Chapter 10 Useful SAS Procedures

Arthur Li

Restructuring Datasets Using the TRANSPOSE Procedure

```
PROC TRANSPOSE <DATA=input-data-set>
                   <DELIMITER=delimiter>
                   <LABEL=label>
                   <LET>
                   <NAME=name>
                   <OUT=output-data-set>
                   <PREFIX=prefix>
                   <SUFFIX=suffix>;
  BY <DESCENDING> variable-1
     <...<DESCENDING> variable-n>;
  COPY variable(s);
  ID variable;
  IDLABEL variable;
  VAR variable(s);
RUN;
```

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

```
data dat1;
   input s_name $ s_id $ e1-e3;
   label e1 = English1
        e2 = English2
        e3 = English3;
   datalines;
John A01 89 90 92
Mary A02 92 . 81
;
```

- Read in data
 - □S_Name & S_ID: character variables
 - ☐ E1 E3: numeric variables
 - ☐ E1 E3: have variable labels

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

- Only the PROC TRANSPOSE statement is used
- OUT=: specifies the name of the transposed data set

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92		81

Obs	_NAME_	_LABEL_	COL1	COL2	
1	e1	English1	89	92	
2	e2	English2	90	•	
3	e3	English3	92	81	

Dat1:

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

By default, without specifying the names of the transposing variables, all the numeric variables from the input data set are transposed.

COL1	COL2	
89	92	
90		
92	81	

Dat1:

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

Default variable names

Obs	_NAME_	_LABEL_	COL1	COL2
1	e1	English1	89	92
2	e2	English2	90	•
3	e3	English3	92	81

	S_NAME	S_ID	E1	E2	E 3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

Obs	_NAME_	_LABEL_	COL1	COL2	
1	e1	English1	89	92	
2	e2	English2	90	•	
3	e3	English3	92	81	

Contain labels

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

Obs	_NAME_	_LABEL_	COL1	COL2	
1	e1	English1		92	
2	e2	English2	90	•	
3	e3	English3	92	81	

Dat1:

	S_NAME	S_ID	E1	E2	E 3
1	John	A01	89	90	92
2	Mary	A02	92		81

Obs	_NAME_	_LABEL_	COL1	COL2
1	e1	English1	89	92
2	e2	English2	90	•
3	e3	English3	92	81

How to control the names of these variables?

	S_NAME	S_ID	E1	E2	E 3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

Obs	varname	labelname	score_1	score_2
1	e1	English1	89	92
2	e2	English2	90	•
3	e3	English3	92	81

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

```
proc transpose data=dat1
    out=dat1_out1
    name=varname
    label=labelname
    prefix=score_;
run;
```

Obs	varname	labelname	score_1	score_2	
1	e1	English1	89	92	
2	e2	English2	90	•	
3	e3	English3	92	81	

Dat1:

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

```
proc transpose data=dat1
    out=dat1_out1
    name=varname
    label=labelname
    prefix=score_;
run;
```

Obs	varname	labelname	score_1	score_2
1	e1	English1	89	92
2	e2	English2	90	•
3	e3	English3	92	81

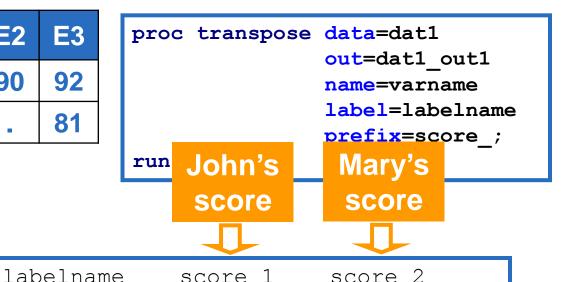
SUFFIX= option: attach a suffix

Dat1:

Obs

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

varname



How can you change the variable names to score_John and score_Mary?

score_r	SCOLE_Z	
89	92	
90	•	
92	81	

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

Obs	varname	labelname	score_ John	score_ Mary	
1	e1	English1	89	92	
2	e2	English2	90	•	
3	e3	English3	92	81	

