

データ統合について

株式会社ライトストーン

2015年11月10日

月次のデータで分析したいのに、特定の変数についてだけは年次のデータしか手に入らない。市町村レベルで分析を行っているのに、特定の変数についてだけは都道府県レベルでしかデータが手に入らない。こうしたことは、非常によくあることです。このように研究や分析に使うデータソースは、一つであるとは限りません。単一のデータソースのみを用いて分析を行う場合、データの構造が一つに定まるため、データの整理はさほど難しくありません。一方で複数のデータソースを組み合わせる場合、それぞれ独自のデータの構造を持っているため、データ整理の段階で特定のデータ構造への規格化が必要になる場合がよくあります。

今回はこのように、異なる構造を持つデータを組み合わせて、一つの分析用ワークファイルを作る方法について解説します。

以下では表記の簡単化のために、分析に使っているワークファイルを移植先ワークファイル、追加したいデータが入っているワークファイルをソースワークファイルと呼びます。

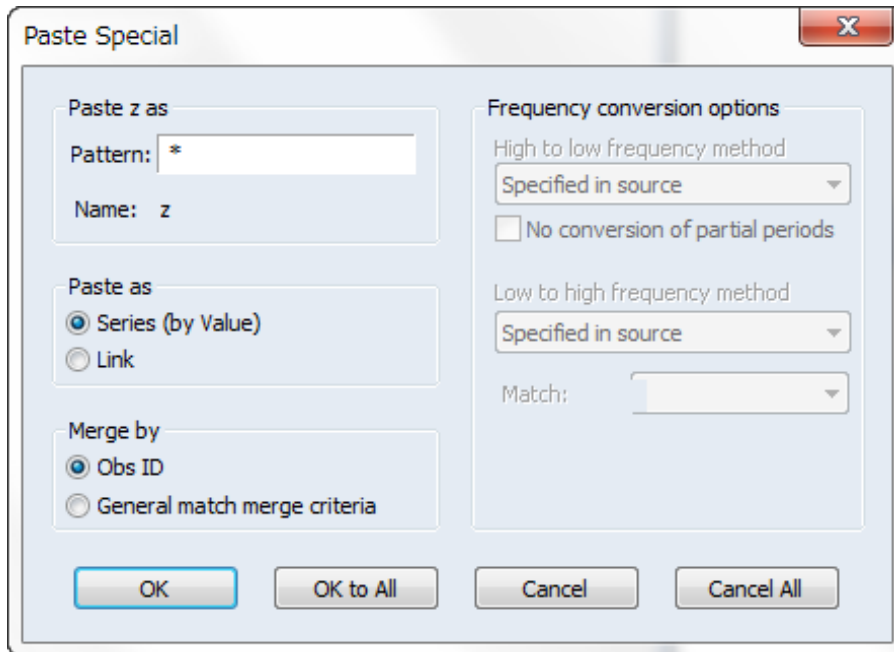
1 One to Many の場合

1.1 基本的な使い方

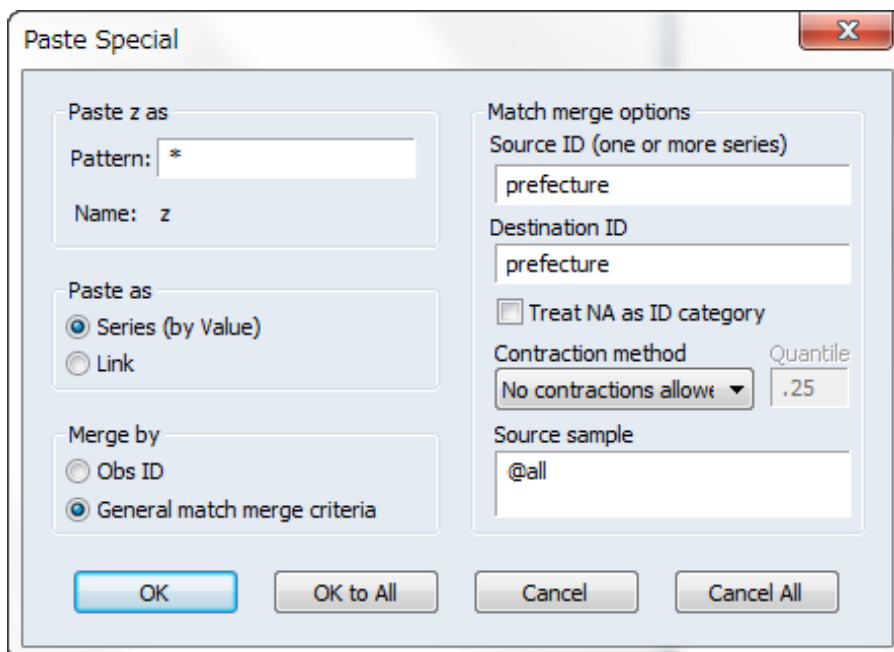
移植先ワークファイルが市町村レベルで、そこに都道府県レベルのソースワークファイルから、変数を移植したいケースを考えます。まずはダブルクリックで citydata.wf1 を開き、File>Open>EViews ワークファイルから prefecdata.wf1 も開いてください^{*1}。citydata には東京都と埼玉県の仮想的な市町村レベルのデータが入力されています。citydata ではいくつかの市町村について X や Y といった変数が利用できますが、ある変数 Z については、市町村レベルのデータが手に入らない状況を想像してください。一方で、別のデータソースから、都道府県レベルでは Z のデータが手に入るものとします。prefecdata には、そのような東京都と埼玉県の仮想的な都道府県レベルのデータがインポートされています。ここでの問題は、各市町村に都道府県レベルの変数 Z を追加したい場合に、EViews 上でどのような操作を行えばよいか？ということなのです。

