STAT40840: Data programming with SAS
Laura Kirwan

Lecture 7

Lecture 7 – Part 2: Combining Data Sets

7.3 Concatenating Data Sets

7.4 Merging Data Sets One-to-One

7.5 Merging Data Sets One-to-Many
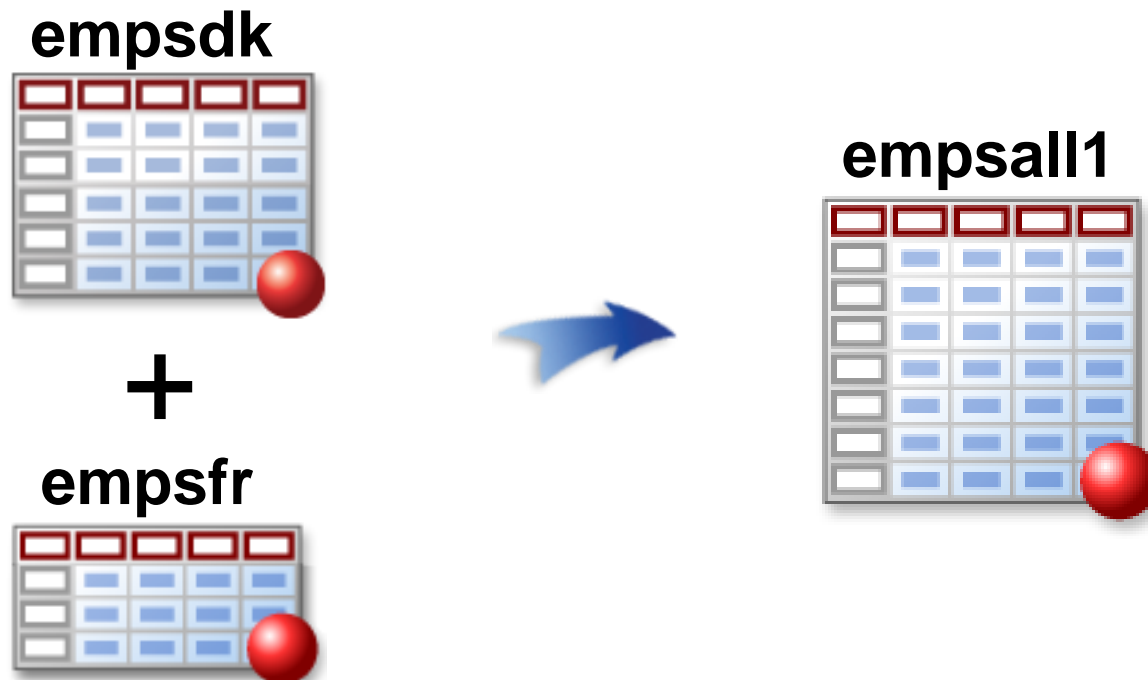
7.6 Merging Data Sets with Nonmatches

Objectives - Part 3

- Concatenate two or more SAS data sets. Use the SET statement in a DATA step.

- Change the names of variables. Use the RENAME= data set option.

Scenario

You were asked to combine the data sets that contain information about Orion Star employees from Denmark and France into a new data set.

empsdk

+

empsfr

empsall1

Considerations

Concatenate like-structured data sets, **empsdk** and **empsfr**, to create a new data set named **empsall1**.

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark
Pierre	M	France
Sophie	F	France

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

Both data sets contain the same variables.

Using a DATA Step

Use a DATA step to concatenate the data sets.
List the data sets in the SET statement.

```
data empsall1;
   set empsdk empsfr;
run;
```

SET *SAS-data-set1 SAS-data-set2* . . .;

- – The SET statement reads observations from each data set in the order in which they are listed.
- – Any number of data sets can be included in the SET statement.

L7_D9.sas

Compilation

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country

empsall1

First	Gender	Country

L7_D9.sas

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

Initialize PDV

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country

empsall1

First	Gender	Country

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country
Lars	M	Denmark

empsall1

First	Gender	Country

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	Country
Lars	M	Denmark

empsall1

First	Gender	Country
Lars	M	Denmark

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country
Lars	M	Denmark

empsall1

First	Gender	Country
Lars	M	Denmark

Data set variables are not reinitialized.

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country
Kari	F	Denmark

empsall1

First	Gender	Country
Lars	M	Denmark

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	Country
Kari	F	Denmark

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country
Kari	F	Denmark

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark

PDV is not reinitialized.

