Stata+α

「新型コロナウィルスの感染データを Stata で入手する（国内編）」

前回、「新型コロナウィルスの感染データを Stata で入手する」ではジョンズ・ホプキンス大学が公開しているは各国のコロナウィルス感染者数を Stata でアクセスし、Stata 形式に変換して整理しました。今回は、「COVID-19 Japan 新型コロナウイルス対策ダッシュボード」で公開されている都道府県別の感染情報にアクセスし、Stata 形式に変換します。

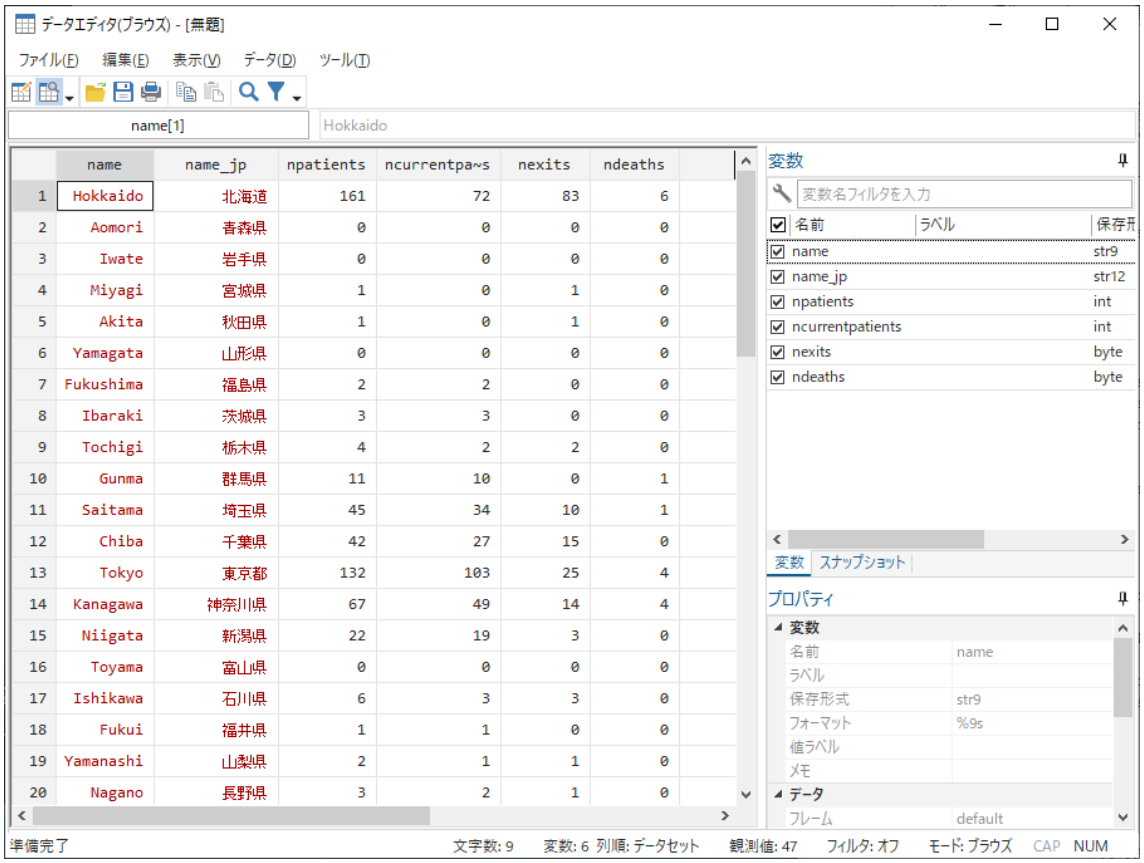
1. データのインポート

「COVID-19 Japan 新型コロナウイルス対策ダッシュボード」(https://www.stopcovid19.jp/)は、前回紹介したジョンズ・ホプキンス大学と同じように https://www.stopcovid19.jp/data/covid19japan/2020-03-22.csv という形で csv ファイルを公開しています。各ファイルは、YYYY-MM-DD.csv となっています。1 日分の csv ファイルを `import delimited` コマンドで Stata に取り込んでみましょう。都道府県名は漢字で入力されているため、`encoding()`オプションで、文字コードを指定します。

```
import delimited "https://www.stopcovid19.jp/data/covid19japan/2020-03-22.csv", encoding(UTF-8)
```

