

中国健康与营养调查数据的合并

1. 实验基本原理：

准确的数据是进行科学实证研究的基础，如何对已取得的数据进行科学、合理的处理，是非常非常重要的。本次主要根据“中国健康与营养调查”数据，演示数据管理的一些基本技巧：数据纵向合并；数据横向合并。

数据纵向合并就是将保存在某个盘符中的 STATA 数据文件的内容追加到当前数据窗口下面，生成一个新数据。原理如下图：

three. dta		
a	b	c
1	2	3
4	5	6
+		
four. dta		
a	b	D
7	8	9
10	11	12
13	14	15

合并后的数据			
a	b	c	D
1	2	3	
4	5	6	
7	8		9
10	11		12
13	14		15

横向合并从一个数据库中提取一些变量到另一个数据库中，故横向合并也就是变量的合并。利用横向合并方法，可以将两个或多个文件整合在一起，形成一个新数据。例如，将一个母亲数据合并到一个儿童数据中，从而使儿童数据包含母亲信息。但是，这个前提是，母亲要是儿童的母亲。横向合并要比纵向合并复杂的多，存在着 3 种情况：“一对一合并”、“一对多合并”、“多对一合并”。

2. 实验目的与要求

- (1) 学会运用 STATA 计量软件，对数据进行纵向合并。
- (2) 学会运用 STATA 计量软件，对数据进行横向合并。

3. 实验内容与数据来源

本次采用的数据是“中国健康与营养调查”部分公开的数据。中国健康与营养调查由美国北卡罗莱纳大学教堂山校区的罗莱纳州人口中心（the Carolina Population Center at the University of North Carolina at Chapel Hill）和中国疾病控制和预防中心的国家营养和食品安全所合作建立的一个数据。

4.实验操作指导

a.关于数据的纵向合并

有 2 个独立的数据：城市儿童的个人信息；农村儿童的个人信息。为了比较在其他条件同等条件下城乡儿童体重升高差异等问题，需要将这 2 个数据进行合并。

第一步：读取使用数据（城市数据），并观察该数据的基本信息。

```
use "D:\work\case2\case2_umerge.dta"
```

```
describe
```

第二步：读取主要数据（农村数据），并观察该数据的基本信息。

```
use "D:\work\case2\case2_rmerge.dta"
```

```
describe
```

第三步：将使用数据合并到主要数据中。

```
append using "D:\work\case2\case2_umerge.dta"
```

```
describe
```

```
list
```

第四步：保持该数据

```
save "D:\work\case2\case2_rumerge.dta",replace
```

b.关于数据的横向合并

第一种情况“一对一合并”。

有丈夫与妻子两个数据。妻子的数据包含：妻子的标识符、丈夫的标识符、妻子的年龄、妻子的教育、妻子的职业。丈夫的数据包含：妻子的标识符、丈夫的标识符、丈夫的年龄、丈夫的教育、丈夫的职业。现在以“丈夫的标识符(husbid)”作为关键变量，将丈夫的资料整合到妻子的数据中，形成一个包含夫妻双方资料的数据。

*1. one -to- one on key variables

```
clear
```

```
use "D:\work\case2\case2_merge_husb.dta"
```

```
sort husbid
```

```
save "D:\work\case2\case2_merge_husb.dta",replace
clear
use "D:\work\case2\case2_merge_wife.dta"
sort husbid
merge husbid using "D:\work\case2\case2_merge_husb.dta",unique
tab _merge
```

第二种情况 “一对多合并”。

这种合并要求主要数据和使用数据具有等级层次，一个数据是高层次数据（比如社区或家庭），另一个数据为低层次数据（比如家庭或个人）。这里使用 case2_rumerge 与 case2_mqmerge 数据来进行演示。前者是低层次的儿童数据（使用数据），后者是高层次的母亲数据（主要数据）。

```
use "D:\work\case2\case2_rumerge.dta", clear
sort mothid kidid
describe
list
save "D:\work\case2\case2_rumerge.dta",replace

use "D:\work\case2\case2_mqmerge.dta", clear
sort mothid
describe
list
merge mothid using "D:\work\case2\case2_rumerge.dta"
describe
sort mothid kidid
list mothid kidid mage age mheight height medu mincome _merge
tab _merge
save "D:\work\case2\case2_one_to_many.dta",replace
```

第三种情况“多对一合并”。

“多对一合并”情况与第二种情况完全相反，类似将“母亲数据（使用数据）”匹配到“儿童数据（主要数据）”上。具体如下：

```
use "D:\work\case2\case2_rumerge.dta", clear
merge mothid using "D:\work\case2\case2_mqmerge", uniquising
sort kidid
list kidid  mothid age mage  height mheight  medu mincome _merge
tab _merge
```