Stata+α

基本的なデータ分析タスクの自動化 1

1. はじめに

基本的なタスクを do ファイルで自動化すると、データ分析にかかる手間を省いて、さらに単純なミスを防ぐことができます。
今回は、do ファイルを作成して、変数 x を標準化した新しい変数を作成しましょう。Stata 上では、x の値と x の平均の差を計算し、それを標準偏差で割ることで求めることができます。平均や標準偏差は summarize コマンドで計算できます。summarize で計算した結果は、マクロとして記録されます。平均は r(mean)、標準偏差は r(sd)として呼び出すことができます。記録されるマクロはヘルプファイルの Stored results に一覧されています。

```
Stored results

    summarize stores the following in r():

    Scalars
      r(N)            number of observations
      r(mean)         mean
      r(skewness)     skewness (detail only)
      r(min)          minimum
      r(max)          maximum
      r(sum_w)        sum of the weights
      r(p1)           1st percentile (detail only)
      r(p5)           5th percentile (detail only)
      r(p10)          10th percentile (detail only)
      r(p25)          25th percentile (detail only)
      r(p50)          50th percentile (detail only)
      r(p75)          75th percentile (detail only)
      r(p90)          90th percentile (detail only)
      r(p95)          95th percentile (detail only)
      r(p99)          99th percentile (detail only)
      r(Var)          variance
      r(kurtosis)     kurtosis (detail only)
      r(sum)          sum of variable
      r(sd)           standard deviation
```

CAP NUM OVR

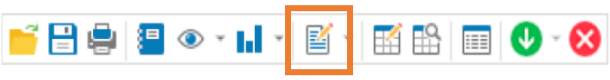
例題データセット auto.dta を sysuse コマンドで読み込んで、summarize で変数 price の記述統計量を求めた後、平均と標準偏差を display コマンドでマクロを呼び出し、確認してみましょう。

```
sysuse auto
```

```
summarize price
display r(mean)
display r(sd)
```

2. do ファイルの作成

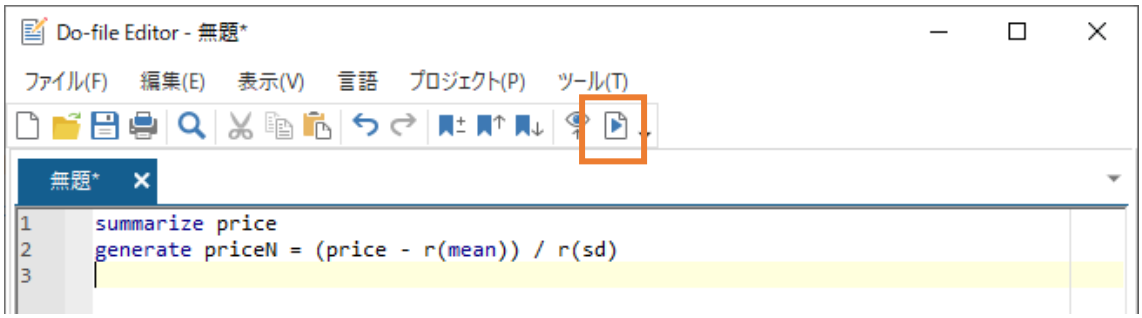
これらを利用して変数 price を標準化する do ファイルを作成します。まずは do ファイルエディタを開くところから始めましょう。do ファイルエディタは、ツールバー内の do ファイルエディタボタンをクリックするか、doedit コマンドを実行します。



do ファイルエディタが開いたら、次の 2 つのコマンドを入力します。2 行目では、generate コマンドで変数 price を標準化した変数 priceN を、前述のマクロを使用して作成します。

```
summarize price
generate priceN = (price - r(mean)) / r(sd)
```

作成した do ファイルを実行するには、do ファイルエディタのツールバー上にある実行ボタンをクリックします。実行出来たら、新しい変数 priceN がデータセットに追加されていることを確認しましょう。