データエディタには次のようなデータセットがインポートされます。

LightStone® 株式会社ライトストーン



	name	name_jp	npatients	ncurrentpa~s	nexits	ndeaths
1	Hokkaido	北海道	161	72	83	6
2	Aomori	青森県	0	0	0	0
3	Iwate	岩手県	0	0	0	0
4	Miyagi	宮城県	1	0	1	0
5	Akita	秋田県	1	0	1	0
6	Yamagata	山形県	0	0	0	0
7	Fukushima	福島県	2	2	0	0
8	Ibaraki	茨城県	3	3	0	0
9	Tochigi	栃木県	4	2	2	0
10	Gunma	群馬県	11	10	0	1
11	Saitama	埼玉県	45	34	10	1
12	Chiba	千葉県	42	27	15	0
13	Tokyo	東京都	132	103	25	4
14	Kanagawa	神奈川県	67	49	14	4
15	Niigata	新潟県	22	19	3	0
16	Toyama	富山県	0	0	0	0
17	Ishikawa	石川県	6	3	3	0
18	Fukui	福井県	1	1	0	0
19	Yamanashi	山梨県	2	1	1	0
20	Nagano	長野県	3	2	1	0

2. ファイルの一括ダウンロード

公開されている csv ファイルを一括でダウンロードするには、import delimited を何度も繰り返します。この操作は、前回と同様です。繰り返しには同じく forvalues と local コマンドを利用します。次のコマンドを do ファイルエディタに貼り付けて実行しましょう。

```
local URL = "https://www.stopcovid19.jp/data/covid19japan/"
forvalues month = 1/12 {
        forvalues day = 1/31 {
                local month = string(`month', "%02.0f")
                local day = string(`day', "%02.0f")
                local year = "2020"
                local today = "`year'-`month'-`day'"
                local FileName = "`URL'`today'.csv"
                clear
                capture import delimited "`FileName'", encoding(UTF-8)
```


LightStone® 株式会社ライトストーン

```
11                if _rc == 0 {
12                        generate date = "`today'"
13                        capture confirm variable ndeath
14                        if _rc == 0 {
15                            destring ndeath, replace force
16                        }
17                        capture confirm variable
nheavycurrentpatients
18                        if _rc == 0 {
19                            destring nheavycurrentpatients, replace
force
20                        }
21                }
22        capture save "`today'", replace
23    }
}
```

基本的には、前回の「3. 複数ファイルを一括ダウンロード」と同様ですが、大きく異なっている点は、11-21 行目まで一連のコマンドです。11 行目以降は if コマンドの括弧で囲まれています。10 行目の capture import delimited コマンドは、ファイルのインポートが成功すれば、システム変数_rc に 0 を代入し、エラーが発生すると該当のエラーコードを代入します。つまり csv ファイルの読み込みが成功した場合のみ、12-21 行目までが実行されます。

まず、ダウンロードしたデータセットにはファイル名にしか調査した日付がありませんので、12 行目で、generate コマンドを使用し日付を新しい変数として作成します。

13-19 行では、5 月 8 日付けデータセットから新たに集計が開始された 2 変数、ndeath と nheavycurrentpatients が文字列変数として読み込まれてしまう問題に対処しています。13 行目の capture confirm ndeath コマンドで読み込んだデータセットに変数 ndeath があれば、15 行目の destring コマンドで数値変数に変換します。17-19 行も同様の処理を変数 nheavycurrentpatients に対して行います。

