	-3.1	Obs		96
25%	7.2	-2.5	Sum of Wgt.		96
50%	16.85		Mean		15.11354
			Std. Dev.		9.480835
		Largest			
75%	22.8	29.5			
90%	27.1	29.8	Variance		89.88623
95%	28.9	29.9	Skewness		-.2596824
99%	30.4	30.4	Kurtosis		2.007305

第3章 基本的な統計操作



コマンド操作では、次の通りです。

```
. summarize temp_mean
```

3.2.1 正規分布

変数の分布が正規分布にしているかを確認したい場合は、コマンド `sktest` を使用します。`sktest` は変数の歪度と尖度が正規分布と異なっているかを検定します。

分布のゆがみを表す統計量が歪度(skewness)です。歪度が正の値ならば、分布は正の方向に、負の値であれば、負の方向に歪んでいます。正規分布では、歪度は0をとります。

分布の尖り具合を示す統計量が尖度(kurtosis)です。尖度が大きいほど、分布曲線は鋭くなります。一方、尖度が小さいほど、平らな分布になります。正規分布では尖度は3をとります。

1. **統計 > 要約/表/検定 > 分布のプロットと検定 > 歪度/尖度による正規検定**、を選択します。
2. 「変数」テキストボックスに検定対象となる変数を入力します。
3. OK をクリックします。

コマンド操作では、次のようになります。

3.3 度数分布表

```
. sktest temp_mean
```

`sktest` は「変数の分布は正規分布に従っている」を帰無仮説として検定を実行します。検定結果は次のように表示されます。

```
. sktest temp_mean
```

Skewness/Kurtosis tests for Normality					
Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	joint	
				adj chi2(2)	Prob>chi2
temp_mean	96	0.2742	0.0003	11.68	0.0029

3.3 度数分布表

カテゴリ変数データを整理する場合には度数分布表が良く用いられます。Stata には 1 元の度数分布表を作成するための `tabulate oneway` コマンドと 2 元の度数分布表(クロス集計表)を作成するための `tabulate twoway` コマンドがあります。

3.3.1 一元配置表

一元配置表は、1 つのカテゴリ変数の各値の頻度を表にまとめます。表の作成は **統計 > 要約/表/検定 > 度数分布表 > 一元配置表**、を選択します。コマンド操作では、`tabulate` を利用し変数名を指定して、次のように入力します。

```
. tabulate summerday
```

```
. tabulate summerday
```

summerday	Freq.	Percent	Cum.
0	42	43.75	43.75
1	6	6.25	50.00
4	6	6.25	56.25
6	1	1.04	57.29
7	2	2.08	59.38
8	2	2.08	61.46
9	3	3.13	64.58
10	1	1.04	65.63
15	1	1.04	66.67
17	2	2.08	68.75
18	1	1.04	69.79
19	3	3.13	72.92
20	2	2.08	75.00
21	2	2.08	77.08
22	1	1.04	78.13
23	1	1.04	79.17
27	2	2.08	81.25
28	4	4.17	85.42
30	7	7.29	92.71
31	7	7.29	100.00
Total	96	100.00	

主なオプション

第3章 基本的な統計操作

- 欠損値を他の値と同様に扱う：`tabulate` コマンドは通常、欠損値を排除して集計します。このオプションでは、欠損値も合わせて集計します。コマンドは `missing` です。

```
. tabulate summerday, miss
```

summerday	Freq.	Percent	Cum.
0	37	38.54	38.54
1	6	6.25	44.79
4	6	6.25	51.04
6	1	1.04	52.08
7	2	2.08	54.17
8	2	2.08	56.25
9	3	3.13	59.38
10	1	1.04	60.42
15	1	1.04	61.46
17	2	2.08	63.54
18	1	1.04	64.58
19	3	3.13	67.71
20	2	2.08	69.79
21	2	2.08	71.87
22	1	1.04	72.92
23	1	1.04	73.96
27	2	2.08	76.04
28	4	4.17	80.21
30	7	7.29	87.50
31	7	7.29	94.79
.	5	5.21	100.00
Total	96	100.00	

- 度数に関して降順で表示する：度数を大きいものから小さいものへ、降順で表示します。コマンドは `sort` です。

```
. tabulate summerday, sort
```

summerday	Freq.	Percent	Cum.
0	37	40.66	40.66
30	7	7.69	48.35
31	7	7.69	56.04
1	6	6.59	62.64
4	6	6.59	69.23
28	4	4.40	73.63
9	3	3.30	76.92
19	3	3.30	80.22
7	2	2.20	82.42
8	2	2.20	84.62
17	2	2.20	86.81
20	2	2.20	89.01
21	2	2.20	91.21
27	2	2.20	93.41
6	1	1.10	94.51
10	1	1.10	95.60
15	1	1.10	96.70
18	1	1.10	97.80
22	1	1.10	98.90
23	1	1.10	100.00
Total	91	100.00	

- 値ラベルではなく数値コードを表示する：値ラベルが設定されている場合、`tabulate` は集計結果をラベルで表示しますが、このオプションでは、数値コードを表示します。

3.3 度数分布表

値ラベルで表示した場合

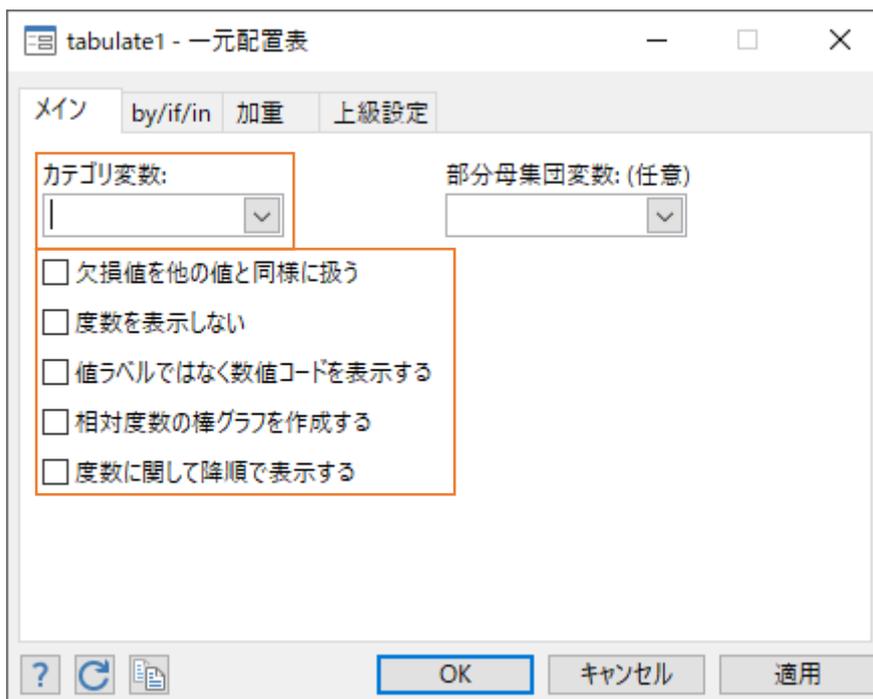
```
. tabulate snow
```

snow	Freq.	Percent	Cum.
降雪なし	75	78.13	78.13
降雪あり	21	21.88	100.00
Total	96	100.00	

数値コードで表示した場合

```
. tabulate snow, nolabel
```

snow	Freq.	Percent	Cum.
0	75	78.13	78.13
1	21	21.88	100.00
Total	96	100.00	



3.3.2 二元配置（クロス集計）表

二元配置表は2つのカテゴリ変数の対応関係を表にまとめます。二元配置表の作成は、**統計** > **要約/表/検定** > **度数分布表** > **二元配置表/関連係数**、を選択します。

1. 「2元配置表」ダイアログで「行の変数」に `snow` と「列の変数」に `fourseasons` を選択します。
2. オプションを選択します。

第3章 基本的な統計操作

3. OK をクリックします。

```
. tabulate snow fourseasons
```

snow	fourseasons				Total
	1	2	3	4	
0	18	24	23	10	75
1	6	0	1	14	21
Total	24	24	24	24	96

tabulate2 - 二元配置表と関連係数

メイン by/if/in 加重 上級設定

行の変数: snow

列の変数: fourseasons

検定統計量

- ピアソンのカイニ乗
- フィッシャーの正確確率検定
- GoodmanとKruskalのガンマ
- 尤度比カイニ乗
- ケンドールのタウb
- クラメールのV

セルの内容

- ピアソンのカイニ乗
- 列内の相対度数
- 行内の相対度数
- 尤度比カイニ乗
- 相対度数
- 期待度数
- 度数を表示しない

親測度数の順に行を表示する

親測度数の順に列を表示する

欠損値を他の値と同様に扱う

表を折り返さない

セルの凡例を表示する

値ラベルを表示しない

計数ログを表示しない

OK キャンセル 適用

- パーセンテージ情報の付加

基本的な 2 元度数分布表は上記のように作成しますが、行・列のパーセンテージが表示されていると、さらに便利です。各セルの数値の意味が左上にキーとして表示されます。

row オプションでは行ごと、column オプションは列ごとのパーセンテージを表示します。

3.3 度数分布表

● 列ごとのパーセンテージを表示した例

```
. tabulate snow fourseasons, row
```

Key
<i>frequency</i>
<i>row percentage</i>

snow	fourseasons				Total
	1	2	3	4	
0	18 24.00	24 32.00	23 30.67	10 13.33	75 100.00
1	6 28.57	0 0.00	1 4.76	14 66.67	21 100.00
Total	24 25.00	24 25.00	24 25.00	24 25.00	96 100.00

● 列ごとのパーセンテージを表示した場合

```
. tabulate snow fourseasons, column
```

Key
<i>frequency</i>
<i>column percentage</i>

snow	fourseasons				Total
	1	2	3	4	
0	18 75.00	24 100.00	23 95.83	10 41.67	75 78.13
1	6 25.00	0 0.00	1 4.17	14 58.33	21 21.88
Total	24 100.00	24 100.00	24 100.00	24 100.00	96 100.00

検定統計量オプション

● ピアソンのカイ二乗

「行と列が独立している」を帰無仮説として仮定し、カイ二乗検定を行います。結果は表の下に表示されます。

```
. tabulate snow fourseasons, chi2
```

snow	fourseasons				Total
	1	2	3	4	
0	18	24	23	10	75
1	6	0	1	14	21
Total	24	24	24	24	96

```
Pearson chi2(3) = 29.9276 Pr = 0.000
```

第3章 基本的な統計操作

● Goodman と Kruskal のガンマ

Goodman と Kruskal のガンマと標準誤差を算出します。結果は表の下に表示されます。

```
. tabulate snow fourseasons, gamma
```

snow	fourseasons				Total
	1	2	3	4	
0	18	24	23	10	75
1	6	0	1	14	21
Total	24	24	24	24	96

```
gamma = 0.4601 ASE = 0.190
```