まずはメニューで操作する方法をご紹介します。それぞれのワークファイルが一度に開かれている状態で、prefecdata 側の Z を citydata 側にドラッグアンドドロップするか、コピーアンドペーストしてください。すると、以下のダイアログが表示されます。

^{*1} EViews は多重起動することが可能なソフトウェアですが、この場合は同一のウィンドウ内に両方のワークファイルを読み込む必要があります。読み込む順番は逆でも構いません。

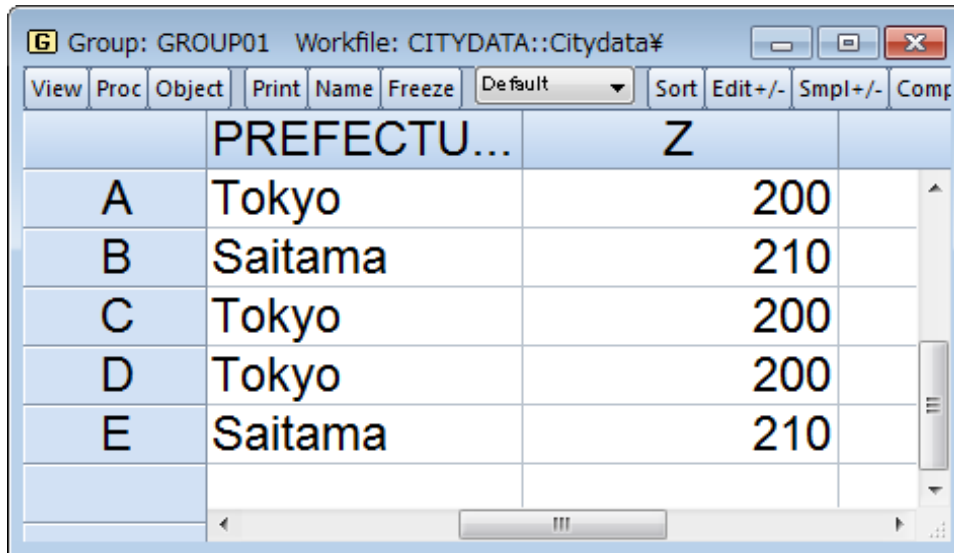


次に、General match merge criteria を選択してください。すると、ダイアログの右側が切り替わりますので、以下のように入力してください。



OK を押すと、Z が citydata 側にコピーされたはずですが、ただし、正しくコピーされていない場合は、もちろん意味がありません。東京の Z の値は 200、埼玉の Z の値は 210 でした。以下のコマンドでグループオブジェクトを作ってみると、その確認が簡単に行えます。