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
     set empsdk empsfr;
run;
```

PDV

First	Gender	Country
Jonas	M	Denmark

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	Country
Jonas	M	Denmark

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country
Jonas	M	Denmark

PDV is not reinitialized.

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
~~nas~~	M	Denmark

EOF

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country
Jonas	M	Denmark

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

UCD School of Mathematics and Statistics

www.ucd.ie/mathstat

...

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
     set empsdk empsfr;
run;
```

Reinitialize PDV

PDV

First	Gender	Country

PDV is reinitialized before processing the next data set.

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country
Pierre	M	France

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
   set emps
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	Country
Pierre	M	France

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark
Pierre	M	France

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country
Pierre	M	France

PDV is not reinitialized.

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark
Pierre	M	France

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country
Sophie	F	France

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark
Pierre	M	France

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
   set empsdk empsfr;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	Country
Sophie	F	France

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark
Pierre	M	France
Sophie	F	France

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country
Sophie	F	France

PDV is not reinitialized.

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark
Pierre	M	France
Sophie	F	France

Execution

empsdk

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

empsfr

First	Gender	Country
Pierre	M	France
Sophie	F	France

EOF

```
data empsall1;
    set empsdk empsfr;
run;
```

PDV

First	Gender	Country
Sophie	F	France

empsall1

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark
Pierre	M	France
Sophie	F	France

Viewing the Log

```
145   data empsall1;
146      set empsdk empsfr;
147   run;

NOTE: There were 3 observations read from the data set WORK.EMPSDK.
NOTE: There were 2 observations read from the data set WORK.EMPSFR.
NOTE: The data set WORK.EMPSALL1 has 5 observations and 3
variables.
```

Unlike-Structured Data Sets

Concatenate **empscn** and **empsjp** to create a new data set named **empsall2**.

empscn

First	Gender	Country
Chang	M	China
Li	M	China
Ming	F	China

empsjp

First	Gender	Region
Cho	F	Japan
Tomi	M	Japan

The data sets do not contain the same variables.

```
data empsall2;
    set empscn empsjp;
run;
```

L7_D10.sas

Exercise 6

How many variables will be in **empsall2** after concatenating **empscn** and **empsjp**?

empscn

First	Gender	Country
Chang	M	China
Li	M	China
Ming	F	China

empsjp

First	Gender	Region
Cho	F	Japan
Tomi	M	Japan

```
data empsall2;
    set empscn empsjp;
run;
```

Compilation

empscn

First	Gender	Country
Chang	M	China
Li	M	China
Ming	F	China

empsjp

First	Gender	Region
Cho	F	Japan
Tomi	M	Japan

```
data empsall2;
    set empscn empsjp;
run;
```

PDV

First	Gender	Country

L7_D10.sas

Compilation

empscn

First	Gender	Country
Chang	M	China
Li	M	China
Ming	F	China

empsjp

First	Gender	Region
Cho	F	Japan
Tomi	M	Japan

```
data empsall2;
    set empscn empsjp;
run;
```

PDV

First	Gender	Country	Region

Final Results

empsall2

First	Gender	Country	Region
Chang	M	China	
Li	M	China	
Ming	F	China	
Cho	F		Japan
Tomi	M		Japan

- **Region** has missing values due to PDV initialization.
- **Country** has missing values due to PDV reinitialization before processing the second data set.

Scenario

Rename variables in one or more data sets to align columns.

empscn

First	Gender	Country
Chang	M	China
Li	M	China
Ming	F	China

empsjp

First	Gender	Region
Cho	F	Japan
Tomi	M	Japan

Rename **Region** to **Country**.

RENAME= Data Set Option

The *RENAME= data set option* changes the name
of a variable.

```
data empsall2;
   set empscn empsjp(rename=(Region=Country));
run;
```

SAS-data-set **(RENAME=(**old-name-1**=**new-name-1
                           old-name-2**=**new-name-2
                                 ...
                          old-name-n**=**new-name-n**))**

- The RENAME= option must be specified in parentheses immediately after the appropriate SAS data set name.

- The name change affects the PDV and the output data set. It has no effect on the input data set.