● ケンドールのタウ b

ケンドールのタウ b と標準誤差を算出します。結果は表の下に表示されます。

```
. tabulate snow fourseasons, taub
```

snow	fourseasons				Total
	1	2	3	4	
0	18	24	23	10	75
1	6	0	1	14	21
Total	24	24	24	24	96

```
Kendall's tau-b = 0.2572 ASE = 0.111
```

3.4 仮説検定

2つの標本が与えられたとき、それぞれの母平均である μ_1 と μ_2 が等しいと言えるかどうかを検定するには t 検定が一般に用いられますが、そのコマンドが **ttest** です。また、コマンド **sdtest** は変数の等分散性を検定します。この2つのコマンドにはいくつかのバリエーションがあり、自分の仮説に適したものを使用する必要があります。

3.4.1 Stata のデータ構成と変形

サンプルデータセットの変数 **temp_mean** は各都市の平均気温を表しています。サンプルでは、24 か月間の 4 都市の平均気温が 1 列に並んでいます。このようなデータ構成を Stata ではロング形式と呼びます。対応関係のある t 検定では、**temp_mean1** と **temp_mean2** のように対応関係にある 2 つの変数が必要になります。

	temp_mean
1.	3.0999999
2.	1.7
3.	3.3
4.	1.4
5.	6.0999999

ロング形式

	temp_me~1	temp_me~2
1.	3.0999999	5.8000002
2.	1.7	4.6999998
3.	3.3	6.9000001
4.	1.4	5.4000001
5.	6.0999999	8.5

ワイド形式

データ構成を変更するには **reshape** コマンドを使用または、メインメニューで **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 > データのワイド形式/ロング形式を**

3.4 仮説検定

変換、を選択します。操作は次のように行います。

1. reshape ダイアログで、「ロング形式からワイド形式へ」を選択します。

reshape - ワイド型とロング型の間でデータを変換する

ワイド形式からロング形式へ

ロング形式からワイド形式へ

ロング形式へ再変換(以前に変換した場合)

ワイド形式へ再変換(以前に変換した場合)

例...

ID変数 - i()オプション:

サブオブザベーション識別子 - j()オプション

変数: 値:(任意)

サブオブザベーションのIDに文字列も利用する

X_jj変数の基準(接頭語)名:

備考: 全ての他の変数はIDごとに一定値とする。

? ↻ 📄 OK キャンセル 適用

第3章 基本的な統計操作

2. 「ID変数」テキストボックスに、変換のキーとなるID変数を指定します。ここでは、`yearmonth` です。

reshape - ワイド型とロング型の間でデータを変換する

ワイド形式からロング形式へ 例...

ロング形式からワイド形式へ

ロング形式へ再変換(以前に変換した場合)

ワイド形式へ再変換(以前に変換した場合)

ID変数 - i()オプション:
yearmonth

サブオブザベーション識別子 - j()オプション

変数: 値: (任意)

サブオブザベーションのIDに文字列も利用する

X_ij変数の基準(接頭語)名:

備考: 全ての他の変数はIDごとに一定値とする。

? ↻ 📄 OK キャンセル 適用

3.4 仮説検定

3. 「サブオブザベーション識別子」を指定します。ここでは、セクション 1.4.3 で作成した `citycode` です。

reshape - ワイド型とロング型の間でデータを変換する

ワイド形式からロング形式へ

ロング形式からワイド形式へ

ロング形式へ再変換(以前に変換した場合)

ワイド形式へ再変換(以前に変換した場合)

例...

ID変数 - i()オプション:

サブオブザベーション識別子 - j()オプション

変数: citycode

値: (任意)

サブオブザベーションのIDに文字列も利用する

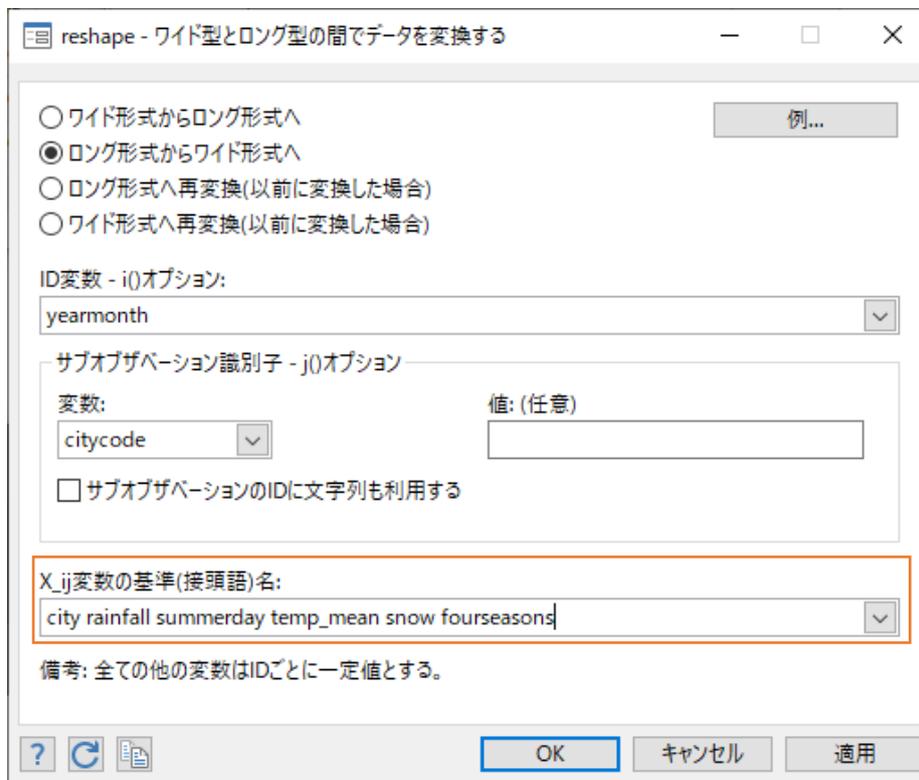
X_j変数の基準(接頭語)名:

備考: 全ての他の変数はIDごとに一定値とする。

? OK キャンセル 適用

第3章 基本的な統計操作

4. 「X_{ij}変数の基準(接頭語)」に変換する変数を指定します。ここでは、city、rainfall、summerday、temp_mean、snow、fourseasons です。



5. 「OK」をクリックして実行します。

以上の手順をコマンドにまとめると次のようになります。

```
. reshape wide rainfall summerday temp_mean snow fourseasons city ,  
i(yearmonth) j(citycode)
```

3.4.2 仮説検定の実行

データの形式変換ができれば、検定を行います。

1. メインメニューの統計 / 要約 / 表 / 検定 > 伝統的な仮説検定 > t検定(平均比較検定)、を選択します。
2. ダイアログのt検定ボックスで「対応のあるデータ」を選択します。
3. 「対応のあるt検定」ボックスで「第1変数」と「第2変数」を、それぞれ指定します。
4. 信頼度を指定します。デフォルトでは95%です。
5. OKをクリックします。

3.4 仮説検定

The screenshot shows the STATA dialog box for a paired t-test. The title is 't検定(平均比較検定)'. Under the 't検定' section, the radio button for '対応のあるデータ' (Matched data) is selected. In the '対応のあるt検定' section, the first variable is 'temp_mean1' and the second variable is 'temp_mean2'. The confidence level is set to 95%.

コマンド操作では次のように第 1 変数と第 2 変数を 2 つのイコール (==) で結んで行います。

```
. ttest temp_mean1 == temp_mean2
```

```
. ttest temp_mean1 == temp_mean2
```

Paired t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
temp_m~1	24	13.89167	1.749326	8.569913	10.27291	17.51042
temp_m~2	24	16.1	1.570874	7.695679	12.8504	19.3496
diff	24	-2.208333	.2523022	1.236024	-2.73026	-1.686406

mean(diff) = mean(temp_mean1 - temp_mean2) t = -8.7527
Ho: mean(diff) = 0 degrees of freedom = 23

Ha: mean(diff) < 0 Ha: mean(diff) != 0 Ha: mean(diff) > 0
Pr(T < t) = 0.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 1.0000

第3章 基本的な統計操作

3.4.2.1 平均値の検定 対応の関係のない場合

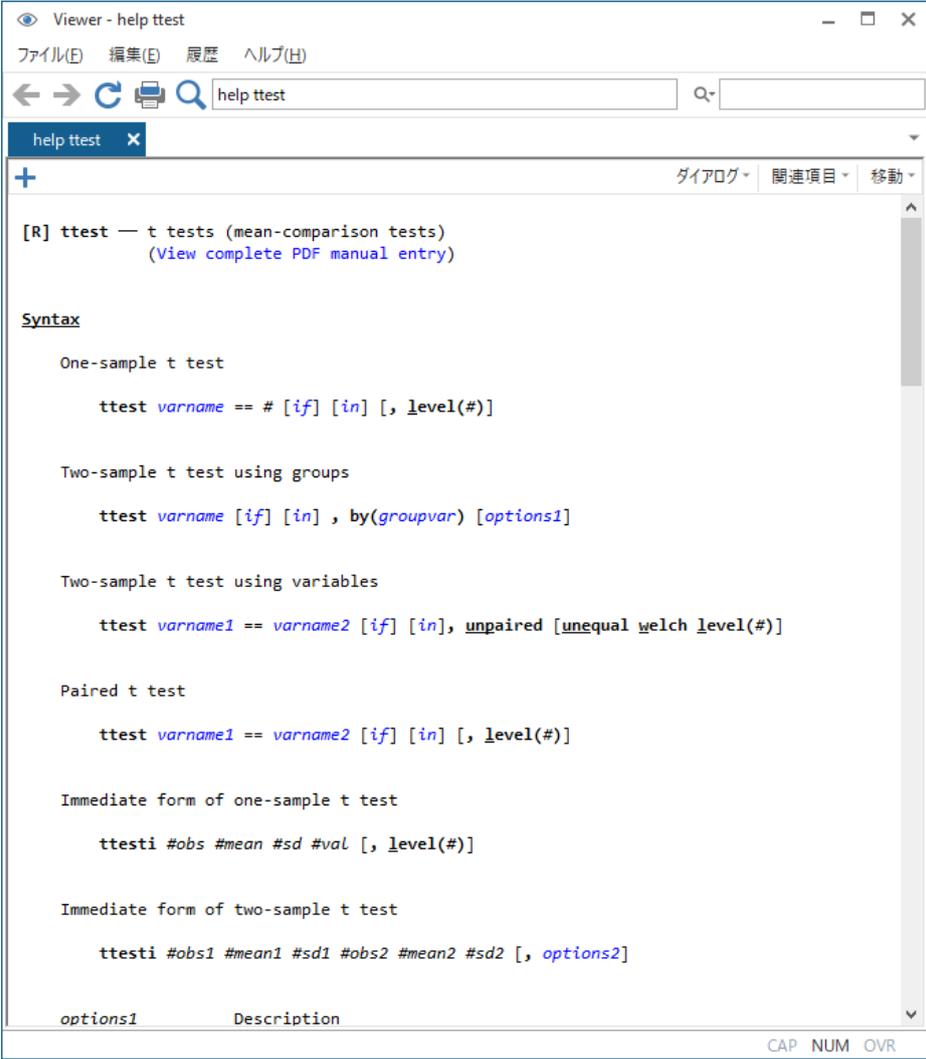
1つの変数を他のカテゴリ変数によって2つのグループに分けることで、ロング形式のデータのまま検定を行います。検定を行う前に、セクション3.4.1でワイド形式に変換したデータセットをロング形式に再変換します。`reshape` ダイアログで、“ロング形式に再変換(以前に変換した場合)”を選択します。