このようにして、変数を整えたら、22 行目の capture save コマンドで YYYY-MM-DD.dta というファイル名の Stata 形式ファイルとして保存します。

3. ファイルの統合

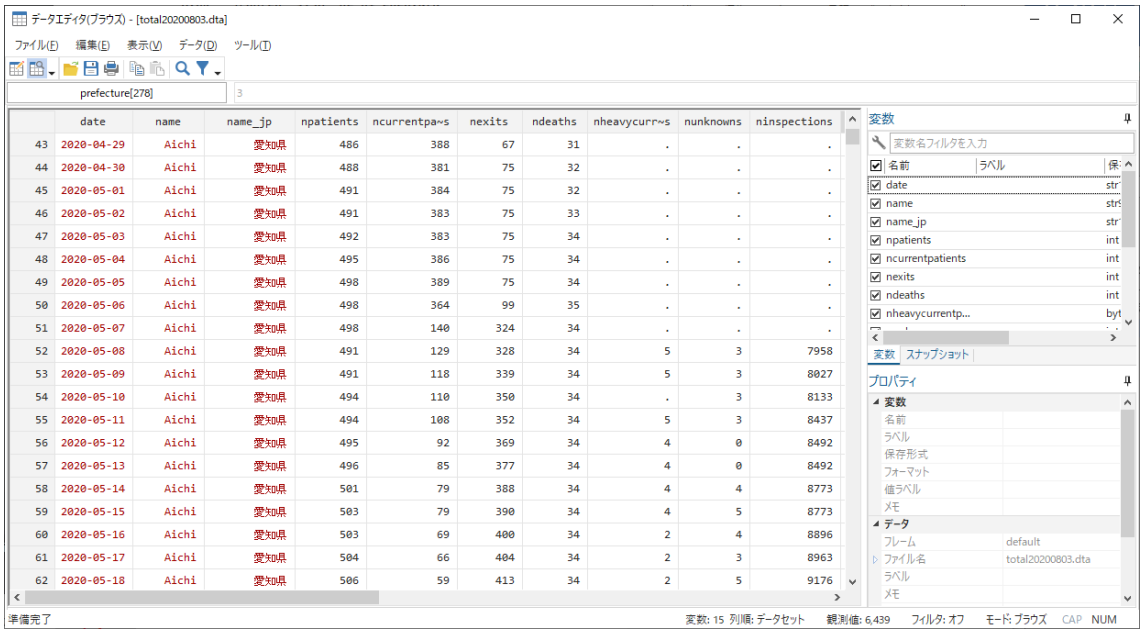
ダウンロードしたファイルを append コマンドで 1 つの dta ファイルに統合します。append

LightStone® 株式会社ライトストーン

も forvalues コマンドと組み合わせて繰り返します。こちらも前回の「4. データをまとめる」で使用したものをほぼ同様です。8 行目の年月日の並び方が異なっていることにご注意ください。

```
1  clear
2  forvalues month = 1/12 {
3     forvalues day = 1/31 {
4        local month = string(`month', "%02.0f")
5        local day = string(`day', "%02.0f")
6        local year = "2020"
7        local today = "`year'-`month'-`day'"
8        capture append using "`today'"
9     }
10 }
```

統合したデータセットは次のようになります。



4. データを整える

3.で統合したデータセットを確認すると、都道府県名の変数 name に入力ミスが見つかります。これは replace コマンドで修正します。

```
1  replace name = "Ehime" if name == "Eihime"
```


LightStone® 株式会社ライトストーン

このデータセットは、時間変数 date と ID 変数 name または name_jp を持つパネル形式です。変数 date と name を文字列から数値に変換して、パネルデータとして宣言しましょう。変数 name または name_jp のような文字列変数に数値を割り振って数値(カテゴリ)変数を作成するには、encode コマンドを使用します。ここでは、変数 name から新しい変数 prefecture を作成しています。