RENAME= Data Set Option

• Multiples variables can be renamed in one or more data sets.

```
set empscn(rename=(Country=Region))
    empsjp;
```

```
set empscn(rename=(First=Fname
                   Country=Region))
    empsjp(rename=(First=Fname));
```

```
set empscn
    empsjp(rename=(Region=Country));
```

Exercise 7

Which statement has correct syntax?

a.
```
set empscn(rename(Country=Location))
    empsjp(rename(Region=Location));
```

b.
```
set empscn(rename=(Country=Location))
    empsjp(rename=(Region=Location));
```

c.
```
set empscn rename=(Country=Location)
    empsjp rename=(Region=Location);
```

Compilation

empscn

First	Gender	Country
Chang	M	China
Li	M	China
Ming	F	China

empsjp

First	Gender	Region
Cho	F	Japan
Tomi	M	Japan

```
data empsall2;
   set empscn empsjp(rename=(Region=Country));
run;
```

PDV

First	Gender	Country

L7_D11.sas

Compilation

empscn

First	Gender	Country
Chang	M	China
Li	M	China
Ming	F	China

empsjp

First	Gender	Region
Cho	F	Japan
Tomi	M	Japan

```
data empsall2;
   set empscn empsjp(rename=(Region=Country));
run;
```

PDV

First	Gender	Country

Compilation

empscn

First	Gender	Country
Chang	M	China
Li	M	China
Ming	F	China

empsjp

First	Gender	Region
Cho	F	Japan
Tomi	M	Japan

```
data empsall2;
   set empscn empsjp(rename=(Region=Country));
run;
```

PDV

First	Gender	Country

Compilation

empscn

First	Gender	Country
Chang	M	China
Li	M	China
Ming	F	China

empsjp

First	Gender	Region
Cho	F	Japan
Tomi	M	Japan

```
data empsall2;
    set empscn empsjp(rename=(Region=Country));
run;
```

PDV

First	Gender	Country

Compilation

empscn

First	Gender	Country
Chang	M	China
Li	M	China
Ming	F	China

empsjp

First	Gender	Region
Cho	F	Japan
Tomi	M	Japan

```
data empsall2;
    set empscn empsjp(rename=(Region=Country));
run;
```

PDV

First	Gender	Country

Final Results

The **Region** values are stored in **Country**.

empsall2

First	Gender	Country
Chang	M	China
Li	M	China
Ming	F	China
Cho	F	Japan
Tomi	M	Japan

Lecture 7 – part 2: Combining Data Sets

7.3 Concatenating Data Sets

7.4 Merging Data Sets One-to-One

7.5 Merging Data Sets One-to-Many

7.6 Merging Data Sets with Nonmatches

Objectives

- – Prepare data sets for merging. Use the SORT procedure.
- – Merge SAS data sets one-to-one based on a common variable.

Match-Merging

A	B	C		C	D	E
		1	↔	1		
		2	↔	2		
		3	↔	3		
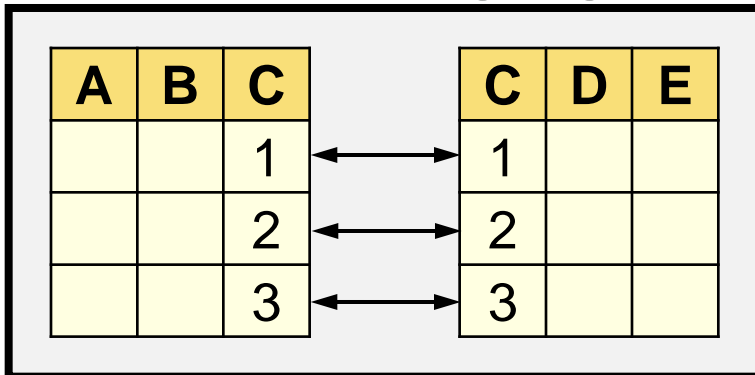
One-to-One

A single observation in one data set is related to exactly one observation in another data set based on the values of one or more selected variables.

A	B	C		C	D	E
		1		1		
		2		1		
				2		
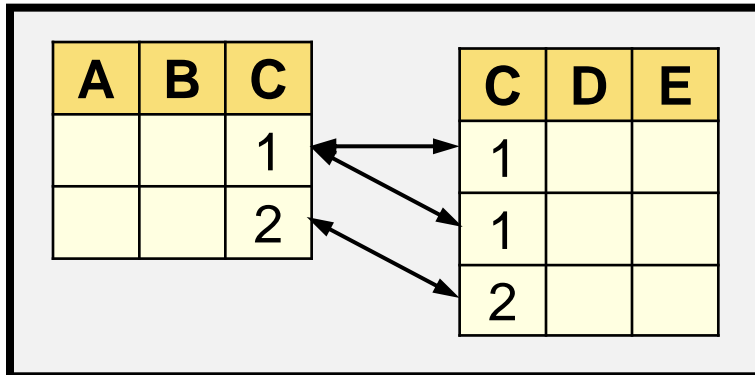
One-to-Many

A single observation in one data set is related to more than one observation in another data set based on the values of one or more selected variables.

