2013 年 11 月

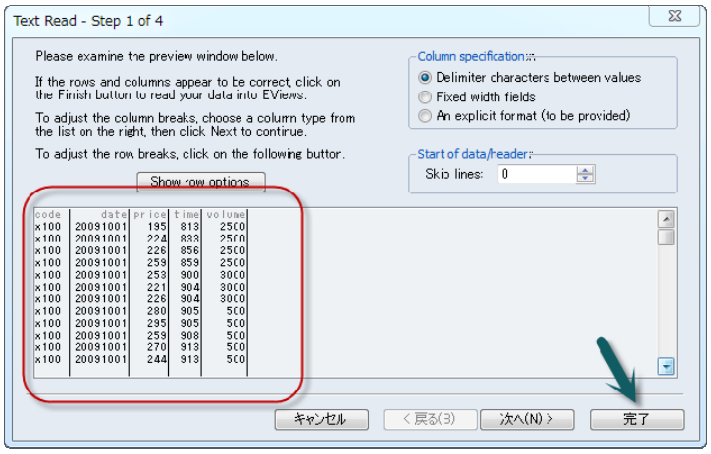
EViews データ分析の前に

今回から数回に分けて、データ分析の前に留意すべき事柄と、そのために利用できる EViews の便利な機能をご紹介します。大容量のデータを EViews プログラムで処理する際に知っておくと便利な機能やテクニックも説明します。また、データの操作に関しては、意識的にプログラム用のコマンドを多用する事にします。実証分析を行うためには、どうしても簡単なプログラムのスキルが必要とされるますので、それに慣れていただくのが理由です。それでは、いよいよ本題にはいりましょう。

■CSV ファイルの読み込み

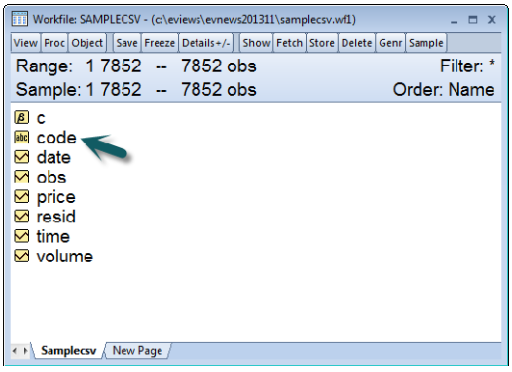
いわゆるテキストデータと呼ばれるものの代表格である CSV ファイルをインポートしてみましょう。CSV ファイルはデータの列をカンマで区切っているというものです。EViews8 で File/Open/Foreign Data as workfile と操作し、サンプルデータ samplecsv.csv を 選択します。

次に示すダイアログでデータの内容が列ごとに正しく表示されていれば、完了ボタンをクリックします。(「次へ」のボタンをクリックして詳細に設定する必要はありません)



サンプルデータの中身は架空の株価データです。

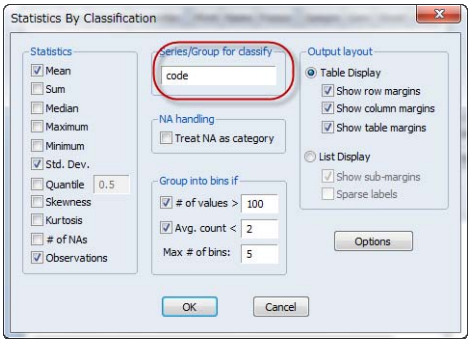
銘柄コードはアルファシリーズ code に格納されています。全データの個数が 7852 個あるようですが、その内訳を確認しましょう。code のアイコンをダブルクリックして開き、View/One-Way tabulation と操作します。



例えば、銘柄 x100 の価格情報は 1118 個あり、全データの中で占める割合(個数)は 14%程度あることがわかります。
x700 という銘柄のデータがやや少なめですが、極端に多かったり、逆に少なすぎるデータは無いことが分かります。

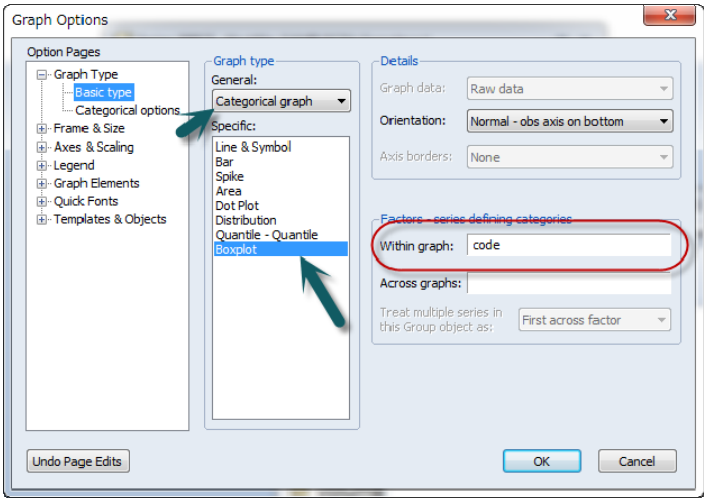
Value	Count	Percent	Cumulative Count	Cumulative Percent
x100	1118	14.24	1118	14.24
x200	866	11.03	1984	25.27
x300	997	12.70	2981	37.96
x400	1052	13.40	4033	51.36
x500	956	12.18	4989	63.54
x600	948	12.07	5937	75.61
x700	777	9.90	6714	85.51
x800	1138	14.49	7852	100.00
Total	7852	100.00	7852	100.00