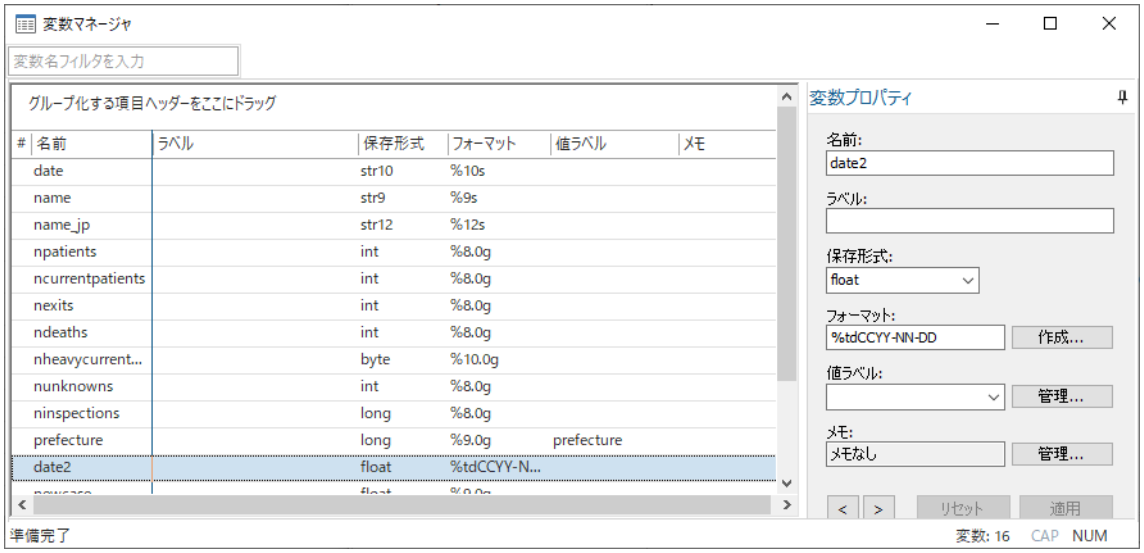
```
encode name, generate(prefecture)
```

続いて、文字列となっている変数 date から generate コマンドで数値変数 date2 を新規作成します。数値変数 date2 はそのままでは人間が認識できる形式ではないので、2 行目の format コマンドで表示形式を変更します。

```
generate date2 = daily(date, "YMD")
format %tdCCYY-NN-DD date2
```

表示形式の変更は、変数マネージャから行うこともできます。変数マネージャを開き、date2 を選択して「フォーマット」の「作成」ボタンをクリックします。「書式作成」ダイアログでフォーマットを変更したら、OK をクリックして閉じ、変数マネージャの「適用」をクリックします。

変数マネージャの画面は次のようになります。



時間変数と ID 変数の準備ができたら、xtset コマンドでパネルデータであることを Stata

LightStone® 株式会社ライトストーン

に認識させます。

```
xtset prefecture date2
```

結果ウィンドウには次のように、データセットの構造が表示されます。パネル ID は prefecture、期間は 3 月 18 日から 8 月 1 日の 1 日毎のデータです。

```
. xtset prefecture date2
       panel variable:  prefecture (strongly balanced)
        time variable:  date2, 2020-03-18 to 2020-08-01
                delta:  1 day
```