A	B	C		C	D	E
		1		2		
		2		3		
		4	↔	4		
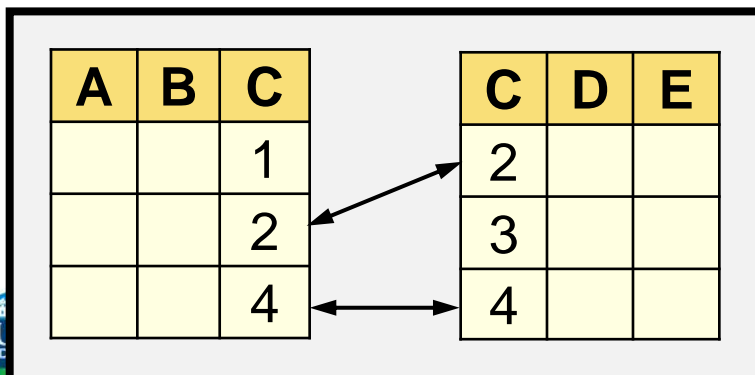
Nonmatches

At least one observation in one data set is unrelated to any observation in another data set based on the values of one or more selected variables.

Match-Merging: Sorting the Data Sets

The data sets in a match-merge must be sorted by the common variable or variables that are being matched.

```
PROC SORT  DATA=input-SAS-data-set
              <OUT=output-SAS-data-set>;
    BY <DESCENDING> by-variable(s);
RUN;
```

Exercise 8

Which of the following BY statements correctly sorts by descending **salary** within **gender**?

a. by descending salary within gender;

b. by descending salary gender;

c. by gender descending salary;

d. by gender salary descending;

Scenario

Merge the Australian employee data set with a phone data set to obtain each employee's home phone number. Store the results in a new data set.

empsau

First	Gender	EmpID

+

phoneh

EmpID	Phone

empsauh

First	Gender	EmpID	Phone

Match-Merging

The *MERGE statement* in a DATA step joins observations from two or more SAS data sets into single observations.

```
data empsauh;
    merge empsau phoneh;
    by EmpID;
run;
```

MERGE *SAS-data-set1 SAS-data-set2 . . .*;
BY <DESCENDING> *BY-variable(s)*;

A *BY statement* indicates a match-merge and lists the variable or variables to match.

L7_D12.sas

MERGE and BY Statements

Requirements for match-merging:

- Two or more data sets are listed in the MERGE statement.
- The variables in the BY statement must be common to all data sets.
- The data sets must be sorted by the variables listed in the BY statement.
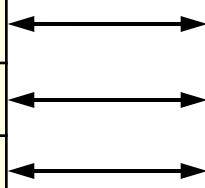
One-to-One Merge

One observation in **empsau** matches exactly one observation in **phoneh**.

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phoneh

EmpID	Phone
121150	+61(2)5555-1793
121151	+61(2)5555-1849
121152	+61(2)5555-1665

The data sets are sorted by **EmpID**.

Final Results

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phoneh

EmpID	Phone
121150	+61(2)5555-1793
121151	+61(2)5555-1849
121152	+61(2)5555-1665
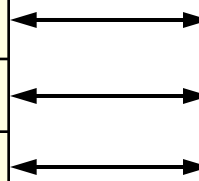
empsauh

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1793
Kylie	F	121151	+61(2)5555-1849
Birin	M	121152	+61(2)5555-1665

Exercise 9

- Complete program **L7_E9.sas t**o match-merge the sorted SAS data sets referenced in the PROC SORT steps.

- Submit the program. Correct and resubmit it, if necessary.

- What are the modified, completed statements?

Lecture 7 part 2: Combining Data Sets

7.3 Concatenating Data Sets

7.4 Merging Data Sets One-to-One

7.5 Merging Data Sets One-to-Many

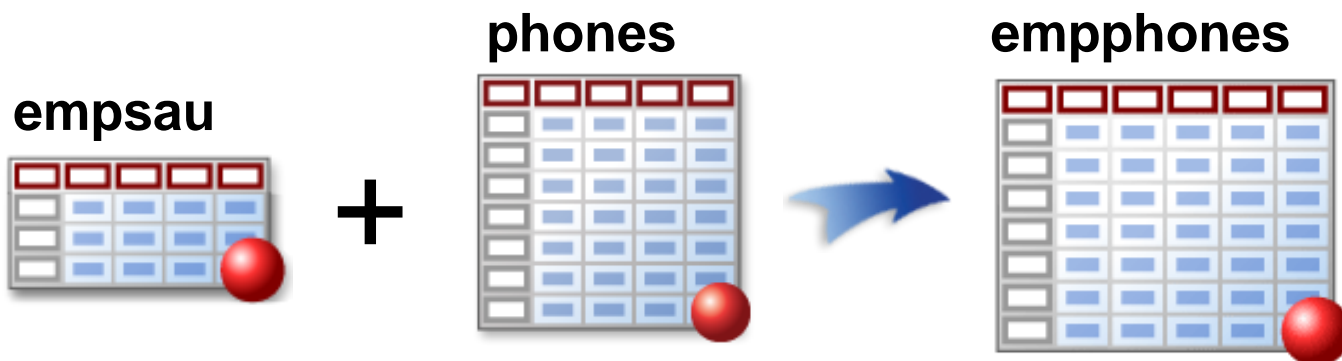**7.6 Merging Data Sets with Nonmatches**