「t 検定」フィールドで「グループ別の二標本」を選択します。「変数名」に、検定する変数、ここでは `temp_mean` を選択し、「グループ変数」にカテゴリ変数 `city` を指定します。

第3章 基本的な統計操作

できます。または `help` コマンドで、特定の箇所を指定して呼び出すこともできます。
t 検定を行うコマンド `ttest` のマニュアルを呼び出してみましょう。次のようにして `help` コマンドでヘルプファイルを呼び出します。

```
. help ttest
```



```
[R] ttest — t tests (mean-comparison tests)
      (View complete PDF manual entry)

Syntax

One-sample t test

      ttest varname == # [if] [in] [, _level(#)]

Two-sample t test using groups

      ttest varname [if] [in] , by(groupvar) [options1]

Two-sample t test using variables

      ttest varname1 == varname2 [if] [in], unpaired [unequal welch _level(#)]

Paired t test

      ttest varname1 == varname2 [if] [in] [, _level(#)]

Immediate form of one-sample t test

      ttesti #obs #mean #sd #val [, _level(#)]

Immediate form of two-sample t test

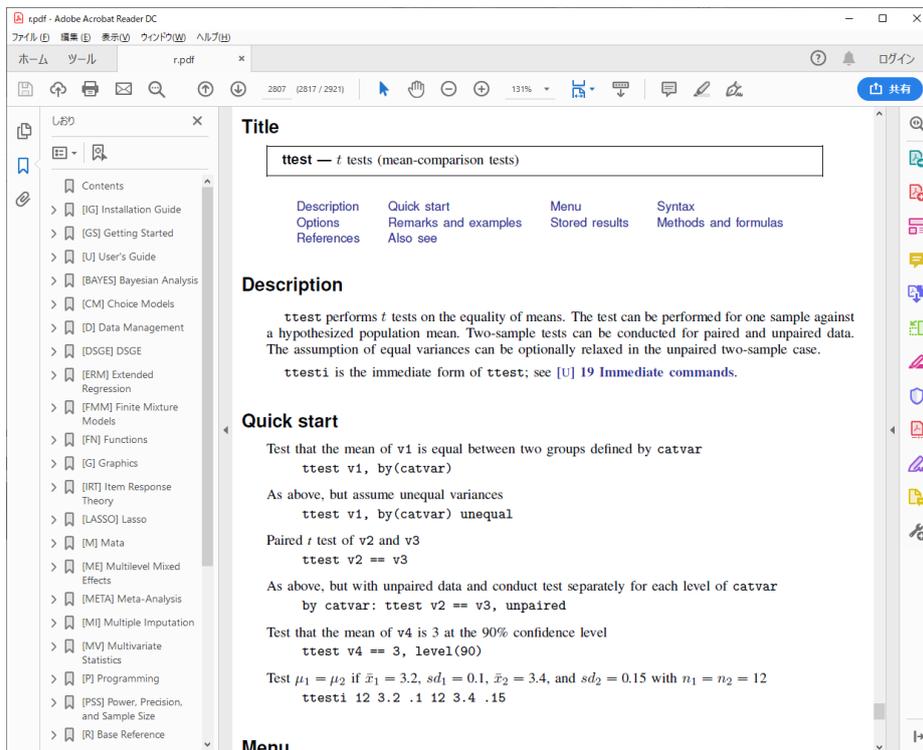
      ttesti #obs1 #mean1 #sd1 #obs2 #mean2 #sd2 [, options2]

options1      Description

CAP  NUM  OVR
```

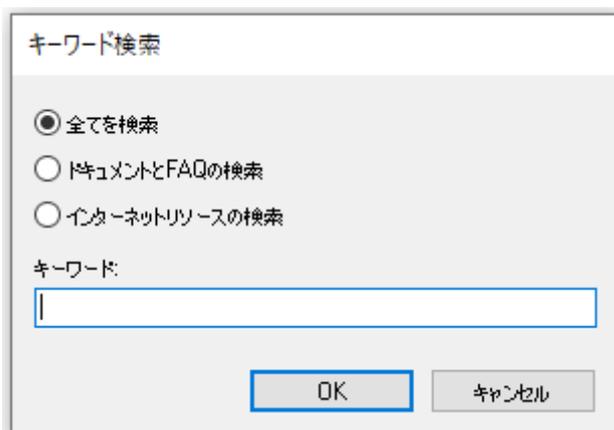
ヘルプファイルは PDF マニュアルを要約したものです。コマンドの概要、使い方、利用可能なオプションや実行例が記述されています。View complete manual entry をクリックして PDF マニュアルを開きます。

3.5 PDF マニュアルの活用



PDF マニュアルではコマンドの概要、クイックスタート、コマンドの構造、オプションの説明、実行例と解釈、技術的な解説などの項目が記載されています。

また、メインメニューの ヘルプ > 検索 を開き、キーワード検索によりヘルプファイルを検索できます。



第4章 線形回帰分析

Stataには推測統計に関連したコマンドが多数用意されています。本章では、代表的な推定コマンドのいくつかについて、その用法を説明するとともに、関連する推定事後機能についても主だったものを紹介します。

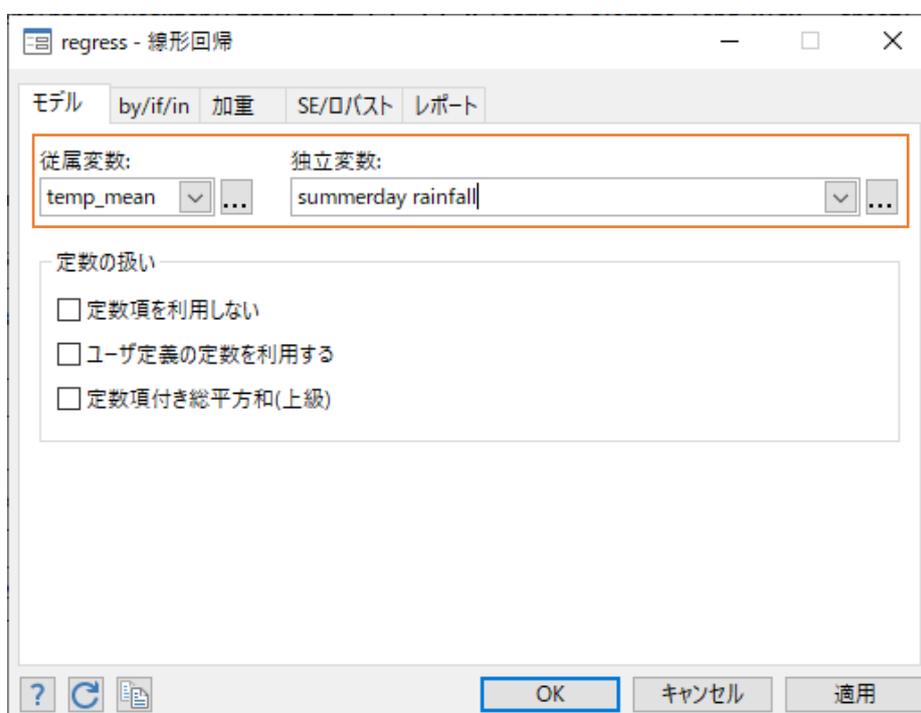
4.1 線形回帰モデル

本セクションでは、線形回帰機能を提供する `regress` コマンドについてその用法を説明します。用例としては、先のサンプルデータを用いて、毎月の平均気温を、夏日の観測回数と合計降水量で説明する、次の回帰モデルを想定します。

$$\text{temp mean} = \beta_0 + \beta_1 \text{summerday} + \beta_2 \text{rainfall} + \epsilon$$

データを読み込んだら、`regress` コマンドにより線形回帰を実行します。

1. 線形回帰機能を利用してフィットを行うには、**統計 > 線形モデル他 > 線形回帰**、を選択します。
2. 線形回帰ダイアログで、「従属変数」ボックスで従属(被説明)変数 `temp_mean` を、「独立変数」ボックスで独立(説明)変数 `summerday` と `rainfall` を指定します。
3. OK をクリックします。



これらをコマンドにまとめると、次の通りです。

4.1 線形回帰モデル

```
. regress temp_mean summerday rainfall
```

推定結果は次のように表示されます。

```
. regress temp_mean summerday rainfall
```

Source	SS	df	MS	Number of obs	=	96
Model	6291.59155	2	3145.79577	F(2, 93)	=	130.17
Residual	2247.60082	93	24.1677508	Prob > F	=	0.0000
				R-squared	=	0.7368
				Adj R-squared	=	0.7311
Total	8539.19237	95	89.8862355	Root MSE	=	4.9161

temp_mean	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
summerday	.6414873	.0447163	14.35	0.000	.5526896 .7302851
rainfall	.0084691	.0049311	1.72	0.089	-.0013232 .0182613
_cons	7.650989	.8201534	9.33	0.000	6.022327 9.279651