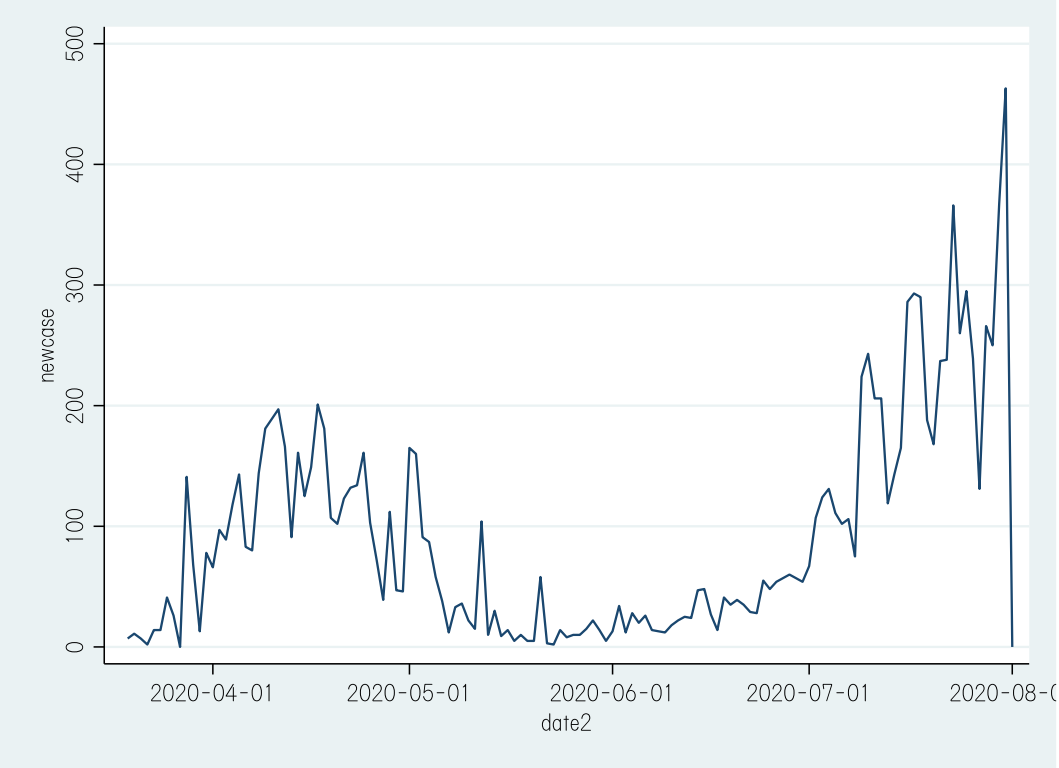
変数 npatients は累積感染者数です。1 日の新規感染者数を求めるには前日の npatients との差を計算します。Stata では差分を d.演算子で簡単に求めることできます。d.演算子と generate コマンドで 1 日あたりの新規感染者数 newcase を作成します。

```
generate newcase = d.npatients
```

新規感染者数は曜日によって変動があることがわかっています。これをグラフで確認してみましょう。東京都の新規感染者数を折れ線グラフにします、if オプションで対象を制限しています。

```
twoway line newcase date2 if name == "Tokyo", sort
```


LightStone® 株式会社ライトストーン



グラフから周期的な変動があること確認できました。このような変動を抑えて、数値の変化を調べるには、移動平均を求めます。変数 newcase の直近 1 週間の平均を新しい変数として作成しましょう。Stata では前期の値を参照するために、l.演算子を使用します。使用しているのデータセットは日次データなので、l.newcase は前日、l2.newcase は 2 日前の新規感染者数を参照しています。パネル ID ごとに計算するために、bysort プリフィックスコマンドを使用しています。