Objectives

– Merge SAS data sets one-to-many based on a common variable.

Scenario

Merge the Australian employee information data set with the **phones** data set to obtain the phone numbers for each employee.

empsau + phones → empphones

Considerations

In this one-to-many merge, one observation in **empsau** matches one or more observations in **phones**.

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

The data sets are sorted by **EmpID**.

**UCD School of Mathematics and Statistics**

Match-Merging

Merge the two data sets by **EmpID** and create a new data set named **empphones**.

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

L7_D13.sas

Match-Merging

BY variable

BY group

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152
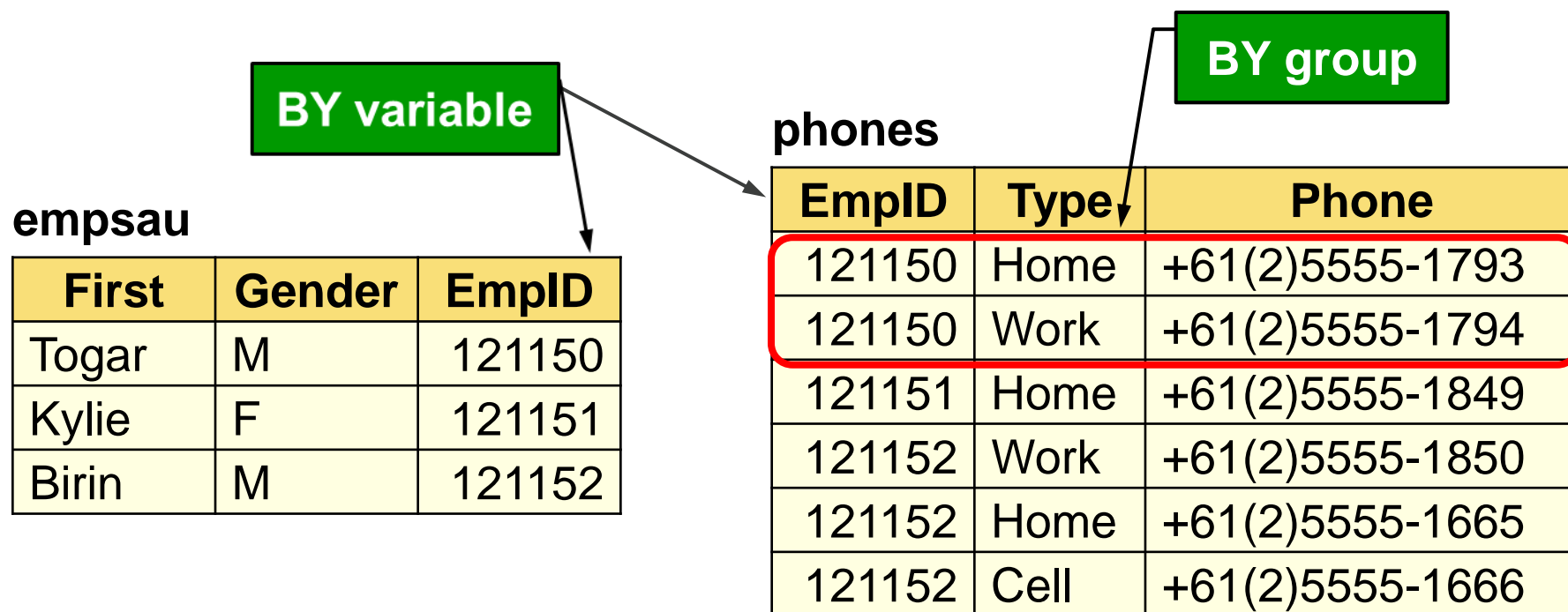
- The common variable is called the *BY variable*.
- The value of the BY variable is the *BY value*.
- A group of observations with the same BY value is a *BY group*.

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Initialize PDV

L7_D13.sas

PDV

First	Gender	EmpID	Type	Phone
		.		

