mwp-471

splitsample - 標本データの分割 【 評価版 】

splitsample はデータをいくつかのランダムなサンプルに分割します。

Example 1
Example 2
Example 3
Example 4

機能概要と用例

splitsample はデータをトレーニング用のサンプルとテスト用のサンプルに分割する際に有用な機能を提供 します。

▷ Example 1: 観測データ単位の分割

本用例では観測データ (observation) 単位の分割例を紹介します。最初に 101 個の観測データからなるデータ セットを生成します。

```
. set obs 101 Number of observations (_N) was 0, now 101.
```

この操作によって 101 個の観測データを含むデータセットが生成されたわけですが、この段階では変数を全く 持たないデータセットとなっています。それでも splitsample コマンドの実行には何ら支障がありません。 最初に標本識別用の変数名のみを指定する形で splitsample を実行してみます。

[ⓒ] Copyright Math 工房; 一部 ⓒ Copyright StataCorp LP (used with permission)

• Data > Create or change data > Other variable-creation commands

▷ Split data into random samples と操作

• Main タブ: Generate sample ID variable: svar

Split into # random samples of equal size: • ($\ensuremath{\vec{\tau}}\ens$

2 (デフォルト)

splitsa	ample - 1	Split data i	into rand	om sample	es	-		×
Main	if/in	Advance	d					
Genera svar	ite samp	ole ID varia	able					
Rep	place an	y variable	with abo	ve name				
Option	IS							
O Spl	it into # 2 🌻	random s	amples o	f equal siz	e			
	ecify nur	nlist of pro	oportions	or ratios f	or splitt	ing		
Rar	ndomly	round sam	ple sizes	when an e	xact spli	t canno	t be made	e
Specif	y numlis	t of values	s for sam	ple ID vari	able:			
Split b	y cluster	rs defined	by variat	ole:				
Across	s the san	nples, bala	ince grou	ps defined	l by varia	ables:		
								<u> </u>
			_					
? C				OK	Cá	ancel	Sub	mit

図1 splitsampleダイアログ - Main タブ

ır)

生成された標本識別用変数 svar の内容(度数)を確認しておきます。

. tabulate svar $^{\ast 1}$

. tabulate	svar		
svar	Freq.	Percent	Cum.
1	51	50.50	50.50
2	50	49.50	100.00
Total	101	100.00	

デフォルトの場合、splitsample はデータを二等分します。この例では観測データ数が奇数であるため、デー タセットは svar = 1 という識別子を持つデータ 51 件と svar = 2 という識別子を持つデータ 50 件に分割さ れたことがわかります。

nsplit(#)オプションを使用すると任意の数のサンプルに分割することができます。ここでは3分割を行ってみます。

• splitsample ダイアログ:

Main タブ: Generate sample ID variable: svar

Replace any variable with above name: \checkmark

Split into # random samples of equal size: • ($\tau 7 \pi \mu F$)

3

. splitsample, generate(svar, replace) nsplit(3)

. tabulate svar

. tabu	. tabulate svar			
	svar	Freq.	Percent	Cum.
	1	34	33.66	33.66
	2	33	32.67	66.34
	3	34	33.66	100.00
	Total	101	100.00	

^{*1} メニュー操作: Statistics \triangleright Summaries, tables, and tests \triangleright Frequency tables \triangleright One-way table

split(*numlist*) オプションを使うと分割するサンプルの割合(相対的サイズ)を指定することができます。 例えばサンプル 1, 2 はそれぞれ 25%、サンプル 3 は 50%の大きさとするには次のように操作します。

評価版では割愛しています。

▷ Example 2: クラスタ単位の分割

評価版では割愛しています。

▷ Example 3: balance() オプション

評価版では割愛しています。

▷ Example 4: 欠損値の扱い

評価版では割愛しています。