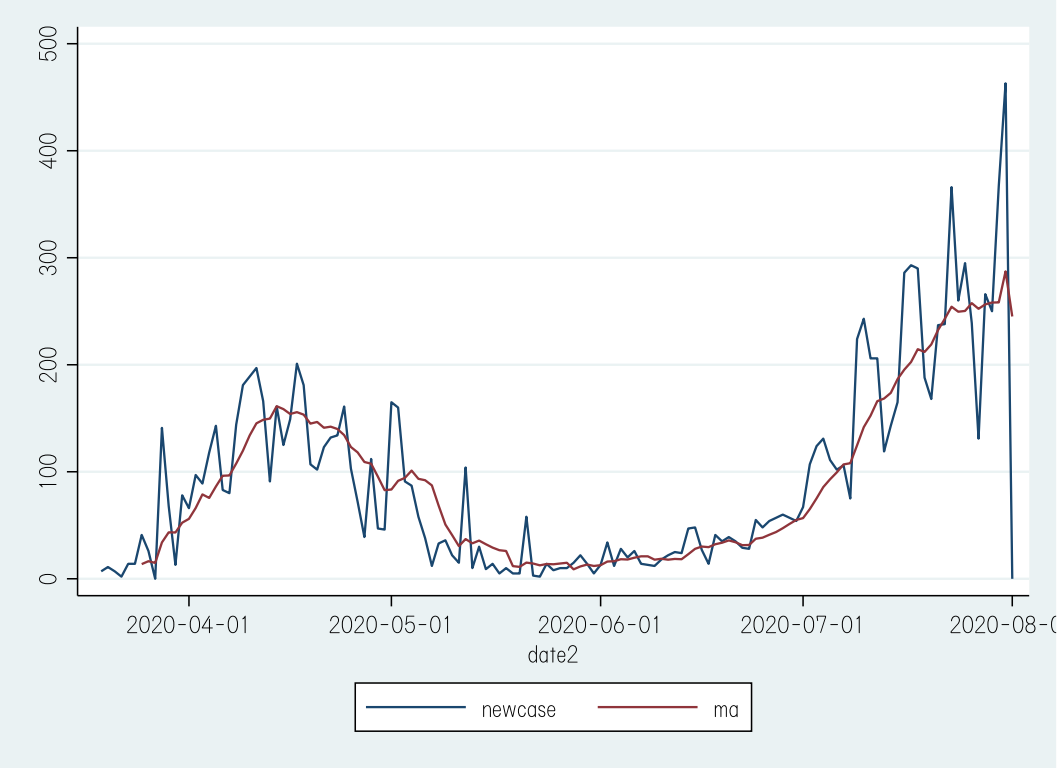
```
bysort prefecture (date2): generate ma = (l6.newcase + l5.newcase +
l4.newcase + l3.newcase + l2.newcase + l.newcase + newcase) / 7
```

新しく作成した直近 1 週間の感染者数平均を示す変数 ma を、先の折れ線グラフに追加します。

```
twoway (line newcase date2 if name == "Tokyo", sort) (line ma date2
if name == "Tokyo", sort)
```

移動平均をプロットしたグラフは次のようになります。周期的な変動がスムージングされていることがわかります。

LightStone® 株式会社ライトストーン



5. まとめ

今回は、「COVID-19 Japan 新型コロナウイルス対策ダッシュボード」が公開している日本国内の新型コロナウィルス感染情報を、Stata からアクセスして保存する方法を紹介しました。ウェブ上で公開されている csv ファイルは、`import delimited "URL"`で Stata に取り込みます。時系列データやパネルデータでは、差分を取る場合は `d.`演算子を、前期の値を参照するには `l.`演算子を使用します。このほかに後期の値を参照する `f.`演算子も用意されています。データセット構造に関係なく 1 つ上のセルを参照する場合は、変数名`[_n-1]`と入力します。

参考文献

Code for Japan. (2020 年 8 月 12 日). COVID-19 Japan　新型コロナウイルス対策ダッシュボード. 参照先: https://www.stopcovid19.jp/