group group01 prefecture z

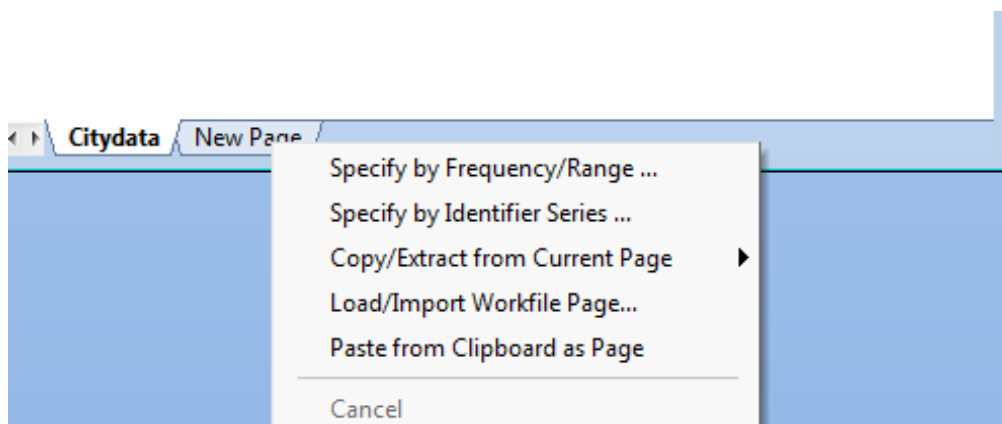


	PREFECTU...	Z	
A	Tokyo	200	
B	Saitama	210	
C	Tokyo	200	
D	Tokyo	200	
E	Saitama	210	

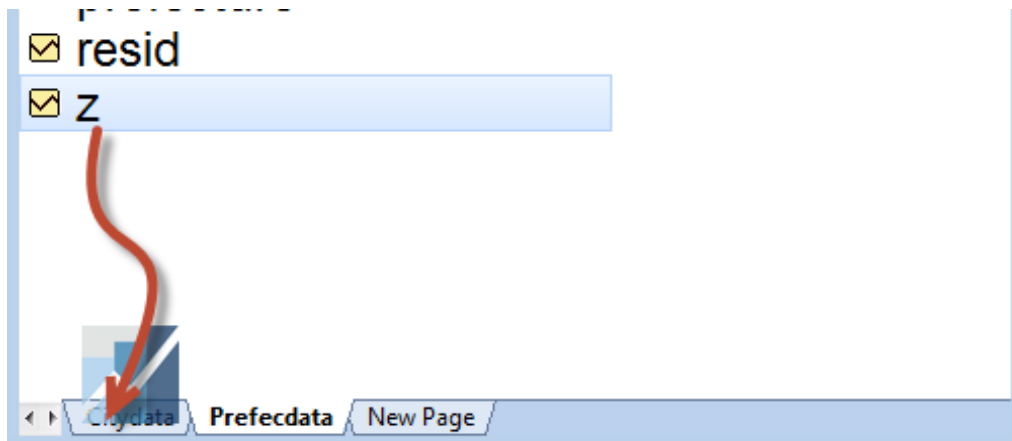
実際、東京都に属する市町村 A, C, D については 200, 埼玉県に属する市町村 B, E については 210 という値が、正しく入力されていることが確認できます。

1.2 ワークページ間で移動を行いたい場合

ワークファイル間だけでなく、ワークページ間でも同じことが可能です。citydata を開き、移動したい変数が入力されているワークページを表示し、以下の場所をクリックしてください。Load/Import Workfile Page を選択し、prefecdata を読み込みます。



これにより、prefecdata をワークページとして読み込むことができました。後は簡単です。prefecdata タブに切り替え、Z をコピーし、citydata にタブを切り替え、ペーストします。あるいは、以下のように Z をドラッグアンドドロップします。



後の操作は同様です。

1.3 複数のオブジェクトをまとめて移動したい場合

複数のオブジェクトをまとめて移動することもできます。prefecdata には、Z の他に alpha という名前で、アルファシリーズオブジェクトが入力されています。alpha と Z をまとめて移動したい場合、単純に alpha と Z をまとめて選択して、同様の操作を行うだけです*2。

移動したいオブジェクトそれぞれに対して、どのような設定で移植を行うかの設定を行います。あるいは、OK to All を選択することで、同じ種類のオブジェクトに対しては、同じ設定を使った一括処理を行うことができます（ただしこの例の場合、シリーズオブジェクトとアルファシリーズオブジェクトという異なる種類のオブジェクトを移動しようとしているため、OK to ALL を選択してもダイアログが 2 回表示されます。）