回帰分析のオプション⁵

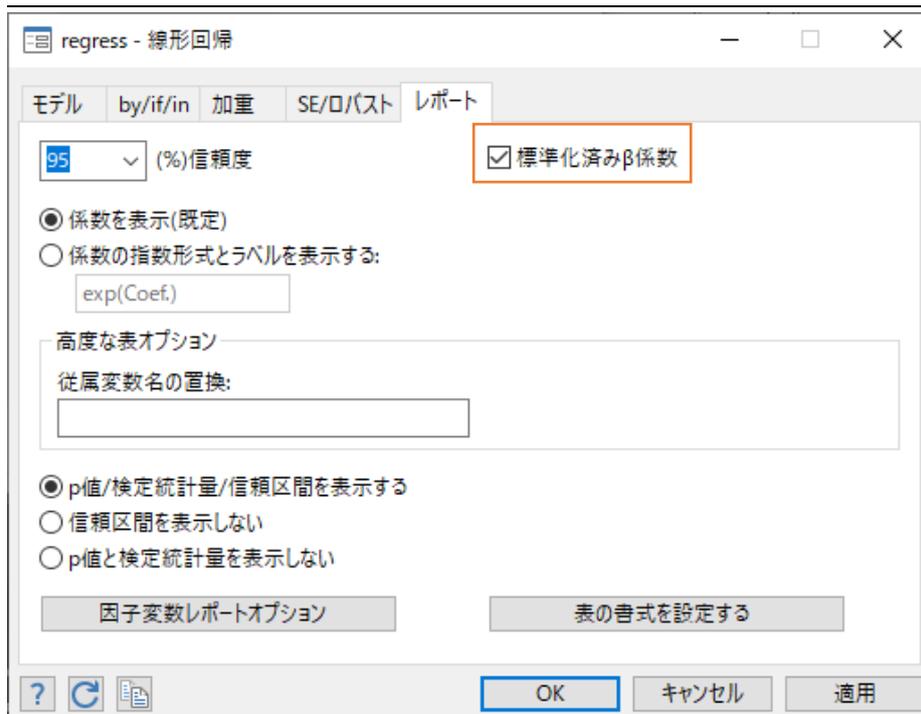
- 標準化済み β 係数

回帰結果に標準化済み β 係数を表示するには、`regress` ダイアログのレポートタブを開き、「標準化済み β 係数」にチェックを入れて OK をクリックします。コマンドオプションは `beta`、構文全体は次のようになります。

```
. regress temp_mean summerday rainfall, beta
```

⁵ その他のオプションとして、“定数項を使用しない”を選択すると、定数項を使用せずに、回帰計算を行います。このオプションのコマンドは `noconstant` です。

第4章 線形回帰分析



4.2 因子変数と交互作用項

Stata では数値がそれぞれに属性を表すようなカテゴリ変数に対して、それを構成する各因子について回帰することができます。

4.2.1 Stata 独自の因子変数

回帰分析にカテゴリ変数を用いる場合、`regress` ダイアログで従属変数に、`i.` プリフィックスを付加して入力します。⁶ここでは、都市を表す `citycode` を `i.citycode` と変更して、従属変数として投入します。コマンドでは次のようになります。

```
. regress temp_mean summerday rainfall i.citycode
```

回帰結果は、次のようにカテゴリ変数の各因子について計算が行われます。つまり、`i.` プリフィックスを使用すると、因子要素の1つを基準として係数を推定します

⁶ Stata ではこの `i.` 以外にもいくつかのプリフィックスコマンドがあります。時系列分析では、ラグを利用する `L.` プリフィックスや階差を利用する `D.` プリフィックスを用いて推定することができます。

4.2 因子変数と交互作用項

```
. regress temp_mean summerday rainfall i.cityenc
```

Source	SS	df	MS	Number of obs	=	91
Model	6128.03257	5	1225.60651	F(5, 85)	=	64.85
Residual	1606.53312	85	18.9003897	Prob > F	=	0.0000
				R-squared	=	0.7923
				Adj R-squared	=	0.7801
Total	7734.56569	90	85.9396188	Root MSE	=	4.3475

temp_mean	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
summerday	.521259	.0466413	11.18	0.000	.4285235	.6139945
rainfall	.0081675	.004406	1.85	0.067	-.0005927	.0169277
cityenc						
東京	1.07989	1.305058	0.83	0.410	-1.514914	3.674694
根室	-2.03947	1.355218	-1.50	0.136	-4.734006	.6550664
那覇	4.649439	1.338845	3.47	0.001	1.987457	7.311421
_cons	8.193636	1.130561	7.25	0.000	5.945779	10.44149

● ベースレベル

各カテゴリの係数はカテゴリ 1 を基準として計算されます。⁷この基準（ベースレベル）を他のカテゴリに変更することもできます。例えば、カテゴリ 3 を基準とする場合は、因子変数を `ib3.citycode`、と指定します。

```
. regress temp_mean summerday rainfall ib3.cityenc
```

Source	SS	df	MS	Number of obs	=	91
Model	6128.03257	5	1225.60651	F(5, 85)	=	64.85
Residual	1606.53312	85	18.9003897	Prob > F	=	0.0000
				R-squared	=	0.7923
				Adj R-squared	=	0.7801
Total	7734.56569	90	85.9396188	Root MSE	=	4.3475

temp_mean	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
summerday	.521259	.0466413	11.18	0.000	.4285235	.6139945
rainfall	.0081675	.004406	1.85	0.067	-.0005927	.0169277
cityenc						
新潟	2.03947	1.355218	1.50	0.136	-.6550664	4.734006
東京	3.11936	1.401203	2.23	0.029	.3333937	5.905326
那覇	6.688909	1.507276	4.44	0.000	3.69204	9.685778
_cons	6.154167	1.022411	6.02	0.000	4.121339	8.186994

4.2.2 交互作用

カテゴリ変数同士の交互作用を回帰式に追加するには、独立変数のテキストボックスで、# で変数同士を繋いで入力します。

セクション 4.1 の回帰式にカテゴリ変数 `citycode` と連続変数 `rainfall` の交差項を追加するには、`cityenc#c.rainfall` と入力します。連続変数をカテゴリ変数と交差させる場合

⁷ ここでは、1 としてエンコードされた新潟が基準となって計算が行われています。

第4章 線形回帰分析

は連続変数に `c.` プリフィックスを付けます。⁸

```
. regress temp_mean summerday rainfall i.cityenc cityenc#c.rainfall
```

```
. regress temp_mean summerday rainfall i.cityenc cityenc#c.rainfall
```

Source	SS	df	MS	Number of obs	=	91
Model	6529.67748	8	816.209685	F(8, 82)	=	55.55
Residual	1204.88821	82	14.6937587	Prob > F	=	0.0000
				R-squared	=	0.8442
				Adj R-squared	=	0.8290
Total	7734.56569	90	85.9396188	Root MSE	=	3.8332

temp_mean	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
summerday	.5540696	.041788	13.26	0.000	.4709399	.6371993
rainfall	-.010577	.0087697	-1.21	0.231	-.0280227	.0068687
cityenc						
東京	-2.288963	1.960941	-1.17	0.246	-6.1899	1.611973
根室	-7.740139	1.934478	-4.00	0.000	-11.58843	-3.891846
那覇	4.529966	2.05611	2.20	0.030	.4397089	8.620223
cityenc#c.rainfall						
東京	.0219059	.0107893	2.03	0.046	.0004424	.0433693
根室	.0500076	.0118272	4.23	0.000	.0264794	.0735358
那覇	-.0009638	.0112658	-0.09	0.932	-.023375	.0214475
_cons	10.7312	1.561378	6.87	0.000	7.625123	13.83728

4.3 限界効果

4.3.1 因子変数と交差項の限界効果

前セクションで推定した因子変数の限界効果を調べるには `margins` コマンドを使用します。メインメニューから **統計 > 推定後の分析**、を選択して推定後の分析セレクトを開きます。開いたダイアログの「限界効果、推定平均、相互作用、…」のツリーを開き、「カスタマイズした高次元分析」を選択します。

「マージンを算出する水準を示す因子変数」のドロップダウンリストで限界効果を算出するカテゴリ変数を選択します。OK をクリックして計算します。

ここでは、セクション 4.2.2 で回帰分析を行った結果を元に交差項の限界効果を求めます。

「マージンを算出する水準を示す因子変数」に、`citycode` を指定します。

⁸ カテゴリ変数同士を組み合わせて交差項とする場合は、プリフィックスを使用せず、#で変数同士を繋げて入力します。

4.3 限界効果

margins - 周辺平均、予測マージン、限界効果

メイン 位置(at) if/in/over Within コントラスト 対比較 加重 SE 上級設定 レポート

マージンを算出する水準を示す因子変数:
cityenc

因子項の無い場合、既定でグランドマージンを追加する

応答を選択する

既定の予測
 予測を指定する

予測:
作成...
編集
無効化
有効化

作成ボタンをクリックして予測を定義します。

推定パラメータを式で指定する

応答の限界効果

限界効果 $d(y)/d(x)$
 弾力性 $d(\ln y)/d(\ln x)$
 半弾力性 $d(y)/d(\ln x)$
 半弾力性 $d(\ln y)/d(x)$

変数:
 因子変数による水準を示す共変量を連続として扱う

OK キャンセル 適用

算出された限界効果は次のように表示されます。

```
. margins cityenc
```

```
Predictive margins          Number of obs   =       91
Model VCE      : OLS
```

```
Expression      : Linear prediction, predict()
```

	Delta-method				
	Margin	Std. Err.	t	P> t	[95% Conf. Interval]
cityenc					
新潟	15.05173	.8122204	18.53	0.000	13.43597 16.6675
東京	15.75641	.8198702	19.22	0.000	14.12543 17.38739
根室	14.1456	.9509276	14.88	0.000	12.2539 16.0373
那覇	19.44999	.8495196	22.90	0.000	17.76003 21.13996

4.3.2 限界効果のグラフ化

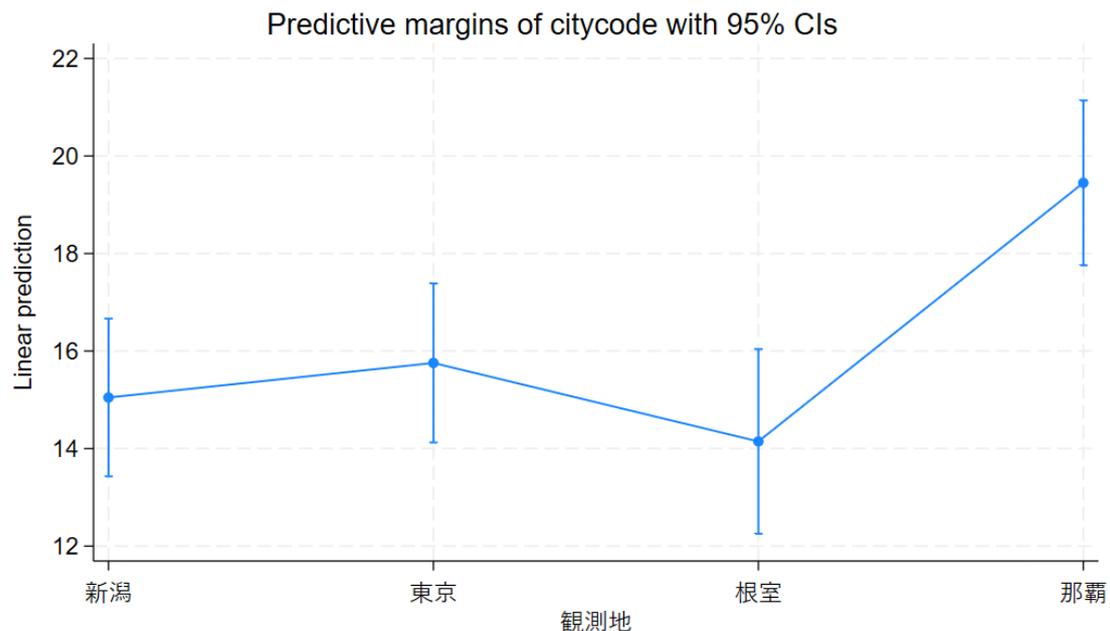
前セクション 4.3.1 で算出した限界効果は各因子を数値化して結果を表示しますが、Stata