LightStone® 株式会社ライトストーン

研究者向けの統計解析ソフトウェア **Stata® 評価版のお申込み**

機能制限なし！30日間無料で使える評価版でまずはお試しください！

https://www.lightstone.co.jp/stata/evaluate.html

※学生の方への評価版の提供はございません。大学の講義で使用するなど単位取得に関わる形でのご利用となる場合は学生版のご購入をご検討ください。

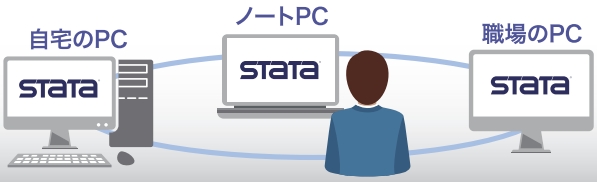
ライセンスはサブスクリプションがおすすめ！

サブスクリプションライセンスを選ぶメリット

- 常にStataの最新バージョンが利用できます
- 技術サポートの対象となります
- 毎年の経費としてStataを導入できます
- 利用者の増減に柔軟に対応可能
- 初期導入費を抑えたい場合にもおすすめ

シングルライセンス（サブスクリプション）

職場のPC・ノートPC・自宅のPC等、個人所有のPCにインストール可能です。
同時に使用できるのは1台のみとなります。



マルチユーザライセンス（サブスクリプション）　2ユーザ以上

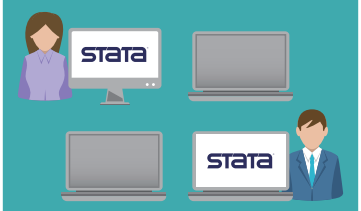
ボリュームライセンス

シングルライセンスを複数人でまとめてご購入いただく際のボリュームディスカウント商品です。

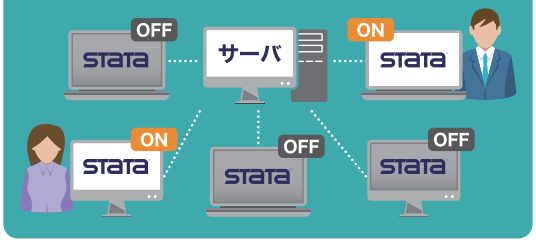
同時起動ライセンス

ご利用環境により下記3通りの運用方法からご選択いただけます。

PC毎にスタンドアロンインストールをして運用
例:2ライセンスの場合



クライアント＆サーバ環境によるネットワーク運用
例:2ライセンスの場合



1台の仮想マシンにユーザがアクセスする運用
例:2ライセンスの場合

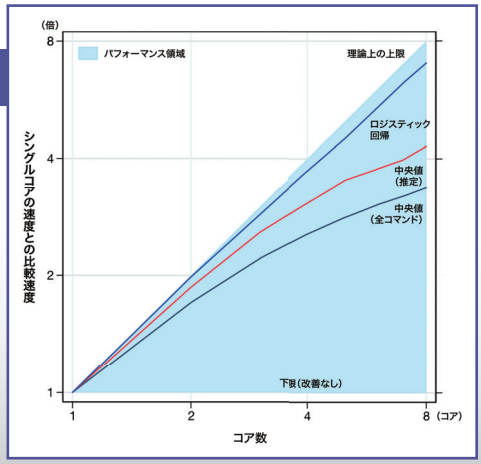
＼データ処理時間の削減ならMP／

Stata／MPの演算能力

Stata MPはPCの持つマルチコアの特性を活かして、処理を分散・並列する機能を備えます。およそ85%以上のコマンドで処理速度が向上し、コア数に応じた計算時間の短縮が期待できます。コマンドごとの処理速度の向上の度合いについては以下のページや資料をご覧ください。

コア数	全てのコマンド	推定コマンド	ロジスティック回帰
2	1.7倍	1.8倍	1.9倍
4	2.6倍	3.1倍	3.8倍
8	3.3倍	4.2倍	6.8倍

弊社Webページ内の『Stata／MP』
https://www.lightstone.co.jp/stata/statamp.html



株式会社 ライトストーン
〒101−0031
東京都千代田区東神田 2−5−12　龍角散ビル 7F
TEL：03−3864−5211　e-mail：sales@lightstone.co.jp
FAX：03−3865−0050　WEB：https://www.lightstone.co.jp/