1.4 コマンドで移動を行いたい場合

Copy コマンドを利用します。まずは両方のデータを開いてください。citydata を選択した状態で、たとえば以下のように入力して実行します。

```
copy(c=none) PREFECDATA::PREFECDATA\Z * @src prefecture @dest prefecture
```

citydata を選択していない状態でも移動を実行できるようにしたい場合は、以下のように入力します*3。

```
copy(c=none) PREFECDATA::PREFECDATA\Z CITYDATA::CITYDATA\Z @src prefecture @dest  
prefecture
```

最初のコマンドではデータの移植先がワイルドカード*で表されているので、デフォルト値が使用され、選択中のワークファイルに対しコピーが実行されます。一方で 2 番目のコマンドではデフォルト値ではなく指定された移植先にコピーが行われます。基本的には、以下のように記述してください。

*2 オブジェクトをまとめて選択するには、ctrl キーを押しながらマウスでオブジェクトをクリックします。あるいは、shift キーを使うと、選択した 2 点を始点と終点とした、範囲選択を行うことができます。

*3 紙面の横幅の関係で最後の Prefecture が別行になってしまっていますが、実際には同じ行内に入力する必要があります。

copy(オプション)- データの出力元- データの出力先_@src_ ソース ID_@dest_ 出力先 ID

アンダーバーは半角スペースを表しています。“ ”や全角スペースを入力するとエラーになります。データの出力元や出力先は、ワークファイル間で作業を行う場合はワークファイル名::ワークページ名 \ コピーしたいオブジェクト名という記述になります。オプションの copy(c=none) は contraction を許可しないという意味です。他の利用できるオプションやより細かいコマンドの記述方法については、EViews Help Topics: Copy をご確認ください。

また、EViews9 の新機能である、コマンドキャプチャ機能を使うと、一度メニューから行った操作について、対応するコマンドを調べることができます（実際、一番最初にご紹介したコマンドはコマンドキャプチャ機能でキャプチャされたものをそのまま張り付けています）。何かと便利な機能ですので、EViews8 以前のバージョンをお使いの方は、もしよろしければアップグレードをご検討ください。

1.5 その他情報

1.5.1 Source ID に重複がある場合

コピーの実行時、“More than one source value with contraction disallowed”というメッセージが出る場合、Source ID に重複があります。例えば東京について、Z=200 というデータと、Z=220 というデータの両方が併存して入力されている場合です。ソフトウェアはどちらの Z の値を使うべきかを判断できないため、このような場合に No contraction allowed オプションを選択していると、冒頭のエラーが生じます。最終的な目的に応じてデータを修正するか、contraction オプションを変更してください。contraction オプションについては次回により詳しく解説します。

1.5.2 Source ID と Destination ID のオブジェクト名が異なる場合

citydata と prefecdata で、ID として用いるオブジェクトの名前が違うケースもありえます。例えば citydata では都道府県名が prefecture というオブジェクト名で格納されている一方、prefecdata には都道府県名が todofuken というオブジェクト名で格納されているようなケースです。このように Source ID と Destination ID が異なる場合も、ダイアログでそれぞれに適切な名前を入力すれば、データの移植が可能です。

1.5.3 Source ID 側に Destination ID 側がない変数がある

例えば prefecdata の Source ID 側に、citydata 側の Destination ID には存在しない神奈川県や千葉県の情報が入っているとします。この場合も、他の設定が正しければ出力には影響しません。

1.5.4 コマンド操作が上手くいかない場合

コマンド名などの記述が間違っていないかまずご確認ください。また、必要な位置にスペースが入っているか、不必要な位置にスペースが入っていないかについてもご確認ください。EViews のプログラミングでは、基本的に意味のまとまり毎に半角スペースを入力する必要があります。例えば、入力先と出力先の間に半角スペースが入っていないと動作しませんし、逆に copy と (c=none) の間にスペースが入っているような場合も動作しません。また、全角スペースなどの全角文字を入力した場合もエラーとなります。

1.5.5 既に同じ名前のオブジェクトが移植先ワークファイルに存在する場合

上書き、部分的な移植などが可能です。より詳しくは次回説明します。

今回は、一つの Destination ID に対応する Source ID が一意に定まる状況の場合の操作を学習しました。EViews によるデータの統合機能は非常に便利ですが、誤解や操作ミス、あるいは未知のバグなどにより、データが正しく移植されない場合も考えられます。データが正しく移植されているか、目視による確認を怠らないようにしてください。次回は Destination ID に対応する Source ID が一意に定まらない状況、言い換えれば移植する候補として複数のデータが考えられる場合の操作方法について解説したいと思います。