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

Obs	varname	labelname	score_ John	score_ Mary
1	e1	English1	89	92
2	e2	English2	90	•
3	e3	English3	92	81

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

0bs	varname	labelname	John_A01	Mary_A02
1	e1	English1	89	92
2	e2	English2	90	-
3	e3	English3	92	81

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

Ol	os varname	labelname	John_A01	Mary_A02	
1	e1	English1	89	92	
2	e2	English2	90		
3	e3	English3	92	81	

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

```
proc contents data=dat1_out2;
run;
```

```
Alphabetic List of Variables and Attributes
Variable
                             Label
             Type
                      Len
John_A01
                             A01
             Num
                        8
             Num
Mary_A02
                             A02
labelname
             Char
                             LABEL OF FORMER VARIABLE
                       40
             Char
                             NAME OF FORMER VARIABLE
varname
```

Dat1:

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

By-variable: S_Name

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

Obs	s_name	_NAME_	_LABEL_	COL1	
1	John	e1	English1	89	
2	John	e2	English2	90	
3	John	e3	English3	92	
4	Mary	e1	English1	92	
5	Mary	e2	English2	•	
6	Mary	e3	English3	81	

Dat1:

S_NAME S_ID E1 E2 E3

John A01 89 90 92

Mary A02 92 . 81

Obs	s_name	_NAME_	_LABEL_	COL1
1 2	John John	e1 e2	English1 English2	89 90
3	John	e3	English3	92
4	Mary	e1	English1	92
5	Mary	e2	English2	.
6	Mary	e3	English3	81

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92		81

				1
Obs	s_name	_NAME_	_LABEL_	COL1
				\sim
1	John	e1	English1	89
2	John	e2	English2	90
3	John	e3	English3	92
4	Mary	e1	English1	92
5	Mary	e2	English2	
6	Mary	e3	English3	81

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92		81

Obs	s_name	s_id	varname	TEST	COL1
1	John John	A01	e1 e2	English1 English2	8 9 9 0
3	John		e3	English3	92
4 5	Mary Mary	A02	e1 e2	English1 English2	92
6	Mary		e3	English3	81

Dat1:

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

```
proc transpose data=dat1 sort
               out=dat1 out3
               name=varname
               label=TEST;
    by s name;
    copy s id;
run;
```

COL1	TEST	l varname	s_i	s_name	Obs
Y stater	Er COP	e1	A01	John	1
y variabl	Er CODY	e2		John	2
	ĿΥ	e3		John	3
n the inp	Er from	e1	A02	Mary	4
	E r	e2		Mary	5
set dire	Er Cata	e3		Mary	6
e trancr	to th				

ment: to the transposed data set

N = 2 from the input data

	S_NAME	S_ID	E1	E2	E 3
1	John	A01	89	90	92
2	Mary	A02	92		81

	0]	os s_nam	ne s_id	varname	TEST	COL1	
	1	John	A01	e1	English1	89	
101	ımber (~£		e2	English2	90	
ПС	illiber	Ji		e3	English3	92	
oh	servat	ions	A02	e1	English1	92	
				e2	English2	•	
ar	e copie	ed = 2		e3	English3	81	

Dat1:

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

ID statement: used to specify the variable from the input data set that contains the values to rename the transposed variables.

ıe	s_id	varname	TEST	COL1	
	A01	e1	English1	89	
		e2 e3	English2 English3	90 92	
	A02	e1	English1	92	
		e2	English2		
		e3	English3	81	

Where the ID Statement Does Not Work for Transposing BY-Groups

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

Obs	s_name	varname	TEST	A01	A02	
1	John	e1	English1	89	•	
3	John John	e2 e3	English2 English3	90 92	•	
4 5	Mary Mary	e1 e2	English1 English2	•	92	
6	Mary	e3	English3	•	81	

Where the ID Statement Does Not Work for Transposing BY-Groups

Dat1:

column.

