Paper PH09-2012

Reproducible Research Two Ways: SASweave vs. StatRep

Shannon M. Morrison, M.S.
Matthew T. Karafa, PhD.
Quantitative Health Sciences – Cleveland Clinic Foundation, Cleveland, OH

Abstract

Reproducible research is becoming increasingly important in biostatistics. It is essential to keep documentation of code used for data management and analyses along a copy of the results for future verification of research. It can also be especially helpful when working as part of a team of researchers, where projects are sometimes completed through a group effort or are handed off to co-workers due to various reasons (going on leave, leaving for a new position, retirement, etc.).

In this paper, ease-of-use and results from SASweave and StatRep are compared using a publicly available SAS® dataset. An example program, which contains basic data manipulation and simple, widely used SAS® Procedures, is looked at using both packages. More challenging, in-house macros are also run using both tools, again to gauge ease-of-use and quality of results. Along the way, pros and cons of both tools are discussed in an attempt to find a better way to generate reproducible research.

Introduction

The Department of Quantitative Health Sciences (QHS) at Cleveland Clinic offers research support in the form of statistical analysis, data management services, and statistical programming. Department members often work on projects in groups of 2 or 3, update work that can be several years old, take over projects begun by someone else (people who have left the department years ago or those on temporary leave, for example) and work on regulated trials. Thus the concept of creating reproducible research is becoming more important than ever. In most cases documentation is seen as an asset simply because it is helpful. In others, it is necessary to have documentation readily available for regulatory compliance purposes. Oftentimes running a SAS® program in batch mode on a validated UNIX server creates sufficient documentation, generating a .log and .lst file from the SAS® program that can be stored along with the results typed up in what the team refers to as a "memo format."

Two methods seem to satisfy the desire for more attractive, easier-to-store and update documentation. Both require the user to change and run any SAS® code in one location in order to generate updated results: they automatically produce .log and .lst files, as well as a SAS® program file and final report that includes SAS® code and results in one document.

An article on SASweave, known as the SAS counterpart to the popular SWeave tool (Leisch 2002) that is widely used for R documentation, first appeared in the Journal of Statistical Software in June 2007. SASweave "is a collection of AWK and shell scripts" that the user must download (Lenth and Højsgaard 2007). This gives the user the ability to "embed SAS® code into a $\LaTeX$ document," producing code, output and other documentation stored in one location (Lenth and Højsgaard 2007). Users can produce a flat SAS® file (the SAS® program, i.e. the ".sas" file) using the SAStangle package, which is part of the SASweave download.

StatRep, a newer package supported by the SAS® institute, was presented at SAS® Global Forum in 2012. Like SASweave, StatRep uses the $\LaTeX$ typesetting system to generate results. A SAS® flat file of the code is automatically produced, along with the output, .log and .lst files, which are all stored in one central location.

Documentation similar to that produced by SWeave was requested for a QHS project in 2012. In an attempt to find the preferable method to produce the desired documentation, a small program was created and run in both SASweave and StatRep to compare and contrast both ease-of-use and satisfaction with the results produced by the two tools. In-house macros `%cattable` and `%conttable` (Karafa 2006), which are regularly used by the team, were also run in the packages in an attempt to gauge how well they would be handled by both SASweave and StatRep.

Set-up

The UNIX version of both tools and the program TeXworks were downloaded to create the files. Both packages were not immediately ready to use upon download: extra set-up in both files included in the tool downloads and in the UNIX environment itself were needed before the tools could be properly used. For this paper, SASweave and

StatRep were run in the UNIX environment; users may have different experiences using the packages in a different system set-up.

SASweave

The SASweave installation file is a .tgz file, which is a compressed file like a .zip or .7z file some users may be unfamiliar with. Once the folder was extracted on the UNIX server, some SASweave files needed to be moved for the tool to work properly. Many of the files included in the SASweave download needed to be opened and modified (mostly to update file locations). The Systems/IT team will be fairly helpful in achieving this install; it is not for the novice UNIX user. For example, figuring out which incarnation of AWK the system uses, and the correct command-line switches, was a tedious and lengthy process. Although a README file with basic instructions was included in the download, a more seasoned UNIX expert needed to be called in for help before the tool could be used. It was quickly concluded that SASweave is not a package that can be downloaded and used immediately, and that the code written for SASweave is not robust: the lack of an installer means the user must configure it to their own environment, at least in this example on the UNIX server.