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Do the **EmpID** values match?

Yes

PDV

First	Gender	EmpID	Type	Phone
		.		

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Reads both observations into the PDV.

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Home	+61(2)5555-1793

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Home	+61(2)5555-1793

...

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

PDV

Data set variables are not reinitialized.

First	Gender	EmpID	Type	Phone
Togar	M	121150	Home	+61(2)5555-1793

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Do the **EmpID** values match?

No

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Home	+61(2)5555-1793

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Did **EmpID** change?

No

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Home	+61(2)5555-1793

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Reads the matching observation into the PDV.

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Work	+61(2)5555-1794

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Work	+61(2)5555-1794

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Do the **EmpID** values match?

Yes

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Work	+61(2)5555-1794

...

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Did **EmpID** change?

Yes

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Work	+61(2)5555-1794

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Reinitialize PDV

PDV

First	Gender	EmpID	Type	Phone
		.		

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Reads both observations into the PDV.

PDV

First	Gender	EmpID	Type	Phone
Kylie	F	121151	Home	+61(2)5555-1849

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
   merge empsau phones;
   by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Type	Phone
Kylie	F	121151	Home	+61(2)5555-1849

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Do the **EmpID** values match?

Yes

PDV

First	Gender	EmpID	Type	Phone
Kylie	F	121151	Home	+61(2)5555-1849

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Did **EmpID** change?

Yes

PDV

First	Gender	EmpID	Type	Phone
Kylie	F	121151	Home	+61(2)5555-1849

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

Reinitialize PDV

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

PDV

First	Gender	EmpID	Type	Phone
		.		

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Reads both observations into the PDV.

PDV

First	Gender	EmpID	Type	Phone
Birin	M	121152	Work	+61(2)5555-1850

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
   merge empsau phones;
   by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Type	Phone
Birin	M	121152	Work	+61(2)5555-1850

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

EOF

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Did **EmpID** change?

No

PDV

First	Gender	EmpID	Type	Phone
Birin	M	121152	Work	+61(2)5555-1850

...

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

EOF

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Reads the matching observation into the PDV.

PDV

First	Gender	EmpID	Type	Phone
Birin	M	121152	Home	+61(2)5555-1665

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

EOF

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Type	Phone
Birin	M	121152	Home	+61(2)5555-1665

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

EOF

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Did **EmpID** change?

No

PDV

First	Gender	EmpID	Type	Phone
Birin	M	121152	Home	+61(2)5555-1665

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

EOF

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

Reads the matching observation into the PDV.

PDV

First	Gender	EmpID	Type	Phone
Birin	M	121152	Cell	+61(2)5555-1666

Execution

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

EOF

```
data empphones;
   merge empsau phones;
   by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Type	Phone
Birin	M	121152	Cell	+61(2)5555-1666

...

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

EOF

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

EOF

```
data empphones;
    merge empsau phones;
    by EmpID;
run;
```

PDV

First	Gender	EmpID	Type	Phone
Birin	M	121152	Cell	+61(2)5555-1666

Final Results

empphones

First	Gender	EmpID	Type	Phone
Togar	M	121150	Home	+61(2)5555-1793
Togar	M	121150	Work	+61(2)5555-1794
Kylie	F	121151	Home	+61(2)5555-1849
Birin	M	121152	Work	+61(2)5555-1850
Birin	M	121152	Home	+61(2)5555-1665
Birin	M	121152	Cell	+61(2)5555-1666

Many-to-One Merge

One or more rows in one data set match exactly one row in the other data set.

phones

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121152	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Cell	+61(2)5555-1666

empsau

EmpID	First	Gender
121150	Togar	M
121151	Kylie	F
121512	Birin	M

```
data phones;
    merge phones empsau;
    by EmpID;
run;
```

L7_D14.sas

Viewing the Output

PROC PRINT Output

Obs	EmpID	Type	Phone	First	Gender
1	121150	Home	+61(2)5555-1793	Togar	M
2	121150	Work	+61(2)5555-1794	Togar	M
3	121151	Home	+61(2)5555-1849	Kylie	F
4	121152	Work	+61(2)5555-1850	Birin	M
5	121152	Home	+61(2)5555-1665	Birin	M
6	121152	Cell	+61(2)5555-1666	Birin	M

– The results are the same as the one-to-many merge.

– The order of variables is different.

UCD School of Mathematics and Statistics

www.ucd.ie/mathstat

Lecture 7 part 2: Combining Data Sets

7.3 Concatenating Data Sets

7.4 Merging Data Sets One-to-One

7.5 Merging Data Sets One-to-Many

7.6 Merging Data Sets with Nonmatches

UCD School of Mathematics and Statistics

www.ucd.ie/mathstat

Objectives

– Control the observations in the output data set by using the IN= data set option.

Scenario

An Orion Star manager in Australia requested an inventory of company phone numbers.
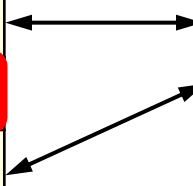
Merge with Nonmatches

There are observations in **empsau** that do not have a match in **phonec**, and some in **phonec** that do not match any observations in **empsau**.

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

The data sets are sorted by **EmpID**.

Match-Merging

Merge **empsau** and **phonec** by **EmpID** to create a new data set named **empsauc**.

