BC Cancer Agency
CARE + RESEARCH
An agency of the Provincial Health Services Authority

Cancer Surveillance & Outcomes

A Beginner's Experience Using SAS

Kimberly Burrus
Cancer Surveillance & Outcomes
Population Oncology
BC Cancer Agency

Overview

- Thoughts as a new SAS user
- Tips or lessons learned
- My favourite SAS tool – merging datasets
  - DATA Step
  - PROC SQL

Thoughts as a New SAS User

- New to SAS (~3 mo)
- User-friendly
- Like most new tasks, challenging
- But, you’re not alone!

Happy to report…

4 Things a Beginner Should Know

1. Document what you're doing (and why)
   - Add comment block to SAS program
     - Using `/* */` or `*;` to denote
     - Example:

```
**Create new dataset for breast cancer cases,
   stage I or II, diagnosis year 2010, 2011.
   Exclude lymphoma cases**;
```

   - Use a standard header to begin each file

Example Header Template

```
/*************************************************************************
Program Name:
Program Objective:

Project Name:
Project Objective:

Main Directory:
Other Directories:

Author:
SAS Version:
Program Created Date:

Datasets used:
Datasets created:

Formats used:
Stored Macros used:

Other Inputs:
Other Outputs:

Notes:

Date Modified         By         Reason

*************************************************************************/
```

4 Things a Beginner Should Know

2. Think before you type
   - Start with a plan
   - Understand your data
   - Understand how SAS 'thinks'

3. Use the SAS tools available to you
   - Specify variables to DROP, KEEP, or RENAME
   - Apply SAS formats

4 Things a Beginner Should Know

4. There isn't just one right way to do things
   - Focus on efficiency
   - Streamline code

Merging Data

DATA Step

- Can easily merge datasets
- Requires data to be in sorted order first
- Requires variables you're merging by to have the same name

PROC SQL

- Can also easily join (merge) datasets
- Can sort its results in the same step
- Can match on variables that are not exactly the same

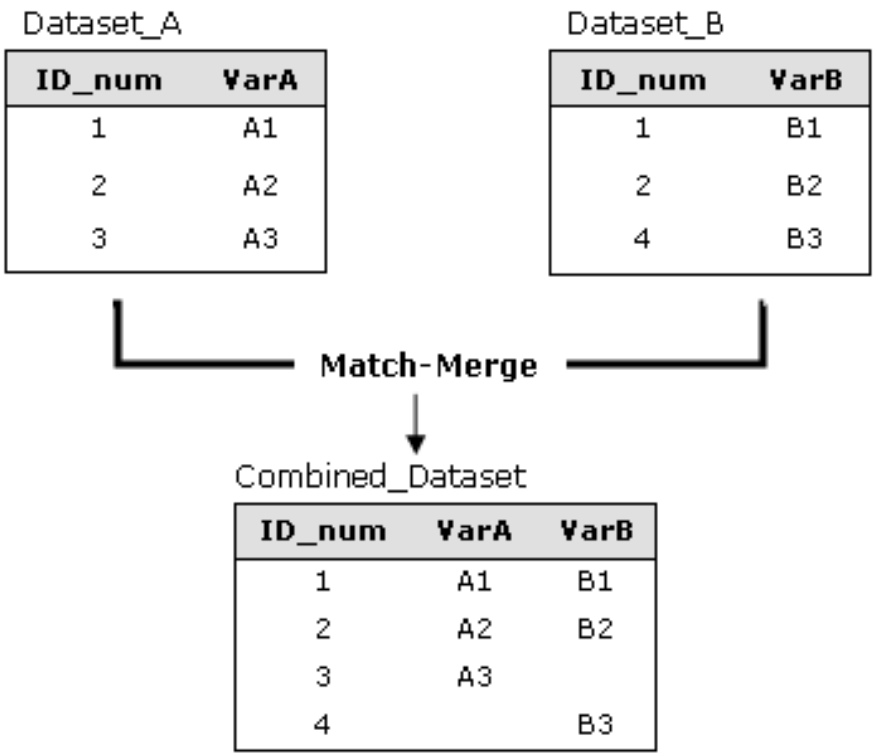
Data Step

- Methods of merging using the DATA Step:
  - One-to-one merging
  - Concatenating
  - Appending
  - Interleaving
  - Match-merging

Data Step: Match-merge

- Matches observations from two or more data sets into a single observation in a new data set according to the values of a common variable.
- Statements:

  MERGE, BY



Example SAS Code

```
proc sort data=dataset_A;
   by ID_num;
run;

proc sort data=dataset_B;
   by ID_num;
run;


data combined_dataset;
   MERGE  dataset_A
                  dataset_B;
   by ID_num;
run;
```

Proc SQL

- Methods of merging using the PROC SQL:
  - Inner Join (default SQL join)
  - Left Join
  - Right Join
  - Full Join

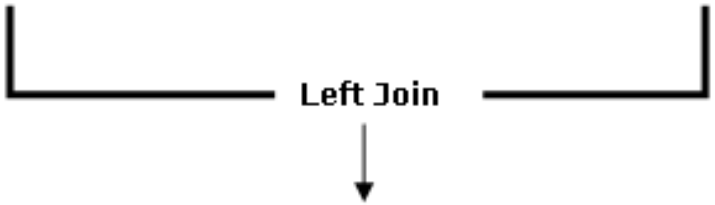
Proc SQL: Left Join

Dataset_A

ID_num	Name
1	Jones, Sam
3	Choi, Sora
6	Smith, Anne

Dataset_B

ID	Surgery_code	Surgery_date
1	1YM87	01Jan2010
1	1YM89	05May2010
3	1YM87	16Feb2010
7	1YM92	20Jun2010



Combined_Dataset

ID_num	Name	Surgery_code	Surgery_date
1	Jones, Sam	1YM87	01Jan2010
1	Jones, Sam	1YM89	05May2010
3	Choi, Sora	1YM87	16Feb2010
6	Smith, Anne		

Example SAS Code

```
PROC SQL;
   create table Combined_dataset as
   select a.*, b.*
   from dataset_A   a
                left join
          dataset_B   b
   on a.ID_num = b.ID
   ;

quit;
```

Thank you!