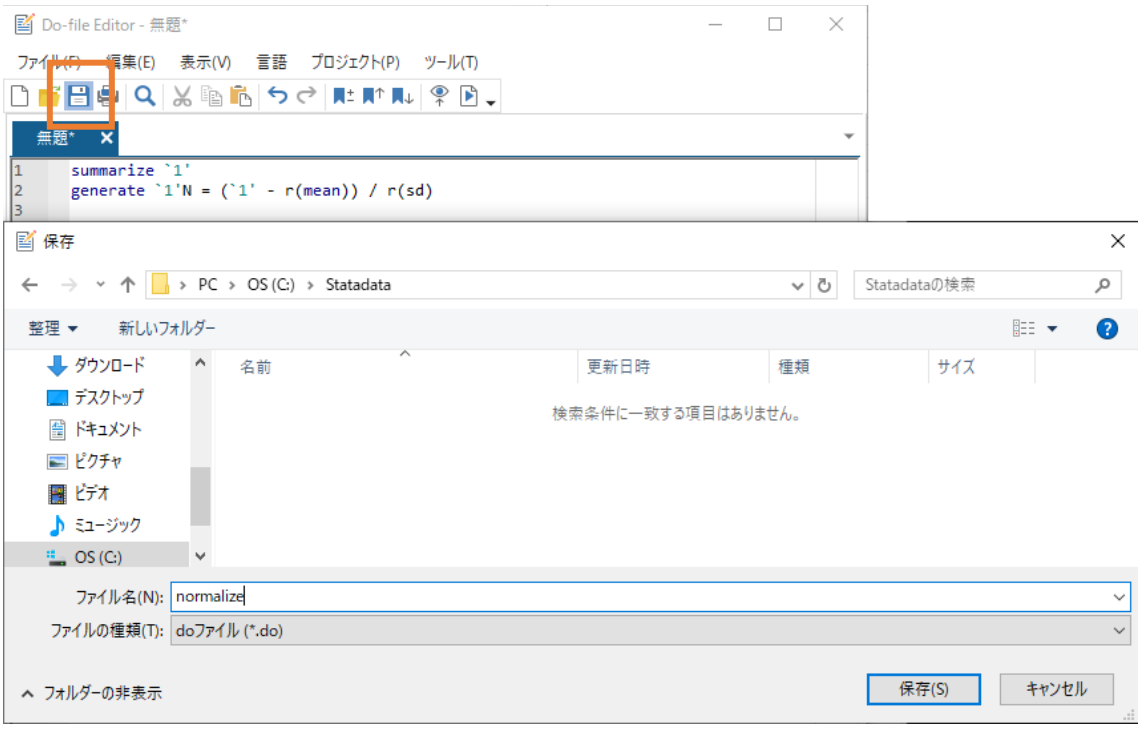
上記のサンプルでは、変数 price の標準化しかできません。これを任意の変数に適用できるように price をローカルマクロ`1'に変更して、do ファイルにしましょう。Stata の do ファイルでは、ローカルなマクロは 1、2、3…となります。これを normalize.do として作業フォルダに保存します。

```
summarize `1'
generate `1'N = (`1' - r(mean)) / r(sd)
```


LightStone® 株式会社ライトストーン

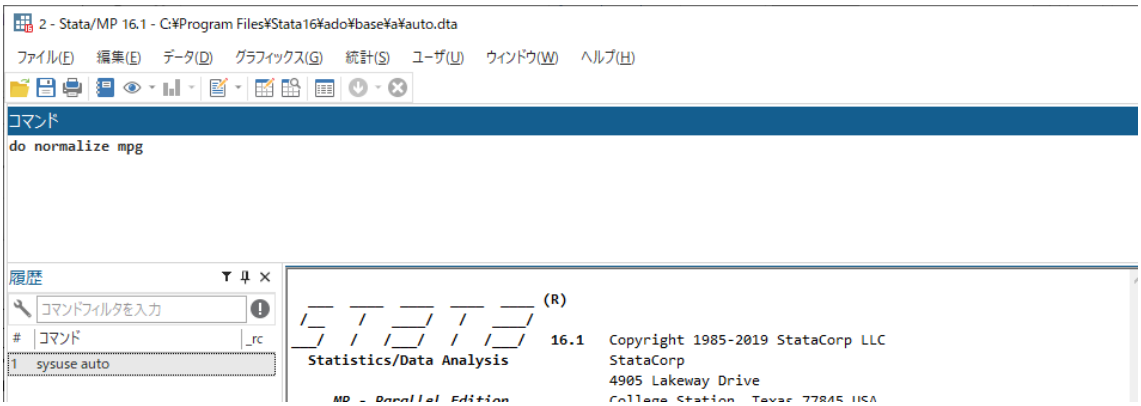
※１つ前のサンプルから変更・追加された点を赤字で示しています。

do ファイルの保存はツールバー上の保存ボタンをクリックし、現在の作業フォルダを選択し、ファイル名を決定、保存ボタンをクリックします。ここでは、ファイル名を normalize として保存しています。



保存した do ファイルをコマンド欄から呼び出し、実行してみましょう。実行するには、do または run コマンドを使用します。コマンド欄に次のように入力して、エンターキーを押しコマンドを実行します。このように normalize.do を実行すると、マクロ`1'には、新たな変数 mpg が代入され、新しく mpgN が作成されます。

```
do normalize mpg
```



LightStone® 株式会社ライトストーン

3. 観測値を制限する

do ファイルは、if や in オプションで観測値を制限できると便利です。do ファイルエディタに戻り、ファイルの 1 行目に syntax コマンドを追加します。syntax に続く、varlist(min=1 max=1)は、normalize コマンドを適用できる変数の数を最小 1、最大 1 に制限し、[if]と[in]は if、in オプションの使用を有効化しています。また、マクロ`1'が`varlist'に置き換わっています。summarize と generate コマンドにも if や in オプションが適用されるよう、行末に`if'と`in'が追加されています。

```
syntax varlist(min=1 max=1) [if] [in]

summarize `varlist' `if' `in'
generate `varlist'N = (`varlist' - r(mean)) / r(sd)   `if' `in'
```

