Love of Errors. When Speed Matters

Adam Amborski

2018-10-12

Agenda

- Why love errors?
- How to show your errors to the world?
- How to get even more errors?

Why love errors?

```
proc compare base=base compare=comp1;
run;
```

```
                        The COMPARE Procedure
                    Comparison of WORK.BASE with WORK.COMP
                                (Method=EXACT)

                              Data Set Summary

 Dataset             Created          Modified  NVar    NObs

 WORK.BASE  18SEP18:13:45:21  18SEP18:13:45:21    16  235478
 WORK.COMP  18SEP18:13:45:21  18SEP18:13:45:21    16  235477


                             Variables Summary

                    Number of Variables in Common: 16.


                            Observation Summary

                        Observation      Base  Compare

                        First Obs           1        1
                        Last  Match    235477   235477
                        Last  Obs      235478        .

Number of Observations in Common: 235477.
Number of Observations in WORK.BASE but not in WORK.COMP: 1.
Total Number of Observations Read from WORK.BASE: 235478.
Total Number of Observations Read from WORK.COMP: 235477.

Number of Observations with Some Compared Variables Unequal: 0.
Number of Observations with All Compared Variables Equal: 235477.

NOTE: No unequal values were found. All values compared are exactly equal.
```

Why love errors?

```
proc compare base=base compare=comp2;
run;
```

```
                        The COMPARE Procedure
                Comparison of WORK.BASE with WORK.COMP2
                             (Method=EXACT)

                          Data Set Summary

Dataset              Created          Modified  NVar    NObs

WORK.BASE   17SEP18:16:22:49  17SEP18:16:22:49     1 2354578
WORK.COMP2  17SEP18:16:25:09  17SEP18:16:25:09     2 2354578


                          Variables Summary

            Number of Variables in Common: 1.
            Number of Variables in WORK.COMP2 but not in WORK.BASE: 1.


                         Observation Summary

                Observation      Base  Compare

                First Obs           1        1
                Last  Obs     2354578  2354578

Number of Observations in Common: 2354578.
Total Number of Observations Read from WORK.BASE: 2354578.
Total Number of Observations Read from WORK.COMP2: 2354578.

Number of Observations with Some Compared Variables Unequal: 0.
Number of Observations with All Compared Variables Equal: 2354578.

NOTE: No unequal values were found. All values compared are exactly equal.
```

Why love errors?

```
proc compare base=base compare=comp3;
  var a;
run;
```

```
                        The COMPARE Procedure
                Comparison of WORK.BASE with WORK.COMP3
                             (Method=EXACT)

                           Data Set Summary

Dataset              Created          Modified  NVar    NObs

WORK.BASE    17SEP18:18:26:19  17SEP18:18:26:19     2 2354578
WORK.COMP3   17SEP18:18:34:54  17SEP18:18:34:54     2 2354578


                          Variables Summary

                Number of Variables in Common: 2.
                Number of VAR Statement Variables: 1.


                        Observation Summary

                    Observation      Base  Compare

                    First Obs           1        1
                    Last  Obs     2354578  2354578

Number of Observations in Common: 2354578.
Total Number of Observations Read from WORK.BASE: 2354578.
Total Number of Observations Read from WORK.COMP3: 2354578.

Number of Observations with Some Compared Variables Unequal: 0.
Number of Observations with All Compared Variables Equal: 2354578.

NOTE: No unequal values were found. All values compared are exactly equal.
```

Why love errors?

```
proc freq data=adsl;
  table hppfl;
run;
```

```
NOTE: There were 11 observations read
      from the data set WORK.ADSL.
```

hppfl	Frequency
N	1
Y	10

```
proc sort data=adsl
          out=happy00 nodupkey;
  by usubjid;
run;
```

```
NOTE: There were 11 observations read
      from the data set WORK.ADSL.
NOTE: 1 observations with duplicate key
      values were deleted.
NOTE: The data set WORK.HAPPY00 has 10
      observations and 2 variables.
```

```
proc freq data=happy00;
  table hppfl;
run;
```

```
NOTE: There were 10 observations read
      from the data set WORK.HAPPY00.
```

hppfl	Frequency
N	1
Y	9

```
proc sql;
  select count(distinct usubjid)
         label="Happy Subjects"
      from adsl where hppfl="Y";
quit;
```

```
NOTE: PROCEDURE SQL used (Total process
      time):
```

Happy Subjects
10

Why love errors?

```
data kot;
  format b c 1.;
  a="GRADE 1";
  b=a;
  c=input(scan(a,2),best.);
  put a= b= c=;
run;
```

```
NOTE: Character values have been converted to numeric values at
      the places given by: (Line):(Column).
      611:5
NOTE: Invalid numeric data, a='GRADE 1' , at line 611 column 5.
a=GRADE 1 b=. c=1
b=. c=1 a=GRADE 1 _ERROR_=1 _N_=1
NOTE: The data set WORK.KOT has 1 observations and 3 variables.
```