	S_NAME	S_ID	E1	E2	E3
1	John	A01	89	90	92
2	Mary	A02	92	•	81

	Obs	s name	varname	TEST	A01	A02
Yo	u are using	the ID				
			e1	English1	89	•
va	riable (cont	ains two	e2	English2	90	•
va	lues) to nar	ne the	e3	English3	92	•
	• • • • • • • • • • • • • • • • • • •		e1	English1	•	92
tra	nsposed va	ariable	e2	English2	•	•
tha	at was supp	osed to	e3	English3	٠	81
oc	cupy only c	ne				

	S_NAME	S_ID	Exam	Score
1	John	A01	1	89
2	John	A01	2	90
3	John	A01	3	92
4	Mary	A02	1	92
5	Mary	A02	3	81

Obs	s_name	_NAME_	COL1	COL2	COL3
1	John	score	89	90	92
2	MARY	SCORE	92	81	•

	S_NAME	S_ID	Exam	Score
1	John	A01	1	89
	John	A01	2	90
3	John	A01	3	92
4	Mary	A02	1	92
5	Mary	A02	3	81

Obs	s_name	_NAME_	COL1	COL2	COL3	
1 2	John MARY	score SCORE	89 92	90 81	92	2

	S_NAME	S_ID	Exam	Score
1	John	A01	1	89
2	John	A01	2	90
3	John	A01	3	92
4	Mary	A02	1	92
5	Mary	A02	3	81

Obs	s_name	_NAME_	COL1	COL2	COL3
1	John	score	89	90	92
2	MARY	SCORE	92	81	•



	S_NAME	S_ID	Exam	Score
1	John	A01	1	89
2	John	A01	2	90
3	John	A01	3	92
4	Mary	A02	1	92
5	Mary	A02	3	81

```
proc transpose data=dat2_sort
    out=dat2_out2(drop=_name_)
    prefix=test_;
    var score;
    by s_name;
    id exam;
run;
```

Obs	s_name	test_1	test_2	test_3
1	John	89	90	92
2	MARY	92		81

Handling Duplicated Observations by Using the LET Option

Dat3:

	S_NAME	S_ID	Exam	Score
1	John	A01	1	89
2	John	A01	2	90
3	John	A01	3	92
4	John	A01	3	95
5	Mary	A02	1	92
6	Mary	A02	3	81
7	Mary	A02	3	85

LET option: keep the last occurrence of a particular ID value within either the entire data set or a BY group.

```
proc transpose data=dat3_sort
    out=dat3_out(drop=_name_)
    prefix=test_
    let;
    var score;
    by s_name;
    id exam;
run;
```

Handling Duplicated Observations by Using the LET Option

Dat3:

	S_NAME	S_ID	Exam	Score
1	John	A01	1	89
2	John	A01	2	90
3	John	A01	3	92
4	John	A01	3	95
5	Mary	A02	1	92
6	Mary	A02	3	81
7	Mary	A02	3	85

```
proc sort data=dat3
         out=dat3 sort;
   by s name exam score;
run;
proc transpose data=dat3 sort
     out=dat3 out(drop= name )
     prefix=test
     let;
               Keep the
    var score;
               maximum
   by s_name;
    id exam;
               score within
run;
               each exam
```

Obs	s_name	test_1	test_2	test_3
1	John	89	90	95
2	Mary	92	•	85

Handling Duplicated Observations by Using the LET Option

Dat3:

	S_NAME	S_ID	Exam	Score
1	John	A01	1	89
2	John	A01	2	90
3	John	A01	3	92
4	John	A01	3	95
5	Mary	A02	1	92
6	Mary	A02	3	81
7	Mary	A02	3	85

```
proc sort data=dat3
          out=dat3 sort;
   by s name exam
       descending score;
run;
proc transpose data=dat3 sort
      out=dat3 out(drop= name )
      prefix=test
      let;
               Keep the
    var score;
                minimum
    by s name;
    id exam;
                score within
run;
```

each exam

Obs	s_name	test_1	test_2	test_3	
1 2	John Mary	89 92	90	92 81	

Dat4_Transpose:

Dat4:

	S_NAME	E1	E2	E3	M1	M2	M3
1	John	89	90	92	78	89	90
2	Mary	92		81	76	91	89

	Test_num	John_E	John_M	Mary_E	Mary_M
1	1	89	78	92	76
2	2	90	89		91
3	3	92	90	81	89

To transpose from
Dat4 → Dat4_Transpose,
we need a "transitional"
data set.

