Title

stata.com

pkshape — Reshape (pharmacokinetic) Latin-square data

Description

`pkshape` reshapes data for use with `anova`, `pkcross`, and `pkequiv`; see [R] **anova**, [R] **pkcross**, and [R] **pkequiv**. Latin-square and crossover data are often organized in a manner that cannot be analyzed easily with Stata. `pkshape` reorganizes the data in memory for use in Stata.

`pkshape` is one of the pk commands. Please read [R] **pk** before reading this entry.

Quick start

Reshape data when string sequence variable `seq` = `TR` or `RT` for patients identified by `idvar` observed at `tvar1` and `tvar2`

```
pkshape idvar seq tvar1 tvar2
```

Same as above, but with numeric `seq` = 1 indicating `TR` and `seq` = 2 indicating `RT`

```
pkshape idvar seq tvar1 tvar2, order(TR RT)
```

Indicate that period 2 is a washout and the second treatment is administered in period 3

```
pkshape idvar seq tvar1 tvar2 tvar3, order(TOR ROT)
```

Menu

Statistics > Epidemiology and related > Other > Reshape pharmacokinetic Latin-square data

Syntax

```
pkshape id sequence period1 period2 [periodlist] [, options]
```

Variable *id* specifies unique subject identifiers. Variable *sequence* specifies the sequence (numeric or string) in which treatments were received. Variables *period1*, *period2*, and so on specify the pharmacokinetic measurements such as AUC in the corresponding periods.

options	Description
`order(`*string*`)`	apply treatments in specified order; required with numeric *sequence*
`outcome(`*newvar*`)`	name for outcome variable; default is `outcome(outcome)`
`treatment(`*newvar*`)`	name for treatment variable; default is `treatment(treat)`
`carryover(`*newvar*`)`	name for carryover variable; default is `carryover(carry)`
`sequence(`*newvar*`)`	name for sequence variable; default is `sequence(sequence)`
`period(`*newvar*`)`	name for period variable; default is `period(period)`

Options

`order(`*string*`)` specifies the order in which treatments were applied when generating the sequence, treatment, and carryover variables in the reorganized data. This option is required if the input sequence variable, *sequence*, is numeric. It is not allowed if *sequence* is a string variable. For crossover designs, any washout periods can be indicated with the number 0.

`outcome(`*newvar*`)` specifies the name for the outcome variable in the reorganized data. By default, `outcome(outcome)` is used.

`treatment(`*newvar*`)` specifies the name for the treatment variable in the reorganized data. By default, `treatment(treat)` is used.

`carryover(`*newvar*`)` specifies the name for the carryover variable in the reorganized data. By default, `carryover(carry)` is used.

`sequence(`*newvar*`)` specifies the name for the sequence variable in the reorganized data. By default, `sequence(sequence)` is used.

`period(`*newvar*`)` specifies the name for the period variable in the reorganized data. By default, `period(period)` is used.

Remarks and examples

stata.com

Often, data from a Latin-square experiment are naturally organized in a manner that Stata cannot manage easily. `pkshape` reorganizes Latin-square data so that they can be used with `anova` (see [R] **anova**) or any pk command. This includes the classic 2×2 crossover design commonly used in pharmaceutical research, as well as many other Latin-square designs. When using `pkshape`, newly created variables will automatically be labeled and system value labels will be created. The value label `_treatlbl` will be attached to the treatment and carrover variables, to indicate which treatment is applied in a given period and which treatment is being carried over from the previous period. The value label `_seqlbl` will be attached to the sequence variable, indicating the sequence of treatments.

`pkshape` expects the data to be organized in the same format as that produced by [R] **pkcollapse**—with variables representing time periods of the study.

▷ Example 1

Consider the example data published in Chow and Liu (2009, 71). There are 24 patients, 12 in each sequence. Sequence 1 is the reference formulation followed by the test formulation; sequence 2 is the test formulation followed by the reference formulation. The measurements reported are the $\mathrm{AUC}_{0,t_{\max}}$ for each patient and for each period.

```
. use https://www.stata-press.com/data/r18/chowliu
. list, sep(4)
```

	id	seq	period1	period2
1.	1	1	74.675	73.675
2.	4	1	96.4	93.25
3.	5	1	101.95	102.125
4.	6	1	79.05	69.45
5.	11	1	79.05	69.025
6.	12	1	85.95	68.7
7.	15	1	69.725	59.425
8.	16	1	86.275	76.125
9.	19	1	112.675	114.875
10.	20	1	99.525	116.25
11.	23	1	89.425	64.175
12.	24	1	55.175	74.575
13.	2	2	74.825	37.35
14.	3	2	86.875	51.925
15.	7	2	81.675	72.175
16.	8	2	92.7	77.5
17.	9	2	50.45	71.875
18.	10	2	66.125	94.025
19.	13	2	122.45	124.975
20.	14	2	99.075	85.225
21.	17	2	86.35	95.925
22.	18	2	49.925	67.1
23.	21	2	42.7	59.425
24.	22	2	91.725	114.05