Why love errors?

```
* Converting to C66731 Sex           *;
* CDISC SDTM Controlled Terminology *;
* 2018-06-29                        *;
data dm_sex_a;
  length sex $1;
  set sexes;
  sex=sex_orig;
  put sex= sex_orig= ;
run;
```

Sex (SEX)	text Extensible: No	C66731	Sex
F		C16576	Female
M		C20197	Male
U		C17998	U;UNK;Unknown

```
sex=F sex_orig=Female
sex=M sex_orig=Male
sex=U sex_orig=Unknown
sex=U sex_orig=Undifferentiated
NOTE: There were 4 observations read from the data set WORK.SEXES.
NOTE: The data set WORK.DM_SEX_A has 4 observations and 3 variables.
```

How to show your errors to the world?

```
proc compare base=A compare=B error;
run;
```

```
ERROR: 1 variables are numeric in one data set but character in the other.
ERROR: 1 variables have conflicting attributes in the two data sets.
ERROR: Data set WORK.B contains 5 observations not in WORK.A.
ERROR: Values of the following 1 variables compare unequal: a
ERROR: The data sets WORK.A and WORK.B contain unequal values.
```

…this will not protect you from the var statement issue!

Proc COMPARE gives you &SYSINFO automatic result code, with a 0 for the perfect match. This can be used easily in automation!

How to show your errors to the world?

- DSNFERR – make sure lack of dataset in set statement causes error.
- MERGENOBY= ERROR – making sure data records are not randomly matched if you forget a *by* statement.
- SOURCE – this one shows the source code. It does make log longer, but makes it easier to debug, e.g. for numeric/character conversions

How to get even more errors?

```
proc sort data=adsl out=happy00;
  by usubjid hppfl;
run;

data happy01 adsl_dups;
  set happy00;
  by usubjid;
  if first.usubjid + last.usubjid ne 2 then
output adsl_dups;
  if last.usubjid then output happy01;
run;

data _NULL_;
  set adsl_dups end=_last;
  if _last then
    put "ER" "ROR: Unexp" "ected duplicates
( " _N_ ") in ADSL. Check adsl_dups.";
run;
```

```
NOTE: There were 11 observations read from the data
      set WORK.ADSL.
NOTE: The data set WORK.HAPPY00 has 11 observations
      and 2 variables.

NOTE: There were 11 observations read from the data
      set WORK.HAPPY00.
NOTE: The data set WORK.HAPPY01 has 10 observations
      and 2 variables.
NOTE: The data set WORK.ADSL_DUPS has 2 observations
      and 2 variables.

ERROR: Unexpected duplicates ( 2 ) in ADSL. Check
       adsl_dups.
NOTE: There were 2 observations read from the data
      set WORK.ADSL_DUPS.
```

How to get even more errors?

```
data _null_;
  if nobs = 0 then
    put "ER" "ROR: Empty happy01 dataset!";
  set happy01 nobs=nobs;
run;
```

Source:
https://communities.sas.com/t5/SAS-Programming/How-to-test-for-zero-observations-and-still-be-in-control/td-p/80447

How to get even more errors?

```
data dm_sex_b;
  length sex $1;
  set sexes;
  select(sex_orig);
    when("Female")   sex="F";
    when("Male")     sex="M";
    when("Unknown")  sex="U";
  end;
run;
```

```
ERROR: Unsatisfied WHEN clause and no OTHERWISE clause at line 1751
column 3.
sex=  usubjid=4 sex_orig=Undifferentiated _ERROR_=1 _N_=4
NOTE: The SAS System stopped processing this step because of errors.
NOTE: There were 4 observations read from the data set WORK.SEXES.
WARNING: The data set WORK.DM_SEX_B may be incomplete.  When this step
         was stopped there were 3 observations and 3 variables.
```

How to get even more errors?

```
data dm_sex_b;
  length sex $1;
  set sexes;
  select(sex_orig);
    when("Female")  sex="F";
    when("Male")    sex="M";
    when("Unknown") sex="U";
    otherwise put "ER" "ROR: Unexpected " sex_orig=;
  end;
run;
```

```
ERROR: Unexpected sex_orig=Undifferentiated
NOTE: There were 4 observations read from the data set WORK.SEXES.
NOTE: The data set WORK.DM_SEX_B has 4 observations and 3 variables.
```

Is it too much to ask?

Thank you!

Adam.Amborski@Statfinn.com