As mentioned in the Introduction section, SASweave does not automatically create a flat SAS® file. If a user wishes to store a flat SAS® file (in addition to the $\LaTeX$ file that contains the code for SASweave), they must run SAStangle in the UNIX terminal window. SAStangle is another file included in the SASweave install that needs to be properly configured, similar to the process for SASweave described above. The resulting file can be run directly in SAS®; SASweave inserts comments identifying what number “code chunk” portions of the program are (“Code chunk 1,” “Code chunk 2,” etc.), which correspond to the different begin/end SAScode sections in the $\LaTeX$ code (see example below).

StatRep

StatRep set-up involved creation of subdirectories to load $\LaTeX$ packages. A README file containing instructions for the set-up of these subdirectories was included in the StatRep download and was fairly easy to follow. The biggest hurdle for properly setting up StatRep was obtaining a full $\LaTeX$ install on the UNIX server in order to use packages specified by the StatRep macros. Such a set-up will most likely require the help of a user’s Systems/IT team, as it did in QHS’ case since QHS users do not have full administrative rights on PCs. This is not a fault of the StatRep tool itself, but it was something that took time during set-up.

Set-up for both tools is non-trivial, especially for a user that does not full administrator rights on their system that is running SAS®. It is advised that users contact their system administrators well in advance of any deadlines to configure their packages accordingly. For the average SAS® user running the tools on a UNIX server, setting up both of the tools may be a challenge due to the need-to-know locations and configurations of $\LaTeX$ packages and the AWK program. Note that this is one example of challenges faced for users that were running the packages on UNIX servers…those running the tools using PC SAS® may have a completely different experience.

Basic SAS® Code Examples

Results from basic SAS® code were generated using both SASweave and StatRep. The code is shown below:

```
proc print data = sashelp.shoes (obs = 10) noobs;
   title 'Basic Print of a portion of the data';
run;

proc format;
   value group 0 = 'Men Casual or Dress'
               1 = 'Women Casual or Dress';

   value continent 1 = 'Africa'
                   2 = 'Asia'
                   3 = 'North America'
                   4 = 'Europe'
                   5 = 'South America';

run;


Data modify_shoes;
   set sashelp.shoes;
   Product_Group = .;
   Continent     = .;

        if Product in ("Men's Casual", "Men's Dress") then Product_Group = 0;
   else if Product in ("Women's Casual", "Women's Dress") then Product_Group = 1;

        if Region = 'Africa' then Continent = 1;
   else if Region in ('Asia', 'Middle East', 'Pacific') then Continent = 2;
   else if Region in ('Canada', 'Central America/Caribbean', 'United States') then
Continent = 3;
   else if Region in ('Eastern Europe', 'Western Europe') then Continent = 4;
   else if Region = 'South America' then Continent = 5;

   label
      Product_group = 'Grouping variable'
      Continent     = 'Continent'
   ;

   format Product_Group group. Continent continent.;
run;

%macro checkcat(newvar = , oldvar = );
proc freq data = modify_shoes;
   table &oldvar. * &newvar. / norow nocol nopercent missing;
   title 'Check variable creation';
run;
%mend;

%checkcat(newvar = Product_Group, oldvar = Product);
%checkcat(newvar = Continent, oldvar = Region);

proc means data = modify_shoes MEDIAN Q1 Q3 MAXDEC = 2;
   var Stores Sales Inventory Returns;
   title 'Median(Q1, Q3)';
run;

title;
```