第4章 線形回帰分析

ではこれらをグラフとして表すことができます。

メインメニューで、**統計 > 推定後の分析**、を選択します。「限界効果、推定平均、相互作用、...」ツリーを開き、「プロファイルプロットとマージン分析」を選択します。

開いたダイアログでOKをクリックします。または、特定の因子のみグラフにする場合、「x軸の変数」ドロップダウンリストで因子変数や交差項を選択して、OK をクリックします。



4.4 推定後の診断

`regress` などの推定コマンドを実行すると検定や診断のための推定事後機能が利用できません。Stata ではそのような推定事後のモデルや変数の診断を行うためのコマンド `estat` があります。

1. メインメニューの**統計 > 回帰診断 > モデル選択のための検定**など、を選択します。
2. 開いたダイアログの「結果および統計量:」ドロップダウンリストで実行する検定を選びます。

4.4.1 F 検定

分散分析に基づく F 検定の結果は `regress` コマンドの実行結果に含まれています。F 検定は回帰分析によって導き出されたすべての係数が 0 であることを、帰無仮説として検定を行っています。

4.4 推定後の診断

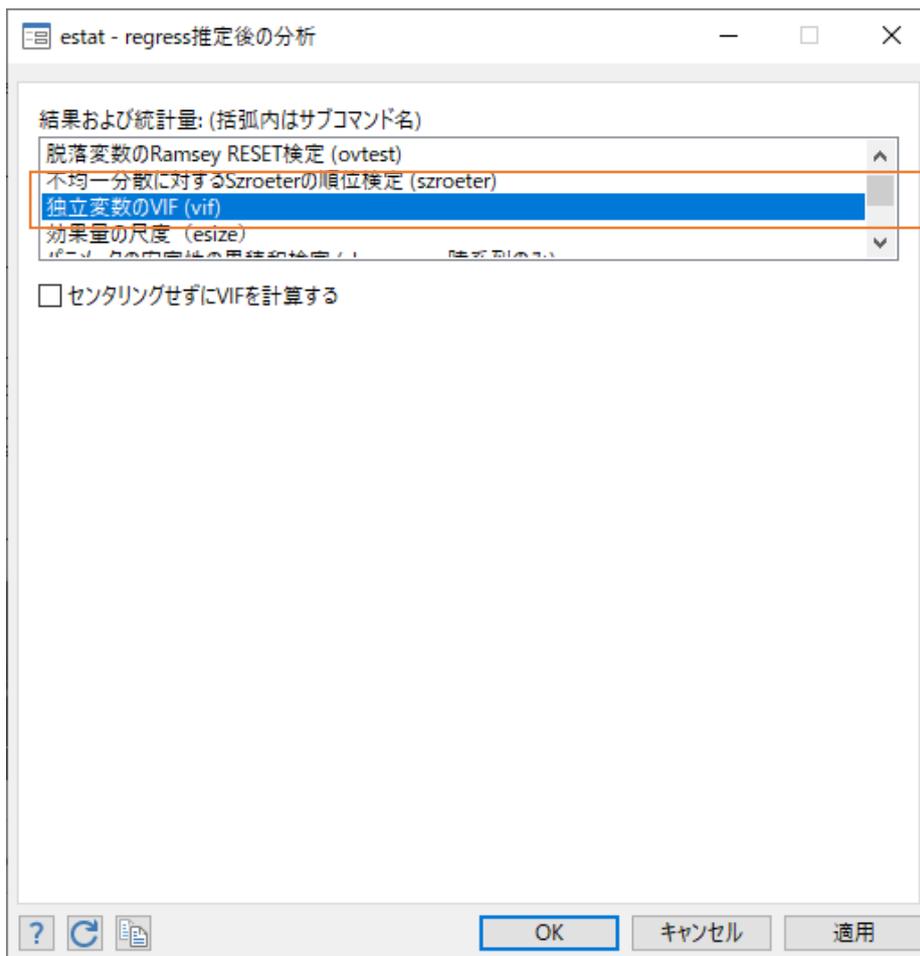
```
. regress temp_mean summerday rainfall
```

Source	SS	df	MS	Number of obs	=	96
Model	6291.59155	2	3145.79577	F(2, 93)	=	130.17
Residual	2247.60082	93	24.1677508	Prob > F	=	0.0000
				R-squared	=	0.7368
				Adj R-squared	=	0.7311
Total	8539.19237	95	89.8862355	Root MSE	=	4.9161

temp_mean	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
summerday	.6414873	.0447163	14.35	0.000	.5526896 .7302851
rainfall	.0084691	.0049311	1.72	0.089	-.0013232 .0182613
_cons	7.650989	.8201534	9.33	0.000	6.022327 9.279651

4.4.2 多重共線性

回帰分析に使用した説明変数の間に強い相関が疑われる場合、`estat` コマンドで `vif` サブコマンドを選択して、実行します。「regress 推定後の分析」ダイアログで、「独立変数の VIF」を選びます。



第4章 線形回帰分析

コマンドは次のようになります。

```
. estat vif
```

結果は次のように表示されます。

```
. estat vif
```

Variable	VIF	1/VIF
summerday	1.72	0.582759
rainfall	1.23	0.815393
fourseasons		
2	2.10	0.475732
3	1.81	0.553135
4	2.74	0.365020
fourseasons#		
snow		
1 1	1.33	0.752961
4 1	1.87	0.534276
Mean VIF	1.83	

4.4.3 標準的仮定

回帰分析の標準的仮定の検定について、`estat` コマンドでは誤差項の不均一分散と系列相関の有無を検定することもできます。不均一分散の検定には、サブコマンド `hettest` を使用して、分散が均一であることを帰無仮説として、Breusch-Pagan 検定および、Cook-Weisberg 検定を行います。コマンドは次のようになります。

```
. estat hettest
```

検定の結果は次のように表示されます。

```
. estat hettest
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of temp_mean

chi2(1)      =    10.23
Prob > chi2  =    0.0014
```

系列相関の検定には、サブコマンド `dwatson` を使用して、Durbin-Watson 検定を行います。この検定を行うには、事前に時系列の設定が必要です。⁹

⁹ 時系列の設定は、メインメニューの統計 > 時系列 > セットアップとユーティリティ > 時系列データ設定、を選択します。

4.5 予測

4.5 予測

一旦モデルがフィットされるとそのモデルに基づいて予測値を計算したり、実際の値との差を計算したりといった操作が可能となります。そのためのコマンドが `predict` です。

4.5.1 フィット値

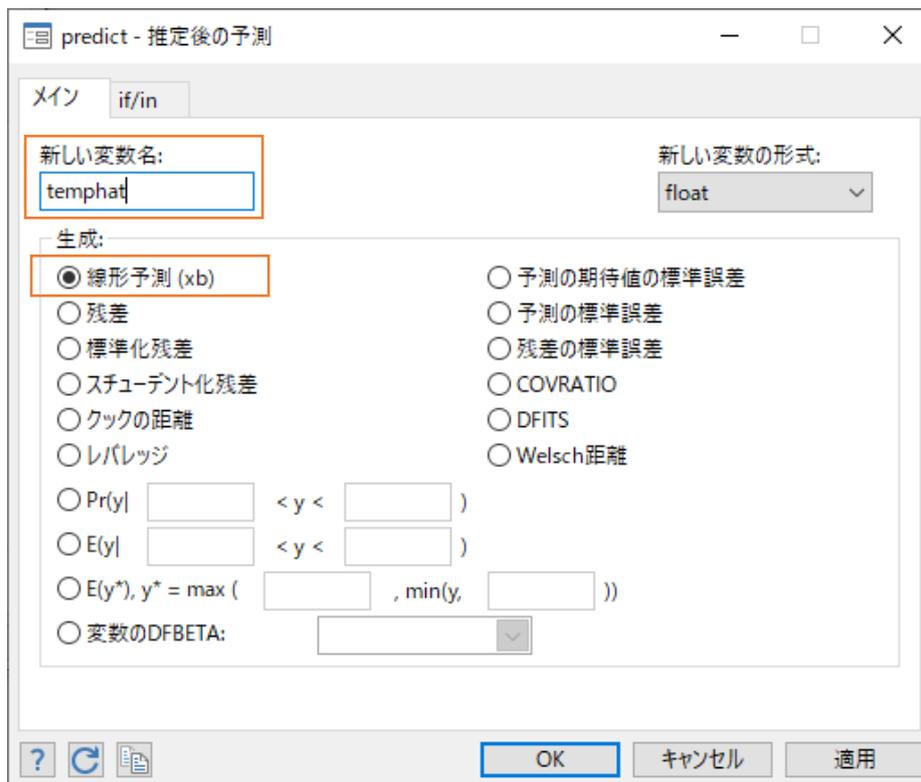
ここではセクション 4.1 でフィットした回帰モデルに基づき、`temp_mean` の予測値、 $\widehat{temp_mean}$ を算出します。`predict` コマンドは計算結果を新たな変数として格納するので、ここではその変数を `temphat` とします。

1. メインメニューの **統計 > 推定後の分析**、を選択し「推定後の分析セレクタ」を開きます。
2. 「予測」ツリーを開いて「予測、その標準誤差、レバレッジ統計量、距離統計量等」を選択します。
3. 「推定後の予測」ダイアログが開きます。メインタブの「生成：」セクションで「線形予測」が選択されていることを確認してください。
4. 「新しい変数名」に `temphat` を入力して、OK をクリックします。

以上の操作で、変数ウィンドウに `temphat` という変数が追加されていることを確認してください。コマンド操作では次のように行います。

```
. predict temphat, xb
```

