

Illustrating Concatenation Operations with PROC APPEND

Kirk Paul Lafler, Software Intelligence Corporation

Abstract

The APPEND procedure provides SAS users with an efficient method for concatenating observations from a smaller data set to a larger data set. This paper explores how PROC APPEND is used, how it works, what happens when variables in the smaller data set are not present in the larger data set, and its advantages over other concatenation methods.

The Essence of PROC APPEND

The APPEND procedure (and APPEND statement of the DATASETS procedure) is an efficient method for concatenating observations from a smaller data set to a larger data set. The BASE= data set option is reserved for the larger of the two data sets with the DATA= option for the smaller data set. Essentially, the APPEND procedure avoids reading any observations in the BASE= data set by positioning the record pointer at the end of the BASE= data set. Each observation from the smaller data set is then applied one at a time to the end of the BASE= data set. In the first example, the BASE= data set identifies a larger data set called MASTER and the DATA= data set identifies the smaller TXN data set.

Code:

```
PROC APPEND
    BASE=master
    DATA=txn;
RUN;
```

When two or more data sets need to be concatenated, multiple APPEND procedures are issued. In the next example, two separate PROC APPEND steps are specified to concatenate the two smaller data sets (TXN1 and TXN2) at the end of the larger BASE= (MASTER) data set.

Code:

```
PROC APPEND
    BASE=master
    DATA=txn1;
RUN;
PROC APPEND
    BASE=master
    DATA=txn2;
RUN;
```

Forcing the Concatenation of Data Sets

When one or more variables in the input data set, represented by the DATA= option, are not present in the BASE= data set, or the variables in the input data set are not of the same type as the variables in the BASE= data set, an optional FORCE option can be specified on the PROC APPEND statement to force the concatenation of the data sets. Using the FORCE option also prevents an error from being generated in these situations. The next example illustrates the FORCE option being specified in the concatenation of the TXN data set to the MASTER data set.

Code:

```
PROC APPEND
    BASE=master
    DATA=txn
    FORCE;
RUN;
```

Selecting Observations for Concatenation

By default, all the rows of an input data set are typically concatenated to the end of the base data set. But when only selected rows of an input data set are desired, an optional WHERE clause can be specified for the associated input data set as a data set option. This is accomplished by specifying an optional WHERE clause. The WHERE clause defines the logical conditions that control which observations will be selected from an input data set, and ultimately concatenated to the end of the base data set. In the next example, a WHERE clause is specified to select only 'PG' and 'PG13' movies from the input data set for concatenation to the end of the base data set.

Code:

```
PROC APPEND  
  
    BASE=master  
  
    DATA=txn (WHERE=(rating in ('PG', 'PG-13')));  
  
RUN;
```

Conclusion

The APPEND procedure is a powerful tool for SAS users to use for data set concatenation purposes. If used, the smaller data set should be specified as the input data set (DATA= option) and the larger data set as the base data set (BASE= option). Users can also concatenate unlike data sets using the FORCE option on the PROC APPEND statement as well as select which observations are captured from the input data set using a WHERE clause and concatenated to the end of the base data set. Users are encouraged to explore the many capabilities the APPEND procedure (or the APPEND statement of the DATASETS procedure) has to offer.

References

Lafler, Kirk Paul (2003), "Concatenating SAS Data Sets with the APPEND Procedure," Kirk's Korner of SAS Tips, Software Intelligence Corporation, Spring Valley, CA, USA.

Acknowledgments

I would like to thank Beth Newsom, WUSS Coders' Corner Section Chair, for accepting my abstract and paper, as well as the WUSS Leadership for their support of a great Conference.

Contact Information

Kirk Paul Lafler, a SAS Certified Professional® and SAS Alliance Partner® (1996 - 2002) with more than 27 years of SAS software experience, provides consulting services and hands-on SAS training around the world. Kirk has written four books including PROC SQL: Beyond the Basics Using SAS by SAS Institute, Power SAS and Power AOL by Apress, and more than one hundred articles in professional journals and SAS User Group proceedings. His popular SAS Tips column, Kirk's Korner, appears regularly in the BASAS, HASUG, SANDS, SAS, SESUG, and WUSS Newsletters and websites. Kirk can be reached at:

Kirk Paul Lafler
Software Intelligence Corporation
P.O. Box 1390
Spring Valley, California 91979-1390
E-mail: KirkLafler@cs.com
Voice: 619-277-7350



SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.