To use the packages, a source file must be created in a text-based editor that can be passed through the LaTeX typesetting system. It is important to note that the LaTeX typesetting system is *case-sensitive,* and the commands used for both tools are also *white space sensitive*. Expected results from the above code are listed below:

```
Basic Print of a portion of the data                                       1
                                                              11:42 Tuesday, August 14, 2012

Region    Product           Subsidiary     Stores           Sales       Inventory         Returns

Africa    Boot              Addis Ababa      12           $29,761        $191,821             $769
Africa    Men's Casual      Addis Ababa       4           $67,242        $118,036           $2,284
Africa    Men's Dress       Addis Ababa       7           $76,793        $136,273           $2,433
Africa    Sandal            Addis Ababa      10           $62,819        $204,284           $1,861
Africa    Slipper           Addis Ababa      14           $68,641        $279,795           $1,771
Africa    Sport Shoe        Addis Ababa       4            $1,690         $16,634              $79
Africa    Women's Casual    Addis Ababa       2           $51,541         $98,641             $940
Africa    Women's Dress     Addis Ababa      12          $108,942        $311,017           $3,233
Africa    Boot              Algiers          21           $21,297         $73,737             $710
Africa    Men's Casual      Algiers           4           $63,206        $100,982           $2,221

                 Check variable creation          11:42 Tuesday, August 14, 2012   2

                The FREQ Procedure
          Table of Product by Product_Group

Product         Product_Group(Grouping variable)

Frequency      |       .|Men Casu|Women Ca|  Total
```

```
               |        |al or Dr|sual or |
               |        |ess     |Dress   |
---------------+--------+--------+--------+
Boot           |     52 |      0 |      0 |     52
---------------+--------+--------+--------+
Men's Casual   |      0 |     45 |      0 |     45
---------------+--------+--------+--------+
Men's Dress    |      0 |     50 |      0 |     50
---------------+--------+--------+--------+
Sandal         |     49 |      0 |      0 |     49
---------------+--------+--------+--------+
Slipper        |     52 |      0 |      0 |     52
---------------+--------+--------+--------+
Sport Shoe     |     51 |      0 |      0 |     51
---------------+--------+--------+--------+
Women's Casual |      0 |      0 |     45 |     45
---------------+--------+--------+--------+
Women's Dress  |      0 |      0 |     51 |     51
---------------+--------+--------+--------+
Total               204       95       96      395

               Check variable creation         11:42 Tuesday, August 14, 2012   3

                         The FREQ Procedure

                    Table of Region by Continent
Region            Continent(Continent)

Frequency        |Africa  |Asia    |North Am|Europe  |South Am|  Total
                 |        |        |erica   |        |erica   |
-----------------+--------+--------+--------+--------+--------+
Africa           |     56 |      0 |      0 |      0 |      0 |     56
-----------------+--------+--------+--------+--------+--------+
Asia             |      0 |     14 |      0 |      0 |      0 |     14
-----------------+--------+--------+--------+--------+--------+
Canada           |      0 |      0 |     37 |      0 |      0 |     37
-----------------+--------+--------+--------+--------+--------+
Central America/ |      0 |      0 |     32 |      0 |      0 |     32
Caribbean        |        |        |        |        |        |
-----------------+--------+--------+--------+--------+--------+
Eastern Europe   |      0 |      0 |      0 |     31 |      0 |     31
-----------------+--------+--------+--------+--------+--------+
Middle East      |      0 |     24 |      0 |      0 |      0 |     24
-----------------+--------+--------+--------+--------+--------+
Pacific          |      0 |     45 |      0 |      0 |      0 |     45
-----------------+--------+--------+--------+--------+--------+
South America    |      0 |      0 |      0 |      0 |     54 |     54
-----------------+--------+--------+--------+--------+--------+
United States    |      0 |      0 |     40 |      0 |      0 |     40
-----------------+--------+--------+--------+--------+--------+
Western Europe   |      0 |      0 |      0 |     62 |      0 |     62
-----------------+--------+--------+--------+--------+--------+
Total                  56       83      109       93       54      395

                    Median(Q1, Q3)            11:42 Tuesday, August 14, 2012   4

                               The MEANS Procedure
                                                          Lower            Upper
Variable     Label                      Median         Quartile         Quartile
--------------------------------------------------------------------------------
Stores       Number of Stores            10.00             3.00            18.00
Sales        Total Sales              38912.00         15312.00        108936.00
Inventory    Total Inventory         118849.00         43155.00        336513.00
Returns      Total Returns             1438.00           578.00          3698.00
-------------------------------------------------------------------------------
```

SASweave

If a user does not wish to use any of the options available in the SASweave package, generating the text file is fairly straight-forward. As with any LaTeX document, `\begin{document}` and `\end{document}` statements are included at the beginning and end of the text document. Other statements indicate that the document is an article, to use the LaTeX package `mathpazo` (which creates mathematical typesetting) and to print the name of the document itself and the author(s):

