



Proc Compare

GHSUG- Proc du Jour
October 2, 2009

Lesley Harschnitz
lesley.harschnitz@arcelormittal.com



Proc Compare

- Compares:

- whether matching variables have different values
- whether one data set has more observations than the other
- what variables the two data sets have in common
- how many variables are in one data set but not in the other
- whether matching variables have different formats, labels, or types.
- a comparison of the values of matching observations.

- Used for

- Data Migration
- Code change validation
- Report validation
- Data quality
- Data editing validation



Proc Compare: Syntax

PROC COMPARE

base= dsname compare= dsname;

BY group varname; *(grouped analysis)*

ID identifier varname; *(matches observations)*

VAR varlist; *(selected variables)*

WITH varlist; *(matching variables- names differ)*

RUN;



Example

- From SAS provided examples:
 - First data set:
 - Student #, Year of Birth, Home State, Grade1, Grade2, Major
 - Second data set:
 - Student #, Year of Birth, Home State, Grade1, Grade2, Major
 - Proc Compare Code:
proc compare base=one compare=two;
id student;
run;



Output Samples-1

Data Set Summary

Dataset	Created	Modified	NVar	NObs	Label
WORK.ONE	01OCT09:12:29:52	01OCT09:12:29:52	5	4	First Data Set
WORK.TWO	01OCT09:12:29:52	01OCT09:12:29:52	6	5	Second Data Set

Variables Summary

Number of Variables in Common: 5.

Number of Variables in WORK.TWO but not in WORK.ONE: 1.

Number of Variables with Differing Attributes: 3.

Number of ID Variables: 1.

Listing of Common Variables with Differing Attributes

Variable	Dataset	Type	Length	Format	Label
year	WORK.ONE	Char	8		Year of Birth
	WORK.TWO	Char	8		
state	WORK.ONE	Char	8		Home State
	WORK.TWO	Char	8		
gr1	WORK.ONE	Num	8	4.1	
	WORK.TWO	Num	8	5.2	



Output Samples-2

Value Comparison Results for Variables

student	Base gr1	Compare gr1	Diff.	% Diff
1000	85.0	84.00	-1.0000	-1.1765
1095	78.0	79.00	1.0000	1.2821

Value Comparison Results for Variables

student	Base gr2	Compare gr2	Diff.	% Diff
1095	72.0000	73.0000	1.0000	1.3889
1187	94.0000	74.0000	-20.0000	-21.2766