How can I randomly assign observations to groups in Stata?

Authored by

stats writer

June 30, 2024

Random assignment of observations to groups in Stata refers to the process of randomly dividing a dataset into distinct groups or categories. This is commonly used in research studies to ensure a balanced distribution of data and avoid bias in the results. To randomly assign observations in Stata, one can use the "sample" command which allows for the selection of a random subset of the data. This command can be customized to specify the number of groups and the size of each group. By using this method, researchers can ensure that their data is evenly distributed among different groups, increasing the validity and reliability of their findings.

How can I randomly assign observations to groups in Stata? | Stata FAQ

The trick here is to create a random variable, sort the dataset by that random variable,
and then assign the observations to the groups. Let's use the hsb2 dataset as an example by
randomly assigning 50 observations to each of four groups.

```
use https://stats.idre.ucla.edu/stat/stata/notes/hsb2, clear

set seed 12345
generate rannum = uniform()
sort rannum
generate grp = .
replace grp = 0 in 1/50
replace grp = 1 in 51/100
```

```
replace grp = 2 in 101/150
replace grp = 3 in 151/200

tabulate grp

grp | Freq. Percent Cum.
------------+-----------------------------------
0 | 50 25.00 25.00
1 | 50 25.00 50.00
2 | 50 25.00 75.00
3 | 50 25.00 100.00
------------+-----------------------------------
Total | 200 100.00

sort id
clist id grp in 1/20

id grp
1. 1 0
2. 2 3
3. 3 2
4. 4 1
5. 5 0
6. 6 3
7. 7 1
```

8. 8 2
9. 9 0
10. 10 0
11. 11 1
12. 12 0
13. 13 3
14. 14 0
15. 15 3
16. 16 3
17. 17 3
18. 18 1
19. 19 3
20. 20 3

Of course, when you try this the grp number for each id will be in a different pattern
because we are using a random process to assign observations to groups.

It is possible to make the code even simpler then the above by using the egen , cut()
command.

use https://stats.idre.ucla.edu/stat/stata/notes/hsb2,

```
clear

generate rannum = uniform()
egen grp2 = cut(rannum), group(4)

sort id
list id grp2 in 1/20
```

	id	grp2
1.	1	0
2.	2	3
3.	3	2
4.	4	1
5.	5	0
6.	6	3
7.	7	1
8.	8	2
9.	9	0
10.	10	0
11.	11	1
12.	12	0
13.	13	3
14.	14	0
15.	15	3
16.	16	3

17. 17 3

18. 18 1

19. 19 3

20. 20 3

For more information see the Stata manual or Stata Help for
functions.