Dat4_Transpose:

Dat4:

	S_NAME	E1	E2	E3	M1	M2	M3
1	John	89	90	92	78	89	90
2	Mary	92	•	81	76	91	89

	Test_num	John_E	John_M	Mary_E	Mary_M
1	1	89	78	92	76
2	2	90	89		91
3	3	92	90	81	89

To transpose from
Dat4 → Dat4_Transpose,
we need a "transitional"
data set.

	S_NAME	Score	Test_num	Class
1	John	89	1	E
2	John	78	1	M
3	Mary	92	1	E
4	Mary	76	1	M
5	John	90	2	E
6	John	89	2	M
7	Mary	•	2	E
8	Mary	91	2	M
9	John	92	3	E
10	John	90	3	M
11	Mary	81	3	E
12	Mary	89	3	M

Dat4_Transpose:

Dat4:

	S_NAME	E1	E2	E3	M1	M2	М3
1	John	89	90	92	78	89	90
2	Mary	92		81	76	91	89

	Test_num	John_E	John_M	Mary_E	Mary_M
1	1	89	78	92	76
2	2	90	89		91
3	3	92	90	81	89

To transpose from
Dat4 → Dat4_Transpose,
we need a "transitional"
data set.



	S_NAME	Score	Test_num	Class
1	John	89	1	Ш
2	John	78	1	M
3	Mary	92	1	Ш
4	Mary	76	1	M
5	John	90	2	Ш
6	John	89	2	M
7	Mary	•	2	Ш
8	Mary	91	2	M
9	John	92	3	Е
10	John	90	3	M
11	Mary	81	3	Ш
12	Mary	89	3	M

Dat4_Transpose:

Dat4:

	S_NAME	E1	E2	E 3	M1	M2	M3
1	John	89	90	92	78	89	90
2	Mary	92		81	76	91	89

	Test_num	John_E	John_M	Mary_E	Mary_M
1	1	89	78	92	76
2	2	90	89		91
3	3	92	90	81	89



	S_NAME	_NAME_	COL1
1	John	E1	89
2	John	E2	90
3	John	E 3	92
4	John	M1	78
5	John	M2	89
6	John	М3	90
7	Mary	E1	92
8	Mary	E2	
9	Mary	E 3	81
10	Mary	M1	76
11	Mary	M2	91
12	Mary	М3	89

	S_NAME	Score	Test_num	Class
1	John	89	1	Е
2	John	78	1	M
3	Mary	92	1	ш
4	Mary	76	1	M
5	John	90	2	Ш
6	John	89	2	M
7	Mary		2	E
8	Mary	91	2	M
9	John	92	3	Е
10	John	90	3	M
11	Mary	81	3	Е
12	Mary	89	3	M

Dat4_Transpose:

Dat4:

	S_NAME	E1	E2	E3	M1	M2	М3
1	John	89	90	92	78	89	90
2	Mary	92		81	76	91	89

	Test_num	John_E	John_M	Mary_E	Mary_M
1	1	89	78	92	76
2	2	90	89		91
3	3	92	90	81	89



	S_NAME	_NAME_	COL1	
1	John	E <mark>1</mark>	89	
2	John	E <mark>2</mark>	90	
3	John	E <mark>3</mark>	92	
4	John	IV 1	78	
5	John	V 2	89	
6	John	N 3	90	
7	Mary	E <mark>1</mark>	92	
8	Mary	E2		
9	Mary	E3	81	
10	Mary	V 1	76	
11	Mary	V 2	91	
12	Mary	V 3	89	