第4章 線形回帰分析



4.5.2 残差の計算

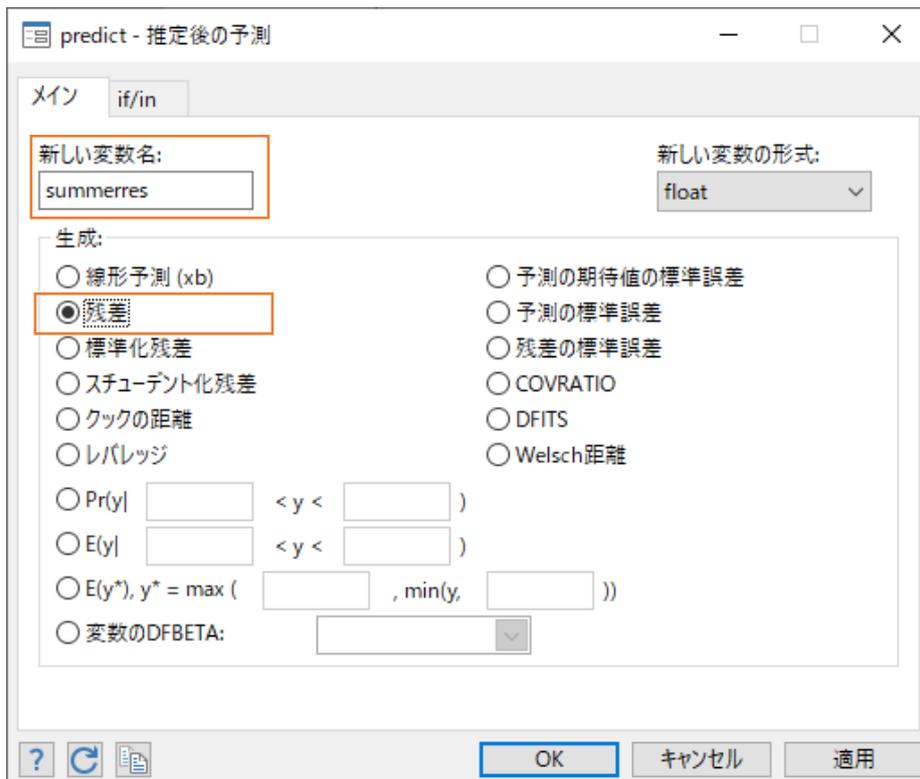
残差の算出も `predict` コマンドで行うことができます。結果を `summerres` という変数名で作成します。

1. `predict` ダイアログのメインタブを開き、「生成:」セクションで「残差」を選択します。
2. 「新しい変数名」に `summerres` を入力して、OK をクリックします。

コマンド操作では `predict` コマンドに `residuals` オプションを追加して、次の通りです。

```
. predict summerres, residuals
```

4.6 推定結果の表

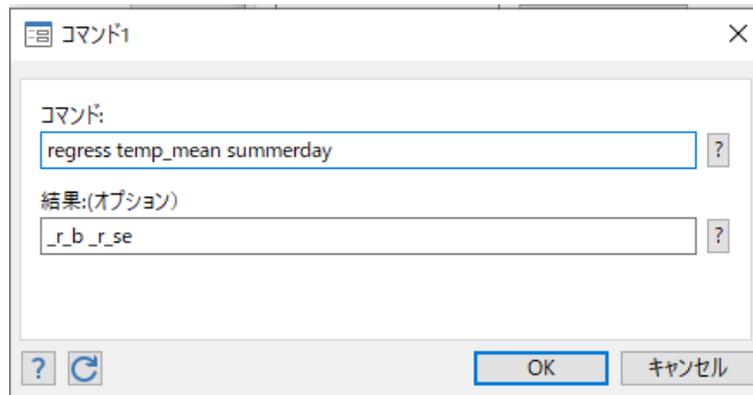


4.6 推定結果の表

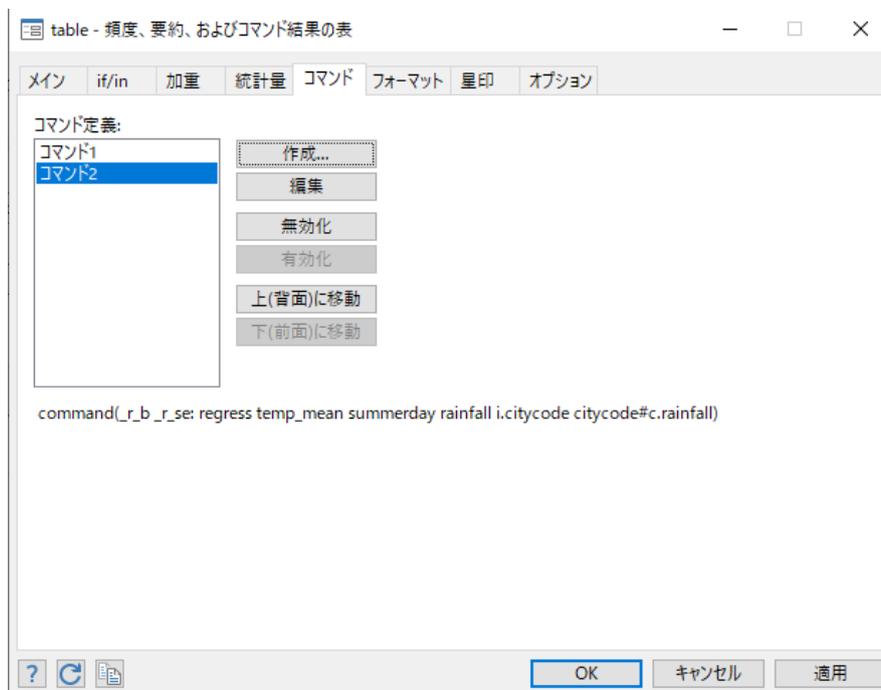
推定結果をもとに表を生成する `table` コマンドについて紹介します。複数の回帰モデルを推定した際に推定した係数や標準誤差を表にまとめて表示することができます。

表を作成するには、**統計(S) > 要約/表/検定 > 頻度、要約、およびコマンド結果の表** を開きます。表示されるダイアログのコマンドタブを開き、作成ボタンをクリックします。コマンド 1 というダイアログが開くので、コマンド欄に推定コマンド、結果欄に表示したい推定結果の識別子を入力します。推定された係数を表示するには `_r_b` と入力し、標準誤差を表示するには `_r_se` と入力します。その他にも z 値、t 値、信頼区間などを指定できます。各識別子については結果欄の横にある「？」をクリックして表示されるヘルプファイルをご覧ください。

第4章 線形回帰分析

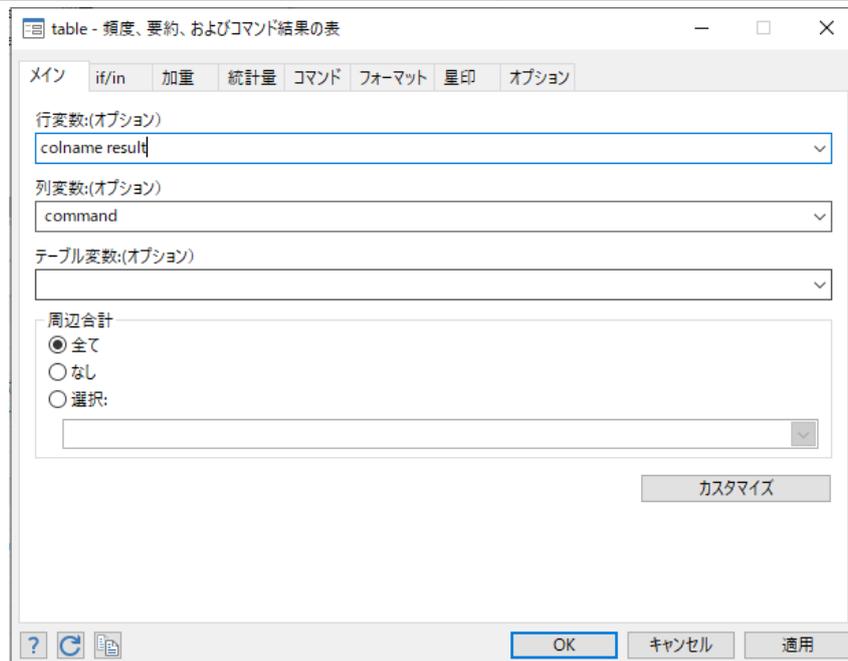


推定したモデル毎に上記の操作を繰り返し行い、コマンド定義を作成します。



メインタブに戻り、行変数に「colname result」、列変数に「command」と指定します。適用ボタンをクリックすると結果ウィンドウに推定結果の表が表示されます。

4.6 推定結果の表



コマンドでは下記のように記述します。

```
. table ( colname result ) ( command ) ( ), command(_r_b _r_se: regress  
temp_mean summerday ) command(_r_b _r_se: regress temp_mean summerday  
rainfall )
```

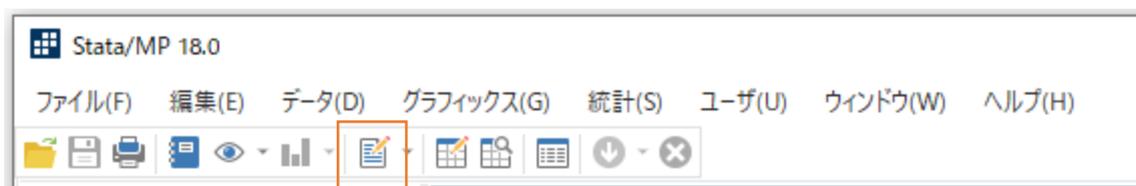

第5章 Do ファイルと ado ファイル

複雑な統計解析を行うときには一連のコマンド・操作を繰り返し実行させることがよくあります。そのような一連の操作をコマンドシーケンスとして Do ファイル化しておくことによって、処理を定型化することができます。本章では、これまで紹介した操作・コマンドを例にとって Do ファイル化のプロセスを説明します。

5.1 Do ファイルの作成

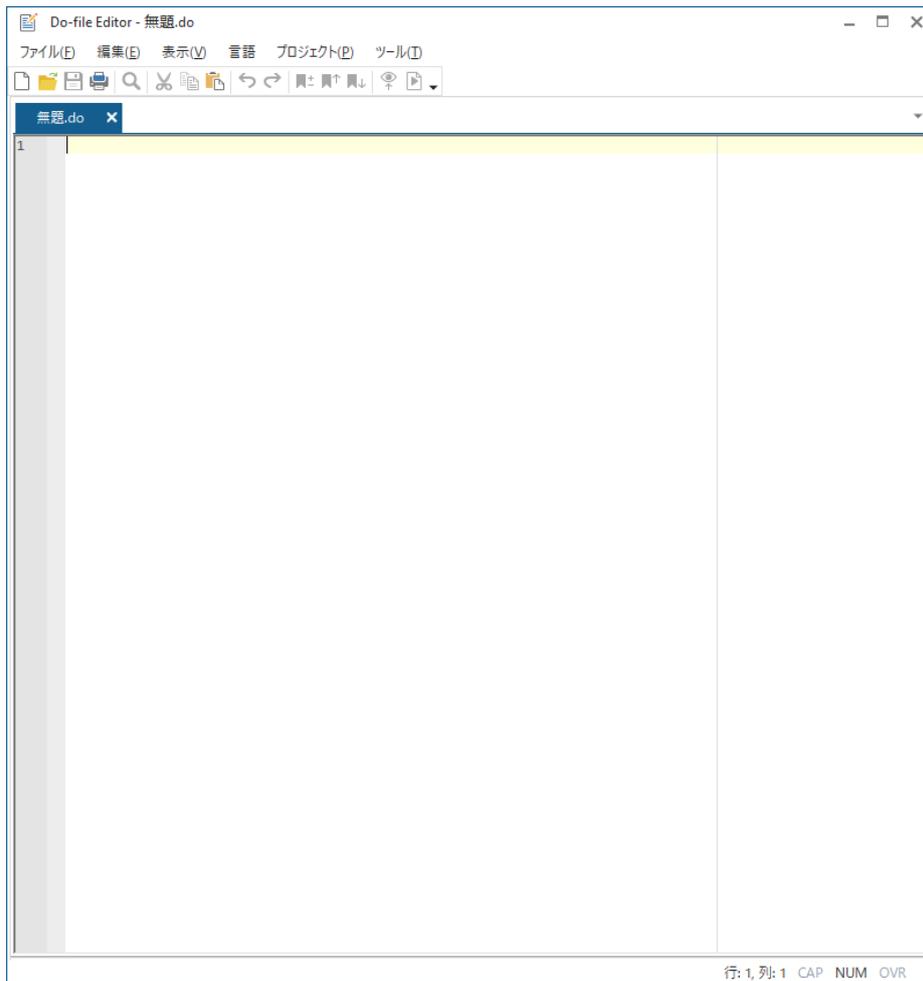
Do ファイルは Do ファイルエディタで作成します。Do ファイルエディタはツールバーの「新規 Do ファイルエディタ」、をクリックします。または、コマンド `doedit` を実行します。