次に株価に関する情報を見てみましょう。
price のシリーズアイコンをダブルクリックしてウィンドウを開きます。続けて View/Descriptive Statistics & Tests/ Stats by classification と操作します。コマンドの意味は階層別にデータの統計量を見てみよう、というものです。図のように設定したら OK ボタンをクリックします。
表の中の Mean に注目すると、銘柄 x500 は大型株で他のデータとは、明らかに大きさの異なることが分かります。
視覚的にグラフとして大きさの違いを、分布も含めて確認しましょう。price のオブジェクトで View/Graph と操作し、Graph Options ダイアログで次のように設定します。ダイアログ中央の Graph type の項目で General を Categorical grpah にし、グラフの種類は Boxplot を選択します。そして、カテゴリ分けに利用するシリーズとして、Within graph の項目に code と入力し、OK ボタンをクリックします。



Sample: 1 7852
Included observations: 7832

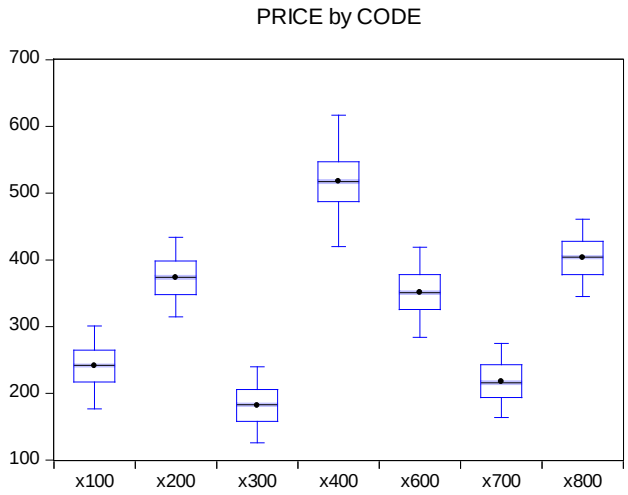
CODE	Mean	Std. Dev.	Obs.
x100	241.3519	29.56753	1114
x200	373.4907	30.01162	864
x300	181.8484	29.18346	996
x400	517.7414	42.80458	1048
x500	285778.7	3462.032	954
x600	351.4000	32.01338	945
x700	217.6000	29.04038	775
x800	403.4049	29.78664	1136
All	35100.51	93373.59	7832



しかし、残念ながら、何やら妙なグラフになってしまいます。これは、x500 だけがこの中では飛びぬけて大きいので、箱ひげ図がどれも潰れてしまっています。このような場合は、標本から x500 の価格を取り除いてみましょう。コマンドウィンドウに次のように入力します。

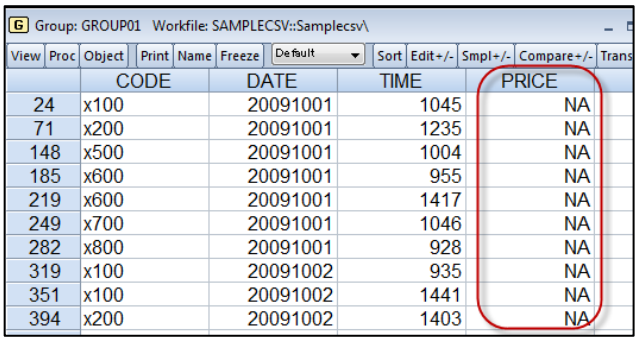
smpl　@all　if　code <> "x500" (Enter キーを押します)

コマンドの意味を確認しておきます。
smpl は標本を設定するための基本的かつ重要なコマンドです。 @all は Range の範囲をそのまま利用し、不等号の向き合った形は not equal です。コード番号をダブルクォーテーションで囲んでいるのは、銘柄が文字列であるためです。
EViews では基本的に、文字列はダブルクォーテーションで囲む、というルールがあります。このコマンドを実行したのち、改めて、同じ操作をすると先のグラフを作成できます。



■欠損値の処理
例えば、x100 の欠損値について調べてみましょう。コマンドウィンドウに次のコマンドを入力します。

group group01 code date time price



	CODE	DATE	TIME	PRICE
24	x100	20091001	1045	NA
71	x200	20091001	1235	NA
148	x500	20091001	1004	NA
185	x600	20091001	955	NA
219	x600	20091001	1417	NA
249	x700	20091001	1046	NA
282	x800	20091001	928	NA
319	x100	20091002	935	NA
351	x100	20091002	1441	NA
394	x200	20091002	1403	NA

```
smpl @all if price=na
show group01
```