Because the outcome for one person is in two different variables, the treatment that was applied to an individual is a function of the period and the sequence. To analyze this treatment using `anova`, all the outcomes must be in one variable, and each covariate must be in its own variable. To reorganize these data, use `pkshape`:

```
. pkshape id seq period1 period2, order(RT TR)
. sort seq id period
. list, sep(8)
```

	id	sequence	outcome	treat	carry	period
1.	1	RT	74.675	R	0	1
2.	1	RT	73.675	T	R	2
3.	4	RT	96.4	R	0	1
4.	4	RT	93.25	T	R	2
5.	5	RT	101.95	R	0	1
6.	5	RT	102.125	T	R	2
7.	6	RT	79.05	R	0	1
8.	6	RT	69.45	T	R	2
9.	11	RT	79.05	R	0	1
10.	11	RT	69.025	T	R	2
11.	12	RT	85.95	R	0	1
12.	12	RT	68.7	T	R	2
13.	15	RT	69.725	R	0	1
14.	15	RT	59.425	T	R	2
15.	16	RT	86.275	R	0	1
16.	16	RT	76.125	T	R	2
17.	19	RT	112.675	R	0	1
18.	19	RT	114.875	T	R	2
19.	20	RT	99.525	R	0	1
20.	20	RT	116.25	T	R	2
21.	23	RT	89.425	R	0	1
22.	23	RT	64.175	T	R	2
23.	24	RT	55.175	R	0	1
24.	24	RT	74.575	T	R	2
25.	2	TR	74.825	T	0	1
26.	2	TR	37.35	R	T	2
27.	3	TR	86.875	T	0	1
28.	3	TR	51.925	R	T	2
29.	7	TR	81.675	T	0	1
30.	7	TR	72.175	R	T	2
31.	8	TR	92.7	T	0	1
32.	8	TR	77.5	R	T	2
33.	9	TR	50.45	T	0	1
34.	9	TR	71.875	R	T	2
35.	10	TR	66.125	T	0	1
36.	10	TR	94.025	R	T	2
37.	13	TR	122.45	T	0	1
38.	13	TR	124.975	R	T	2
39.	14	TR	99.075	T	0	1
40.	14	TR	85.225	R	T	2
41.	17	TR	86.35	T	0	1
42.	17	TR	95.925	R	T	2
43.	18	TR	49.925	T	0	1
44.	18	TR	67.1	R	T	2
45.	21	TR	42.7	T	0	1
46.	21	TR	59.425	R	T	2
47.	22	TR	91.725	T	0	1
48.	22	TR	114.05	R	T	2

Now, the data are organized into separate variables that indicate each factor level for each of the covariates, so the data may be used with `anova` or `pkcross`; see [R] **anova** and [R] **pkcross**.

Initially, the output from `list` displayed sequence values 1 and 2, but now we see sequences `RT` and `TR` listed for the individuals. `pkshape` used the information we provided in the `order()` option to assign value labels to the numeric variables `sequence`, `treat`, and `carry`. Because we did not specify any new variable names, the default names were used.

◁

▷ Example 2

Consider the study of background music on bank teller productivity published in Kutner et al. (2005). The data are

Week	Monday	Tuesday	Wednesday	Thursday	Friday
1	18(D)	17(C)	14(A)	21(B)	17(E)
2	13(C)	34(B)	21(E)	16(A)	15(D)
3	7(A)	29(D)	32(B)	27(E)	13(C)
4	17(E)	13(A)	24(C)	31(D)	25(B)
5	21(B)	26(E)	26(D)	31(C)	7(A)