複数の変数を一度に変換できるとさらに便利です。1 行目の varlist(min=1 max=1)から(min=1 max=1)を削除して、複数の変数を指定できるようにしましょう。複数の変数に summarize と generate を適用するには、foreach コマンドで繰り返し処理を行います。ここで、foreach コマンドは、`varlist'として指定された複数の変数を、1 つずつ`var'に代入して、summarize と generate を繰り返します。

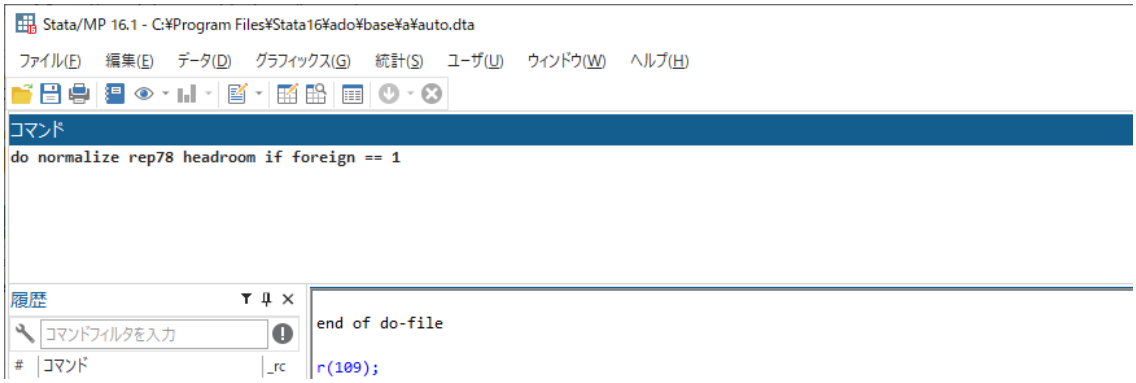
```
syntax varlist [if] [in]

foreach var in `varlist' {
    summarize `var' `if' `in'
    generate `var'N = (`var' - r(mean)) / r(sd)   `if' `in'
}
```

変更したファイルを上書き保存して、コマンド欄に次のように入力して再度 do ファイルを実行し、foreign が 1 を取る観測値に対して、変数 rep78 と headroom を変換してみましょう。

```
do normalize rep78 headroom if foreign == 1
```


LightStone® 株式会社ライトストーン



4. まとめ

今回は簡単なデータ処理を do ファイルにまとめまて、自動化する方法を紹介しました。summarize などで計算した値を後に参照するには、r(mean)などのマクロを使用します。do ファイルでは、syntax varlist [if] [in]を挿入することで、任意の変数・観測値に適用します。作成した do ファイルは作業フォルダに保存して、do または run コマンドで実行します。

今回紹介した情報は Stata 社公式ブログ、The Stata Blog (https://blog.stata.com/)に掲載されている内容を用いて作成しております。

参照文献

Wiggins, V. (2018, October 9). *How to automate common tasks*. Retrieved from The Stata Blog: https://blog.stata.com/2018/10/09/how-to-automate-common-tasks/

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。

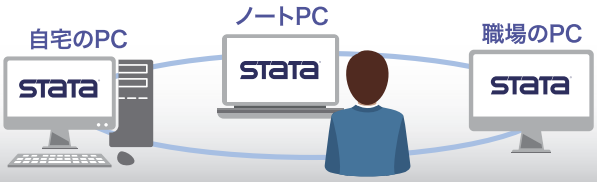
ライセンスはサブスクリプションがおすすめ！

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

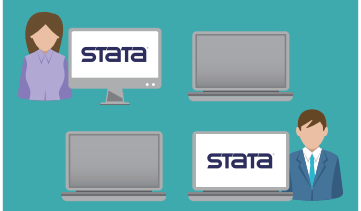
ボリュームライセンス

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。

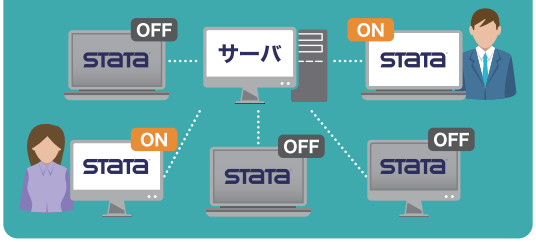
同時起動ライセンス

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

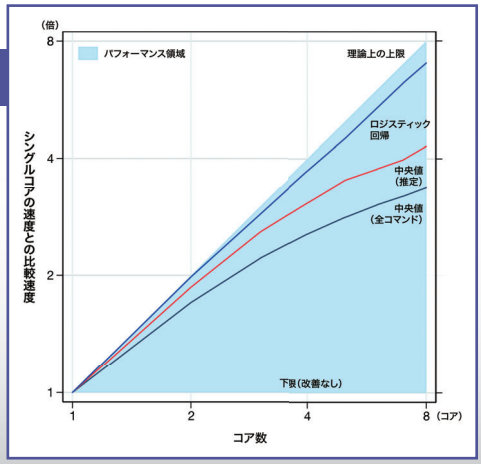
＼データ処理時間の削減ならMP／

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/