1 行目は group オブジェクト group01 を作成するコマンドです。グループメンバは code, date, time,price です。次に標本を価格が欠損値になっているデータだけに限定します。EViews では欠損値は na で示し、属性は数値になりますので、文字ではありますが、ダブルクォーテーションで囲みません。最後に、show コマンドで group 01 のウィンドウを開きます。
ここで、欠損値を埋めることを考えます。データ処理の方針は次の通りです。

方針)コードごとに、価格の中の欠損値をチェックします。欠損値の場合は 1 行上のデータを代入します。

ポイントはコードごとにこの作業を行うところにあります。コードの区別なく、この作業を行うと、ある日の銘柄の最初の価格が欠損値の場合に、他の銘柄の最後の値を代入することになってしまいます。また、コードごとに操作することで、EViews の元から備えている便利な機能を活用することができます。

```
series price2=price
smpl @all if code="x100"    and    price=na
price2=price2(-1)
group01.add    price2
```

元のデータは確認用にとっておきたいので、編集用データを price2 として複製します。標本を銘柄 x100 で、かつ、欠損値のあるデータのみとします。そして、欠損値に対して、その一行上のデータを代入します。最後に、グループオブジェクトに price2 を加えて、視覚的に欠損値に値が代入できた事を確認します。

```
smpl @all if code="x100"
```

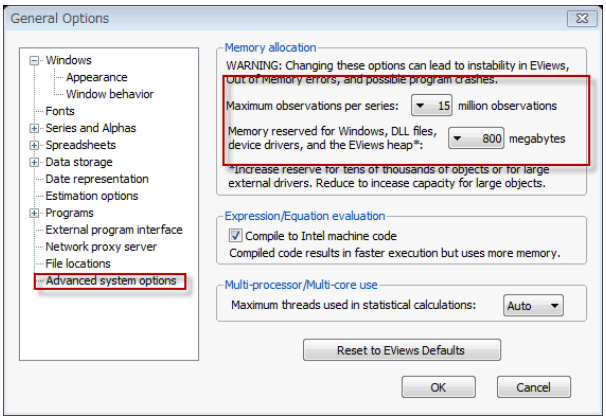
このようにしてコード x100 について price と price2 の値を確認してください。

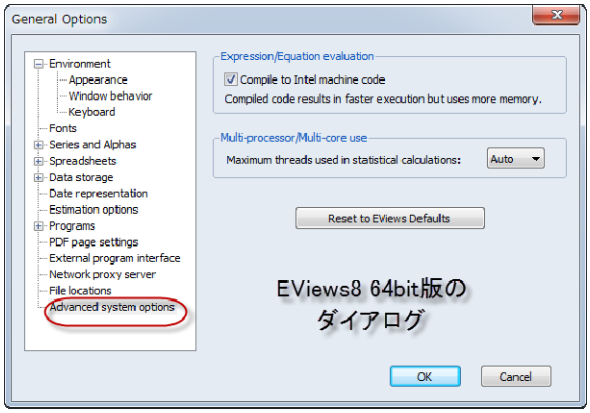
ただ、ここでは欠損値は 1 行だけ、まばらに存在していました。もし、連続して複数行に欠損値が存在していたら、どうなったでしょう。答えは、それらのセルにも直前の値が連続して入ります。ぜひ、ご自身で確認してください。
ここで紹介したコマンドは EViews の機能をよく理解していないと思いつかないかもしれ

ません。直感的には loop と if 文を利用してプログラムを書き、7852 回、チェック作業を行いたくなるかもしれません。

コードは x100 から x800 まで規則的な名前で存在します。したがって、一連の操作をコードの違いに気を付けて 8 回行えば、欠損値の処理を終えることができます。しかし、名前に規則性があるので、次回は、文字列の操作方法に関する EViews のプログラム機能を利用して、ループによってこの処理を 1 回で終えるようなプログラムを作成してみたいと思います。

[補足情報]





最後に EViews8 に関する補足の情報をご案内します。EViews は v8 ではようやく、64bit アプリケーションとなりました。まずは従来の 32bit 版との違いから確認します。

EViews7 まではメインメニューの Options/General Options と操作して表示される General Options のタブで、EViews に取り込み可能なデータの個数を設定していました。

最新の EViews8 はインストール時に 32bit 版と 64bit 版を選択できるようになりました。つまり、64bit 版の Windows をご利用であれば、64bit 版の EViews を利用できます。従って、64bit 版の EViews8 では上図の変数名の設定項目は存在しません。この事は 64bit 版 Windows の稼働している PC で、ハードディスクに収納できるどんな大きさのファイルも、64bit 版の EViews なら開けることを意味しています。

ただし、ソフトウェアの一般的な話として、データを開けることと、処理速度は全く関係ありませんので、ご注意ください。せっかく大容量のデータを開くことができても、データ処理に想像以上の時間がかかるようでは、意味がありません。

■