



# ***RESHAPING & COMBINING SAS® DATA SETS***

---

1

- SAS® offers many ways to analyze and present data
- data may not be stored in a format required for a particular analysis and/or mode of presentation
- SAS tools that reorganize data must be used first

those that reshape (turn variables into observations and vice-versa)

those that combine data sets

---



## ***RESHAPING & COMBINING SAS® DATA SETS***

---

2

- reshaping and combining of data sets may be done in several different ways
- focus on...

PROC TRANSPOSE to reshape data sets

data step match-merge to combine data sets

---



## ***RESHAPING & COMBINING SAS® DATA SETS***

---

**3**

- PROC TRANSPOSE options (PREFIX) and statements (BY, ID, VAR)
- various types of match-merge situations (one-to-one, one-to-many, many-to-many)

how to control the results of match-merge using the IN= data set option

---



# ***RESHAPING & COMBINING SAS® DATA SETS***

---

**4**

- Alternative methods...

arrays to reshape data

PROC SQL to combine data sets

---



# ***RESHAPING & COMBINING SAS® DATA SETS***

---

**5**

## **RESHAPING DATA**

- analysis across observations - SAS procedures
  - analysis within observations - SAS functions
  - occasions when neither a procedure nor a function will suffice
  - data may be arranged in such a way that the only way to complete a given task is to first rearrange the data set
-



## ***RESHAPING & COMBINING SAS® DATA SETS***

---

6

### ■ **Problem**

SAS data set with observations that contain up to five occurrences of a 3-digit medical diagnosis

each observation represents one person...

create a table that shows how often each diagnosis occurs within the data set

---



## ***RESHAPING & COMBINING SAS® DATA SETS***

---

7

**datalines;**

**01 647 641 650 428**

**02 428 416**

**03 642 674 648**

**04 641 416 648 647 641**

**;**  
**/**

- PROC FREQ on each of the five diagnoses and add the results of the five tables
  - create a new data set that --- one variable (DIAG), one observation for each diagnosis in the original data set
-



## ***RESHAPING & COMBINING SAS® DATA SETS***

---

**8**

- Examples...variables-to-observations...
    - 1: create a data set with all character variables
    - 2: data step approach
    - 3: TRANSPOSE with no options, VAR statement
    - 4: add a BY statement
    - 5: add data set options to OUT= data set
    - 6: create a data set with all numeric variables
    - 7: TRANSPOSE with no options
-





## ***RESHAPING & COMBINING SAS® DATA SETS***

---

9

### ■ Problem...

SAS data set - observations contain an account number, a month (in the form of 1 for January, 2 for February, etc.), and a dollar value indicating the amount deposit in money market account

create a report showing the monthly deposits for each person in your data set (as indicated by the account number)

---



## *RESHAPING & COMBINING SAS® DATA SETS*

---

10

account	month	deposit
01	1	100
01	4	50
01	6	200
02	2	50
02	3	100
03	1	50
03	2	50
03	3	50
03	4	50
03	5	50
03	6	50

- easier to create the report if the data set had one observation per person with all the deposits in that observation
-



## ***RESHAPING & COMBINING SAS® DATA SETS***

---

**11**

### ■ Examples...observations-to variables...

8: create a data set

9: TRANSPOSE with no options

10: add a BY statement

11: add data set option to OUT= data set, a VAR statement, PREFIX option

12: add an ID statement

13: change MONTH to a character variable

14: use MONTH as an ID variable, no PREFIX option

---



## ***RESHAPING & COMBINING SAS® DATA SETS***

---

**12**

- Examples...alternative...

15:     ARRAY, variables-to-observations

16:     ARRAY, observations-to-variables



## **COMBINING DATA**

- stack observations...

concatenate, interleave, PROC APPEND

- combine observations...

one-to-one merge, matched-merge, update, modify

- PROC SQL - either stack or combine observations
-



## ***RESHAPING & COMBINING SAS® DATA SETS***

---

14

### ■ **Problem (one-to-one)...**

two SAS data sets (one with information collected on individuals in the month of January, the other with similar information collected in February)

create a new data set with one observation per person that contains all the information from both months

each data set contains a weight for each person,  
compute the weight change from January to February  
examples

---



## ***RESHAPING & COMBINING SAS® DATA SETS***

---

**15**

### ■ Examples...one-to-one...

- 1: create two data sets, JAN and FEB
  - 2: match-merge data sets
  - 3: MERGENOBY option
  - 4: add a RENAME data set option
  - 5: add IN= data set option
  - 6: create multiple data sets (matched, unmatched JAN, unmatched FEB)
-



## ***RESHAPING & COMBINING SAS® DATA SETS***

---

16

- **Problem (one-to-many, many-to-one)...**

two SAS data sets, one has demographic data on a group of individuals while the other has medical data

create a single data set containing the demographic and medical data for each a single observation

---





## ***RESHAPING & COMBINING SAS® DATA SETS***

---

**17**

- Examples...one-to-many (many-to-one)...
  - 1: create two data sets, MEDICAL and DEMOGRAPHIC
  - 2: match-merge data sets
  - 3: add an IF statement to restrict matches
-



## ***RESHAPING & COMBINING SAS® DATA SETS***

---

**18**

### ■ **Problem (many-to-many)...**

two SAS data sets, one has a name and an age while the other has a name, an age, and a heart rate

create a single data set containing the name, age from both data sets, and the heart rate

---



## ***RESHAPING & COMBINING SAS® DATA SETS***

---

**19**

- Examples...many-to-many...
    - 1: create two data sets, DATASET1 and DATASET2
    - 2: match-merge data sets with an IF statement to restrict matches
    - 3: use PROC SQL to get correct results (INNER and FULL joins)
-



# RESHAPING & COMBINING SAS® DATA SETS

20

OBS	DATASET1		DATASET2	AGE WITHIN ONE YEAR
1	ADAMS/20	>-----<	ADAMS/21	YES
	BROWN/21	>-----<	BROWN/15	NO
	BROWN/30	>-----<	BROWN/21	NO
		-----<	BROWN/40	NO
	JONES/45	>-----<	JONES/48	NO
	JONES/46	>-----		NO
2	JONES/47	>-----		YES
3	NO MATCH		KELLY/57	UNKNOWN
	LAWRENCE/10	>-----<	LAWRENCE/16	NO
	LAWRENCE/14	>-----<	LAWRENCE/20	NO
	LAWRENCE/16	>-----		NO
	SMITH/50	>-----<	SMITH/30	NO
4		-----<	SMITH/50	YES
5	WALTERS/29		NO MATCH	UNKNOWN