```
data empsauc;
   merge empsau phonec;
   by EmpID;
run;
```

L7_D15.sas

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Initialize PDV

PDV

First	Gender	EmpID	Phone
		.	

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Do the **EmpID** values match?

Yes

PDV

First	Gender	EmpID	Phone
		.	

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Reads both observations into the PDV.

PDV

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
   merge empsau phonec;
   by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Do the **EmpID** values match?

No

PDV

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Did **EmpID** change?

Yes

PDV

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

> Reinitialize PDV

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

PDV

First	Gender	EmpID	Phone
		.	

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Which **EmpID** value sequentially comes first?

121151

PDV

First	Gender	EmpID	Phone
		.	

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Reads that observation into the PDV.

PDV

First	Gender	EmpID	Phone
Kylie	F	121151	

UCD School of Mathematics and Statistics

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Phone
Kylie	F	121151	

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Do the **EmpID** values match?

Yes

PDV

First	Gender	EmpID	Phone
Kylie	F	121151	

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Did **EmpID** change?

Yes

PDV

First	Gender	EmpID	Phone
Kylie	F	121151	

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

Reinitialize PDV

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

PDV

First	Gender	EmpID	Phone
		.	

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Reads both observations into the PDV.

PDV

First	Gender	EmpID	Phone
Birin	M	121152	+61(2)5555-1667

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Phone
Birin	M	121152	+61(2)5555-1667

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

EOF

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Did **EmpID** change?

Yes

PDV

First	Gender	EmpID	Phone
Birin	M	121152	+61(2)5555-1667

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

EOF

Reinitialize PDV

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

PDV

First	Gender	EmpID	Phone
		.	

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

EOF

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Reads the observation into the PDV.

PDV

First	Gender	EmpID	Phone
		121153	+61(2)5555-1348

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

EOF

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Phone
		121153	+61(2)5555-1348

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

EOF

EOF

```
data empsauc;
    merge empsau phonec;
    by EmpID;
run;
```

PDV

First	Gender	EmpID	Phone
		121153	+61(2)5555-1348

UCD School of Mathematics and Statistics

www.ucd.ie/mathstat

Final Results

empsauc

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795
Kylie	F	121151	
Birin	M	121152	+61(2)5555-1667
		121153	+61(2)5555-1348

The final results include both matches and nonmatches.

Exercise 10

Consider the data set **empsauc** created by the program in the previous example. Which input data sets contributed information to the last observation?

a. empsau

b. phonec

c. both empsau and phonec

d. There is insufficient information.

empsauc

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795
Kylie	F	121151	
Birin	M	121152	+61(2)5555-1667
		121153	+61(2)5555-1348

Scenario

An Orion Star manager requested three phone inventory reports.
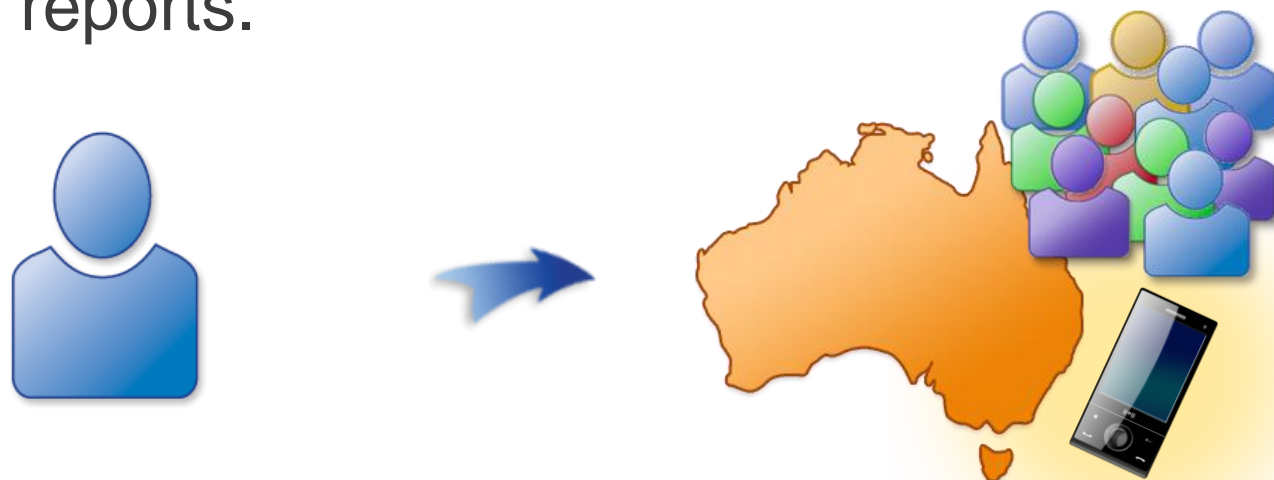
- – employees with company phones
- – employees without company phones
- – phones with an invalid employee ID

IN= Data Set Option

The *IN= data set option* creates a variable that indicates whether the data set contributed to building the current observation.

> **MERGE** *SAS-data-set* (**IN=***variable*) …

variable is a temporary numeric variable that has two possible values:

0	Indicates that the data set did **not** contribute to the current observation.
1	Indicates that the data set **did** contribute to the current observation.

IN= Data Set Option

MERGE statement examples:

