



Five SAS® questions we are afraid to ask



Adrien VALLEE
SAS consultant & trainer
adrien.vallee@keyrus.com

PhUSE SDE Brussels
02 April 2009

www.keyrusbiopharma.com

OVERVIEW

✓ How does a DATA step work?

✓ What does this NOTE mean?

MERGE statement has more than one data set with repeats of BY values.

✓ What is an ARRAY? How is it used?

✓ DSD, MISSOVER, TRUNCOVER ...

How should we use them?

✓ How to debug a DATA step?

✓ DATA step

✓ MERGE statement

✓ ARRAY statement

✓ Import file's options

✓ DEBUG option

How does a DATA step work?

The DATA step process:

✓ COMPILATION

➤ PDV: Program Data Vector

- In-memory array
- Structure of the output dataset

➤ Creation of the PDV

- Get variables from the input dataset
- Detect and add new calculated variables

✓ EXECUTION

➤ Fill the PDV with the current observation values

➤ Write in the output dataset

How does a DATA step work?

COMPILATION

```
DATA work.bmi;  
  SET work.myClass;  
  Bmi=round(weight/height**2,.1);  
  IF sex='M' THEN Gender='MALE';  
  ELSE IF sex='F' THEN Gender='FEMALE';  
RUN;
```

Input dataset: work.myClass

	Name	Sex	Age	Height	Weight
1	Mary	F	15	1.65	48
2	Philip	M	16	1.76	57

How does a DATA step work?

COMPILATION

Name of output dataset: work.bmi

```
DATA work.bmi;  
  SET work.myClass;  
  Bmi=round(weight/height**2,.1);  
  IF sex='M' THEN Gender='MALE';  
  ELSE IF sex='F' THEN Gender='FEMALE';  
RUN;
```

Input dataset: work.myClass

	Name	Sex	Age	Height	Weight
1	Mary	F	15	1.65	48
2	Philip	M	16	1.76	57

Output dataset: work.bmi

How does a DATA step work?

COMPILATION

Get variables from myClass
and add in the PDV

```
DATA work.bmi;  
  SET work.myClass;  
  Bmi=round(weight/height**2,.1);  
  IF sex='M' THEN Gender='MALE';  
  ELSE IF sex='F' THEN Gender='FEMALE';  
RUN;
```

Input dataset: work.myClass

	Name	Sex	Age	Height	Weight
1	Mary	F	15	1.65	48
2	Philip	M	16	1.76	57

PDV

Name \$ 8	Sex \$ 1	Age N 8	Height N 8	Weight N 8
		.	.	.

Output dataset: work.bmi

How does a DATA step work?

COMPILATION

```
DATA work.bmi;  
  SET work.myClass;  
  Bmi=round(weight/height**2,.1);  
  IF sex='M' THEN Gender='MALE';  
  ELSE IF sex='F' THEN Gender='FEMALE';  
RUN;
```

Bmi: New calculated variable

Add in the PDV

Weight and Height already exists in the PDV

Input dataset: work.myClass

	Name	Sex	Age	Height	Weight
1	Mary	F	15	1.65	48
2	Philip	M	16	1.76	57

PDV

Name \$ 8	Sex \$ 1	Age N 8	Height N 8	Weight N 8	Bmi N 8
		.	.	.	.

Output dataset: work.bmi