

Stata Learning Notes

Ethan Deng

Version 1.0

September 16, 2015

1 数据操作

1.1 将 Excel 文件数据导入 Stata 的方法之比较

这一点可能也是大家在实证研究中遇到的最常见的问题，一般找来的数据都是 Excel 文件，但是要做计量肯定要导入到 Stata 中去。可能大多数菜鸟最常见的办法就是直接复制粘贴了，简单明了，不需要任何技术，但潜在的问题是如果数据比较多，拉框选择还是一件很麻烦的事情，有些数据在 Excel 中可能是显示小数点后两位，但实际储存的并不止这么多。如果是复制粘贴了，可能只粘贴过去小数点后两位，这样就损失了一部分精度。因此我不太推崇这种方法，就我的经验来看，直接调用 Stata 的内部数据导入命令会更好，起码不会丧失精度。Stata 数据导入的命令有以下几种，下面分别加以评析：

1. `insheet`

```
1 insheet using filename, [option]
```

这个命令是专门用来导入像 Excel 之类的以电子表格形式存储的数据。在导入之前，首先要将 Excel 文件转存为 Stata 可以识别的格式。其中我最常用的就是另存为 csv 逗号分隔符格式。当然一个 xls 文件中会包含多个 sheets，但 csv 只能保存一个 sheet，在另存时 Excel 会提示你是否只另存当前 sheet，并且有些功能会丧失掉，不管那么多，毕竟 csv 只是一个中介，直接点确定就行了。当然你在保存 csv 时最好让变量名位于第一行，这样 Stata 就好识别了。然后在 Stata 中使用 `insheet` 读取 csv 文件，在 `option` 中指定为 `comma` 告诉 Stata 你读取的是 csv 文件。这种方法有个不足在于如果你的数据中包含中文而且里面含有逗号时，stata 是无法区别这些逗号是数据中的逗号还是 csv 文件自己生成的逗号，因而会出现同一个变量的数据被拆散到两个变量的存储空间中。解决的办法是不要用逗号标示分隔符了，在 Excel 中另存为 txt (制表符分隔)，这样就不会与逗号相混淆了。然后再在 `insheet` 命令中在 `option` 里指定是 `tab`，就完事了。

2. `infile`

```
1 infile using filename
```

这个 `infile` 命令分两类，一种是处理固定格式 (fixed format) 的 txt 或 raw，另一种是处理自由格式 (free format)，两类文件的命令语法不太相同，适用的类型也有所不同，当然你在用这个命令里还需要定义一个 dictionary，这个 dictionary 是用来描述数据的组织方式的，需要自己根据要导入的数据文件自己编写代码，然后嵌套到数据文件 txt 的前面去，或者是单独地存为一个 dct 文件，并且告诉 Stata 你要导入的数据在保存在哪里。由于使用起来的学习成本比较高，用得也比较少。在此就不详解了，有需要的请自行 `help infile`。

3. `xmluse`

这个命令首先要把 xls 文件另存为 xml 格式，然后用 `xmluse` 命令去读取，当然在读取时你也要在 `option` 中声明你的 xml 文件是 Excel 保存的而不是 Stata 保存的，这样就不会弄错。这个命令相比于前面的 `insheet` 的一大好处在于你不必为 xls 中的每个 sheet 单独另存为一个文件，而只需要整体另存一个 xml 格式就行了，在用 `xmluse` 时在 `option` 中告诉 Stata 从哪个 sheet 中的哪些 cell 中读取数据就行了。但是经本人的经验发现如果你的 xls 文件中如果有汉字的话，Stata 读取后对应的变量会出现乱码，这一点用 `insheet` 就不会有这个问题。

4. odbc

这个命令是专门读取数据库文件的，并且支持 SQL 命令，这样如果你的数据比较多的话，可以先用 SQL 语句进行筛选，然后而导入。当然这个命令也能导向 excel 文件，只是本人目前还没用过，在此也不详解了。

更新：Stata12 可以用 `import excel` 命令直接导入。Stata13 中 `insheet` 已经被 `import delimited` 取代：“insheet has been superseded by import delimited. insheet continues to work but, as of Stata 13, is no longer an official part of Stata.”

补充：可以用 Stat/Transfer 软件转换成 Stata 格式。