・ `doedit`



次のように Do ファイルエディタが開きます。ここに定型化するコマンドを入力していきます。

第5章 Do ファイルと ado ファイル



このセクションでは、第1章で行ったデータの読み込みから、3章で扱った変数の検定までを Do ファイルとして作成します。

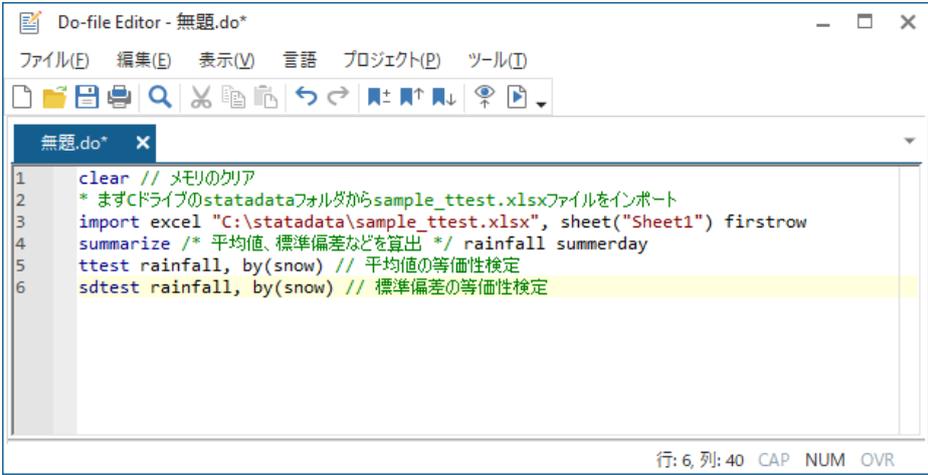
	コマンド	備考
1	<code>clear</code>	メモリのクリア
2	<code>import excel</code> <code>"C:\\$statadata\sample_ttest.xlsx",</code> <code>sheet("Sheet1") firstrow</code>	cドライブの statadata フォルダから sample_ttest.xlsx ファ イルをインポート
3	<code>summarize rainfall summerday</code>	平均値、標準偏差な どを算出
4	<code>ttest rainfall, by(snow)</code>	平均値の等価性検定
5	<code>sdtest rainfall, by(snow)</code>	標準偏差の等価性検 定

5.1 Do ファイルの作成

5.1.1 コメントの付与

Do ファイルは複数のコマンドをまとめたものですので、コマンドだけではどのような処理を行っているか分かりにくいことがあります。このような時のために、Do ファイルエディタではコメントを付ける機能があります。コメントは/*と*/でコメントを囲んで、コマンドと区別します。1行全体をコメントとする場合、*または//で行を開始します。コメントアウトされた部分はエディタ上では緑色で表示されます。¹⁰

上記のコマンド表の備考欄をコメントとして入力すると次のようになります。

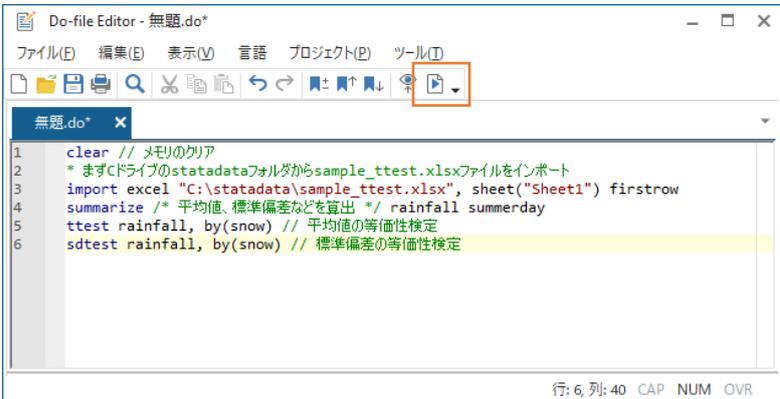


```

1 clear // メモリのクリア
2 * まずCドライブのstatadataフォルダからsample_ttest.xlsxファイルをインポート
3 import excel "C:\statadata\sample_ttest.xlsx", sheet("Sheet1") firstrow
4 summarize /* 平均値、標準偏差などを算出 */ rainfall summerday
5 ttest rainfall, by(snow) // 平均値の等価性検定
6 sdtest rainfall, by(snow) // 標準偏差の等価性検定

```

入力したコマンドが、それぞれが正しく機能するかを確認しましょう。エディタ上で確認したいコマンドを行単位で選択して、実行ボタンをクリックして、選択したコマンドを実行させます。



```

1 clear // メモリのクリア
2 * まずCドライブのstatadataフォルダからsample_ttest.xlsxファイルをインポート
3 import excel "C:\statadata\sample_ttest.xlsx", sheet("Sheet1") firstrow
4 summarize /* 平均値、標準偏差などを算出 */ rainfall summerday
5 ttest rainfall, by(snow) // 平均値の等価性検定
6 sdtest rainfall, by(snow) // 標準偏差の等価性検定

```

¹⁰ コマンドは青色、””で囲った部分は赤色になります。これらの配色は、メインメニューで編集 > ユーザ設定 > 一般的なユーザ設定、を選択すると変更できます。

第5章 Do ファイルと ado ファイル

5.1.2 演算子

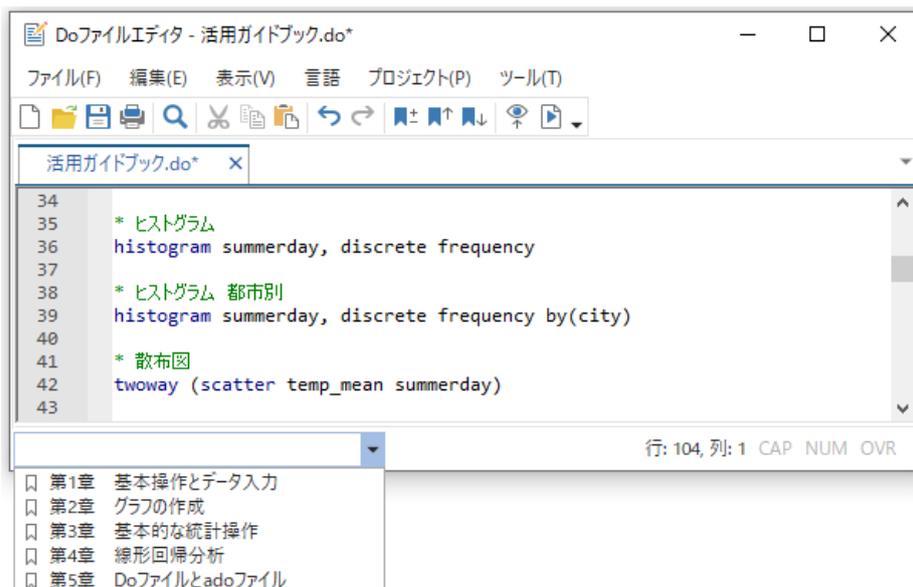
単独のコマンドや、Do ファイルでは演算子を使用することもありますので、ここにまとめておきます。

算術演算子		論理演算子		関係演算子	
+	加算	&	かつ	>	より大きい
-	減算		または	<	より小さい
*	乗算	!	否定	>=	以上
/	除算	~	否定	<=	以下
^	累乗			==	等しい
				!=	不等
				~=	不等

5.1.3 ブックマークとナビゲーションコントロール

ツールバーのブックマークボタンをクリックすると選択している行にブックマークが追加されます。Do ファイル内の処理の区切りのよいところでブックマークを挿入して整理するのに役立ててください。

Do ファイルの画面の左下にあるプルダウンメニューをクリックするとブックマークの一覧が表示され、クリックすると当該箇所へ移動できます。



5.2 Do ファイルの実行

5.2 Do ファイルの実行

全てのコマンドを入力できたら、Do ファイルに名前を付けて保存します。Do ファイルエディタ上でファイル > 名前を付けて保存、を選択します。現在のワーキングディレクトリに **test.do** という名前で保存します。

保存した Do ファイルを実行します。Stata のメインメニューで **ファイル > Do ファイルの実行、**を選択します。開いたダイアログで先ほど保存した Do ファイルを選択して、OK をクリックします。コマンドでは次のように行います。

```
. do test
```

Do ファイルに記入したコマンドが実行されます。最後には Do ファイルの終了を意味する **end of do-file** が表示されます。

```
Ha: diff < 0          Ha: diff != 0          Ha: diff > 0
Pr(T < t) = 0.9761    Pr(|T| > |t|) = 0.0477    Pr(T > t) = 0.0239
```

```
. sdtest rainfall, by(snow) // 標準偏差の等価性検定
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	75	145.66	13.23171	114.5899	119.2953	172.0247
1	21	92.19048	17.16213	78.64675	56.3909	127.99
combined	96	133.9635	11.19159	109.6547	111.7454	156.1816

```
ratio = sd(0) / sd(1)          f = 2.1229
Ho: ratio = 1                  degrees of freedom = 74, 20
```

```
Ha: ratio < 1          Ha: ratio != 1          Ha: ratio > 1
Pr(F < f) = 0.9697    2*Pr(F > f) = 0.0605    Pr(F > f) = 0.0303
```