```
merge empsau(in=Emps)
      phonec(in=Cell);
```

```
merge empsau(in=E)
      phonec(in=P);
```

```
merge empsau(in=AU)
      phonec;
```

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau(in=Emps)
          phonec(in=Cell);
    by EmpID;
run;
```

match

L7_D16.sas

PDV

First	Gender	EmpID	D Emps	Phone	D Cell
Togar	M	121150	1	+61(2)5555-1795	1

UCD School of Mathematics and Statistics

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau(in=Emps)
          phonec(in=Cell);
    by EmpID;
run;
```

nonmatch

PDV

First	Gender	EmpID	D⯈Emps	Phone	D⯈ Cell
Kylie	F	121151	1		0

Execution

empsau

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

phonec

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data empsauc;
    merge empsau(in=Emps)
          phonec(in=Cell);
    by EmpID;
run;
```

match

PDV

First	Gender	EmpID	D Emps	Phone	D Cell
Birin	M	121152	1	+61(2)5555-1667	1

PDV Results

PDV

First	Gender	EmpID	D Emps	Phone	D Cell
Togar	M	121150	1	+61(2)5555-1795	1
Kylie	F	121151	1		0
Birin	M	121152	1	+61(2)5555-1667	1
		121153	0	+61(2)5555-1348	1

The variables created with the IN= data set option are available only during DATA step execution.

– They are not written to the SAS data set.

– Their value can be tested using conditional logic.

Matches Only

Add a subsetting IF statement to select the employees that have company phones.

```
data empsauc;
    merge empsau(in=Emps)
          phonec(in=Cell);
    by EmpID;
    if Emps=1 and Cell=1;
run;
```

empsauc

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795
Birin	M	121152	+61(2)5555-1667

L7_D17.sas

Nonmatches from empsau

Select the employees that do not have company phones.

```sas
data empsauc;
    merge empsau(in=Emps)
          phonec(in=Cell);
    by EmpID;
    if Emps=1 and Cell=0;
run;
```

empsauc

First	Gender	EmpID	Phone
Kylie	F	121151	

L7_D17.sas

Nonmatches from phonec

Select the phones associated with an invalid employee ID.

```
data empsauc;
    merge empsau(in=Emps)
          phonec(in=Cell);
    by EmpID;
    if Emps=0 and Cell=1;
run;
```

empsauc

First	Gender	EmpID	Phone
		121153	+61(2)5555-1348

All Nonmatches

```
data empsauc;
    merge empsau(in=Emps)
          phonec(in=Cell);
    by EmpID;
    if Emps=0 or Cell=0;
run;
```

empsauc

First	Gender	EmpID	Phone
Kylie	F	121151	
		121153	+61(2)5555-1348

⚠️ Use the OR operator, not the AND operator.

L7_D17.sas

Alternate Syntax

When checking a variable for a value of *1* or *0* as in the previous scenario, you can use the following syntax:

Instead of	You can use
if Emps=1 and Cell=1;	if Emps and Cell;
if Emps=1 and Cell=0;	if Emps and not Cell;
if Emps=0 and Cell=1;	if not Emps and Cell;
if Emps=0 or Cell=0;	If not Emps or not Cell;