The numbers are the productivity scores, and the letters represent the treatment. We entered the data into Stata:

```
. use https://www.stata-press.com/data/r18/music, clear
(Background music and teller productivity)
. list
```

	id	seq	day1	day2	day3	day4	day5
1.	1	dcabe	18	17	14	21	17
2.	2	cbead	13	34	21	16	15
3.	3	adbec	7	29	32	27	13
4.	4	eacdb	17	13	24	31	25
5.	5	bedca	21	26	26	31	7

We reshape these data with pkshape:

```
. pkshape id seq day1 day2 day3 day4 day5
. sort id period
. list, sep(0)
```

	id	sequence	outcome	treat	carry	period
1.	1	dcabe	18	d	0	1
2.	1	dcabe	17	c	d	2
3.	1	dcabe	14	a	c	3
4.	1	dcabe	21	b	a	4
5.	1	dcabe	17	e	b	5
6.	2	cbead	13	c	0	1
7.	2	cbead	34	b	c	2
8.	2	cbead	21	e	b	3
9.	2	cbead	16	a	e	4
10.	2	cbead	15	d	a	5
11.	3	adbec	7	a	0	1
12.	3	adbec	29	d	a	2
13.	3	adbec	32	b	d	3
14.	3	adbec	27	e	b	4
15.	3	adbec	13	c	e	5
16.	4	eacdb	17	e	0	1
17.	4	eacdb	13	a	e	2
18.	4	eacdb	24	c	a	3
19.	4	eacdb	31	d	c	4
20.	4	eacdb	25	b	d	5
21.	5	bedca	21	b	0	1
22.	5	bedca	26	e	b	2
23.	5	bedca	26	d	e	3
24.	5	bedca	31	c	d	4
25.	5	bedca	7	a	c	5

Here the sequence variable is a string variable that specifies how the treatments were applied. The characters in this string variable are used to assign value labels to the newly created sequence, treat, and carry variables. We could now produce an ANOVA table:

```
. anova outcome sequence period treat
```

```
                   Number of obs =         25    R-squared     =  0.8666
                   Root MSE      =    3.96232    Adj R-squared =  0.7331
```

Source	Partial SS	df	MS	F	Prob>F
Model	1223.6	12	101.96667	6.49	0.0014
sequence	82	4	20.5	1.31	0.3226
period	477.2	4	119.3	7.60	0.0027
treat	664.4	4	166.1	10.58	0.0007
Residual	188.4	12	15.7		
Total	1412	24	58.833333		

◁

Example 3

Consider the Latin-square crossover example published in Kutner et al. (2005). The example is about apple sales given different methods for displaying apples.

Pattern	Store	Week 1	Week 2	Week 3
1	1	9(B)	12(C)	15(A)
	2	4(B)	12(C)	9(A)
2	1	12(A)	14(B)	3(C)
	2	13(A)	14(B)	3(C)
3	1	7(C)	18(A)	6(B)
	2	5(C)	20(A)	4(B)

We entered the data into Stata:

```
. use https://www.stata-press.com/data/r18/applesales, clear
(Display impact on apple sales)
. list, sep(2)
```

	id	seq	p1	p2	p3	square
1.	1	1	9	12	15	1
2.	2	1	4	12	9	2
3.	3	2	12	14	3	1
4.	4	2	13	14	3	2
5.	5	3	7	18	6	1
6.	6	3	5	20	4	2

Now, the data can be reorganized using descriptive names for the outcome variables.

```
. pkshape id seq p1 p2 p3, order(bca abc cab) seq(pattern) treat(displays)
. anova outcome pattern period displays id|pattern
```

```
                  Number of obs =         18    R-squared     =  0.9562
                  Root MSE      =    1.59426    Adj R-squared =  0.9069
```

Source	Partial SS	df	MS	F	Prob>F
Model	443.66667	9	49.296296	19.40	0.0002
pattern	.33333333	2	.16666667	0.07	0.9370
period	233.33333	2	116.66667	45.90	0.0000
displays	189	2	94.5	37.18	0.0001
id|pattern	21	3	7	2.75	0.1120
Residual	20.333333	8	2.5416667		
Total	464	17	27.294118		

These are the same results reported by Kutner et al. (2005).

References

Chow, S.-C., and J.-P. Liu. 2009. *Design and Analysis of Bioavailability and Bioequivalence Studies*. 3rd ed. Boca Raton, FL: Chapman and Hall/CRC.

Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li. 2005. *Applied Linear Statistical Models*. 5th ed. New York: McGraw–Hill/Irwin.

Also see

[R] **pk** — Pharmacokinetic (biopharmaceutical) data