```
.
end of do-file
```

5.3 Do ファイルの運用方法

Stata 操作画面上の履歴ウィンドウ内には実行されたコマンドの履歴が表示されています。SHIFT キーを押しながらコピーしたいコマンドの範囲を特定した後、右クリック > コピーと操作して、クリップボードにその内容をコピーします。

または、履歴ウィンドウでコマンドを選択して、右クリック > **選択範囲を Do ファイルエディタへ送る、**をクリックすることで、選択したコマンドを新しい Do ファイルエディタに送ることができます。

第5章 Do ファイルと ado ファイル

#	コマンド	_rc
1	import excel "C:\Users\nagai\De...	
2	summarize rainfall summerday	
3	ttest rainfall, by(snow)	
4	sdtest rainfall, by(snow)	

1	21	92.1
combined	96	133.
diff		53.4

diff = mean(0) - mean(
Ho: diff = 0

切り取り(I)
コピー(C)
削除(D)
すべて選択(A)
すべてクリア
選択範囲を実行
選択範囲をDoファイルエディタへ送る
すべて保存...
選択範囲を保存...
フォント...

ワーキングディレクトリ以外に保存されている Do ファイルも do コマンドで同じように実行できます。ファイル > do ファイルの実行、を選択して開いたダイアログで Do ファイルを選ぶ、または、コマンド操作では Do ファイルの位置(ファイルパス)をダブルクォーテーション(” ”)で指定します。

```
. do "D:\StataExample\test.do"
```

5.4 ado の探し方と試し方

ado ファイルは世界中の Stata ユーザが作成した便利なコマンドです。ウェブ上で公開されている ado ファイルをインストールすると、Stata で利用できるコマンドが追加されます。利用可能なコマンドが増えると同時に、普通はその利用方法を示す情報も提供されます。ここでは、ado ファイル利用者としての立場にたって、その利用方法を紹介します。

5.5 ado ファイルの検索とインストール

キーワード「regression output word」を用いて、前セクションで行った回帰分析の結果をワードファイルに出力する ado ファイルを検索してみましょう。ウェブ上の情報を検索するにはコマンド findit を利用します。コマンド欄に次のように入力します。

```
. findit regression output word
```

5.5 ado ファイルの検索とインストール

```

Viewer - search regression output word, all
ファイル(F) 編集(E) 履歴 ヘルプ(H)
search regression output word, all
search regression output word...
+
Search of web resources from Stata and other users
(contacting http://www.stata.com)
6 packages found (Stata Journal and STB listed first)
-----
asdoc from http://fmwww.bc.edu/RePEc/bocode/a
'ASDOC': module to create high-quality tables in MS Word from Stata output
/ asdoc sends Stata output to Word / RTF format. asdoc creates /
high-quality, publication-ready tables from various Stata / commands such
as summarize, correlate, pwcorr, tab1, tab2, / tabulate1, tabulate2,
estout from http://fmwww.bc.edu/RePEc/bocode/e
'ESTOUT': module to make regression tables / estout produces a table of
regression results from one or / several models for use with spreadsheets,
LaTeX, HTML, or a / word-processor table. eststo stores a quick copy of
the active / estimation results for later tabulation. esttab is a wrapper
estout1 from http://fmwww.bc.edu/RePEc/bocode/e
'ESTOUT1': module to export estimation results from estimates table /
-estout1- is a wrapper for -estimates table- and produces a / table of
regression results for use with spreadsheets, TeX, / HTML, or a
word-processor table. In addition, -estout1- / overcomes some of the
outreg2 from http://fmwww.bc.edu/RePEc/bocode/o
'OUTREG2': module to arrange regression outputs into an illustrative table
/ outreg2 provides a fast and easy way to produce an illustrative / table
of regression outputs. The regression outputs are produced / piecemeal and
are difficult to compare without some type of / rearrangement. outreg2
reformat from http://fmwww.bc.edu/RePEc/bocode/r
'REFORMAT': module to reformat regression output / The output from the
last regression command is re-displayed in a / more readable format using
variable and value labels for clarity. / The columns to be displayed can
be controlled by the user and / extra options to show the number of
reg2docx from http://fmwww.bc.edu/RePEc/bocode/r
CAP NUM OVR

```

ここで表示されている「web resources from Stata and other users」とあるのが ado ファイルです。例えば、4 番目のリンク先の情報を見てみましょう。

outreg2 from <http://fmwww.bc.edu/RePEc/bocode/o>

このリンクを開くと、回帰分析の結果をワードやエクセルに変換するファイルが提供されていることがわかります。

第5章 Do ファイルと ado ファイル

```
Viewer - net describe outreg2, from(http://fmwww.bc.edu/RePEc/bocode/o)
net describe outreg2, from(http://fmwww.bc.edu/RePEc/bococ

package outreg2 from http://fmwww.bc.edu/RePEc/bocode/o

TITLE
'OUTREG2': module to arrange regression outputs into an illustrative table

DESCRIPTION/AUTHOR(S)

outreg2 provides a fast and easy way to produce an illustrative
table of regression outputs. The regression outputs are produced
piecemeal and are difficult to compare without some type of
rearrangement. outreg2 automates this process by concatenating
the successive regression outputs in a vertical format. The
resulting table is saved to the disk in ASCII format, which can
be read by other programs. outreg2 thus facilitates the
conversion of regression outputs to a standard format suitable
for inclusion in a scholarly publication. The functionality of
outreg2 is based on the earlier package outreg, by John Luke
Gallup. Unlike outreg, outreg2 is capable of writing LaTeX-format
tables, as well as ASCII, MS Word and MS Excel.

KW: regression
KW: output
KW: tables
KW: tab-delimited output
KW: LaTeX
KW: Word
KW: Excel

Requires: Stata version 7

Distribution-Date: 20140817

Author: Roy Wada
Support: email roywada@hotmail.com

CAP NUM OVR
```

ado ファイルの解説、作者と問い合わせ先が記載されています。最後の「click here to install」をクリックすると、ado ファイルとヘルプファイルがインストールできます。インストールされるファイルは「INSTALLATION FILES」に一覧表示されています。

5.5 ado ファイルの検索とインストール

```
Requires: Stata version 7
Distribution-Date: 20140817
Author: Roy Wada
Support: email roywada@hotmail.com
```

INSTALLATION FILES

```
outreg2.ado
outreg2_prf.ado
outreg2.hlp
../s/shellout.ado
../s/shellout.hlp
../s/seeout.ado
../s/seeout.hlp
```

[\(click here to install\)](#)

[\(click here to return to the previous screen\)](#)

公開されている ado ファイルの名前がわかっている場合、`ssc install` コマンドで直接インストールできる場合もあります。outreg2 の場合は次のようになります。

```
. ssc install outreg2
```

インストールした ado ファイルは Stata フォルダとは異なる独自のフォルダに格納されます。次のコマンドを実行して ado ファイルの保存先を確認しましょう。

```
. adopath
```

Stata は 7 つのフォルダ情報を表示しています。上記の操作によってインターネット経由でインストールされた ado ファイルは [5] (PLUS) に格納されます。

```
. adopath
[1] (BASE)      "C:\Program Files\Stata16\ado\base/"
[2] (SITE)     "C:\Program Files\Stata16\ado\site/"
[3]           "."
[4] (PERSONAL) "c:\ado\personal/"
[5] (PLUS)    "c:\ado\plus/"
[6] (OLDPLACE) "c:\ado/"
```

※シングルライセンス Windows OS の例 ¹¹

5.5.1 ado ファイルの実行

ado ファイルがインストールされたら、outreg2 コマンドを試してみましょう。outreg2 は、セクション 4.3 で紹介した限界効果を求める margins のような推定後コマンドと同様に、最後に行った推定結果を出力します。ほとんどの ado ファイルはダイアログボックスを持たず、コマンドのみで操作します。

ここでは、outreg2 を利用して、セクション 4.1 で行った回帰分析の結果を word ファイル

¹¹ adopath コマンドで表示されるフォルダ構成は Stata インストール時の設定およびシングル/ライセンスライセンスによって異なります。

第5章 Do ファイルと ado ファイル

として出力します。まずは、セクション 4.1 と同じように回帰分析を行います。その結果を作業中のフォルダに `myreg.doc` というワードファイルで保存します。コマンドは次の通りです。

```
. outreg2 using myreg.doc, replace ctitle(Model 1) label
```

ここでは、`ctitle` オプションを利用して回帰係数の列に、`Model 1` という列タイトルを挿入します。さらに、`label` オプションで変数のラベルを出力します。実行すると、作業中のディレクトリに `myreg` という名前でワードファイルが作成されます

VARIABLES	(1) Model 1
summerday	0.628*** (0.0441)
rainfall	0.00728 (0.00481)
Constant	8.198*** (0.826)
Observations	91
R-squared	0.740

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.6 ado ファイルの更新

ado ファイルは世界中のユーザによって作成されているため、不定期に更新されます。この更新状況を確認するコマンドが `ado update` です。`ado update` を実行すると、インストールされている全ての ado ファイルの更新状況を確認して表示します。

```
. adoupdate
(note: ado update updates community-contributed files; type -update- to check for updates to official Stata)
```

```
Checking status of installed packages...
```

- ```
[1] probtab1 at https://stats.idre.ucla.edu/stat/stata/ado/teach:
 installed package is up to date

[2] mibeta at http://www.stata.com/users/ymarchenko:
 installed package is up to date
```

## 5.6 ado ファイルの更新

---

アップデートが公開されている ado ファイルがあれば、最後にまとめて表示されます。最後に表示された **adoupdate, update** をクリックまたは、コマンド欄で **adoupdate, update** コマンドを実行して、更新を適用します。

Packages that need to be updated are...

```
[27] estout -- 'ESTOUT': module to make regression tables
[34] plotmatrix -- 'PLOTMATRIX': module to plot values of a matrix as different coloured
 blocks
[88] asrol -- 'ASROL': module to generate rolling window / groups descriptive
 statistics
[89] cibarc -- 'CIBAR': module to plot bar graphs and confidence intervals over groups
[91] asdoc -- 'ASDOC': module to create high-quality tables in MS Word from Stata
 output
```

Recommendation: type  
**ado update, update**

## 分析機能例題集

Stata に搭載されているマニュアルや開発元 StataCorp のウェブサイトを元に、よく使用される解析について例題を交えて解説しています。分析の考え方やコマンドの用法を理解するのに最適です。

現在、以下のトピックについて解説しています。例題集は順次追加されます。

- 潜在クラスモデル
  - ダイナミックレポート
  - 差分の差分モデル
  - クラスター分析（階層型）
  - 主成分分析
  - クラスター分析（非階層型）
  - ガルブレイスプロット
  - Leave-one-out メタ分析
  - 表作成機能
  - Lasso
  - 有限混合モデル
  - DSGE モデル
  - 空間自己回帰モデル
- ほか