		S_NAME	Score	Test_num	Class
1		John	89	1	E
2		John	78	1	M
3		Mary	92	1	E
4		Mary	76	1	M
5		John	90	2	E
6		John	89	2	M
7	,	Mary		2	E
8		Mary	91	2	M
9		John	92	3	E
10	0	John	90	3	M
11	1	Mary	81	3	E
12	2	Mary	89	3	M

Dat4_Transpose:

Dat4:

	S_NAME	E1	E2	E3	M1	M2	M3
1	John	89	90	92	78	89	90
2	Mary	92		81	76	91	89

	Test_num	John_E	John_M	Mary_E	Mary_M
1	1	89	78	92	76
2	2	90	89		91
3	3	92	90	81	89

	S_NAME	_NAME_	COL1
1	John	E1	89
2	John	E2	90
3	John	E 3	92
4	John	M1	78
5	John	M2	89
6	John	М3	90
7	Mary	E1	92
8	Mary	E2	
9	Mary	E 3	81
10	Mary	M1	76
11	Mary	M2	91
12	Mary	М3	89

Step1:

Dat4_Transpose:

Dat4:

34 (1.								
	S_NAME	E1	E2	E3	M1	M2	M3	
1	John	89	90	92	78	89	90	
2	Mary	92	•	81	76	91	89	

	Test_num	John_E	John_M	Mary_E	Mary_M
1	1	89	78	92	76
2	2	90	89		91
3	3	92	90	81	89

	S_NAME	_NAME_	COL1	Test_num	Class
1	John	E1	89	1	Е
2	John	E2	90	2	Е
3	John	E 3	92	3	Е
4	John	M1	78	1	M
5	John	M2	89	2	M
6	John	М3	90	3	M
7	Mary	E1	92	1	Е
8	Mary	E2		2	Е
9	Mary	E 3	81	3	Е
10	Mary	M1	76	1	M
11	Mary	M2	91	2	M
12	Mary	М3	89	3	M

Step2:

```
data dat4_out1a;
    set dat4_out1;
    test_num=substr(_name__,2);
    class=substr(_name__,1,1);
run;
```

Dat4_Transpose:

Dat4:

34(1.							
	S_NAME	E1	E2	E3	M1	M2	M3
1	John	89	90	92	78	89	90
2	Mary	92		81	76	91	89

	Test_num	John_E	John_M	Mary_E	Mary_M
1	1	89	78	92	76
2	2	90	89		91
3	3	92	90	81	89

	S_NAME	_NAME_	COL1	Test_num	Class
1	John	E1	89	1	E
2	John	M1	78	1	M
3	Mary	E1	92	1	Е
4	Mary	M1	76	1	M
5	John	E2	90	2	E
6	John	M2	89	2	M
7	Mary	E2		2	E
8	Mary	M2	91	2	M
9	John	E 3	92	3	Е
10	John	M3	90	3	M
11	Mary	E 3	81	3	Е
12	Mary	М3	89	3	M

Step3:

```
proc sort data=dat4_out1a
          out=dat4_sort2;
    by test_num s_name;
run;
```

Dat4_Transpose:

Dat4:

	S_NAME	E1	E2	E3	M1	M2	М3
1	John	89	90	92	78	89	90
2	Mary	92		81	76	91	89

	Test_num	John_E	John_M	Mary_E	Mary_M
1	1	89	78	92	76
2	2	90	89		91
3	3	92	90	81	89

	S_NAME	_NAME_	COL1	Test_num	Class
1	John	E1	89	1	Е
2	John	M1	78	1	M
3	Mary	E1	92	1	Е
4	Mary	M1	76	1	M
5	John	E2	90	2	Е
6	John	M2	89	2	M
7	Mary	E2	•	2	E
8	Mary	M2	91	2	M
9	John	E 3	92	3	Е
10	John	M3	90	3	M
11	Mary	E 3	81	3	Е
12	Mary	М3	89	3	M

Step4:

```
proc transpose
    data=dat4_sort2
    out=dat4_out2(drop=_name_)
    delimiter=_;
    by test_num;
    var col1;
    id name class;
run;
```