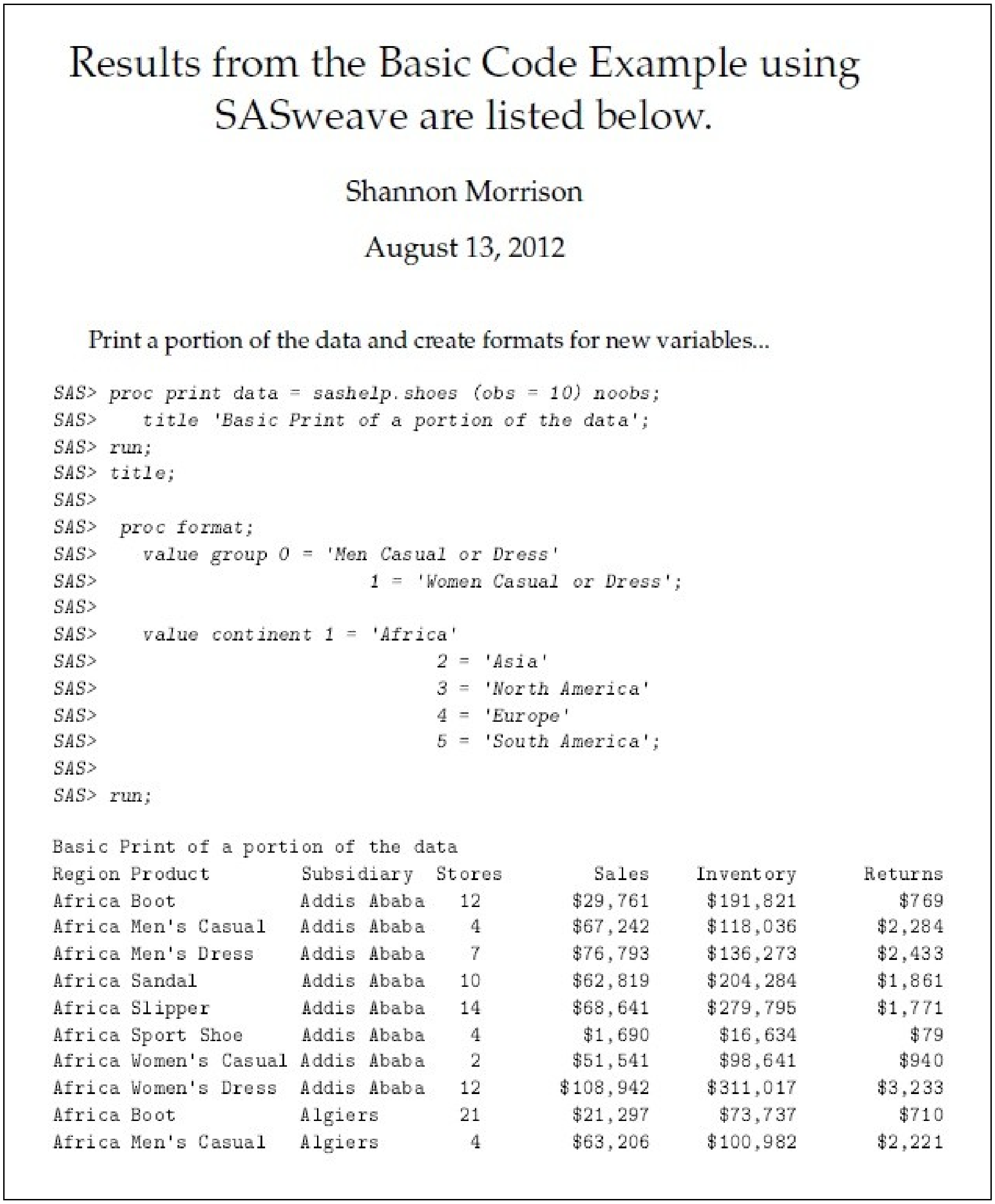
```
\documentclass{article}
\usepackage{mathpazo}

\title{Results from the Basic Code Example using SASweave are listed below.}
\author{Shannon Morrison}

\begin{document}
```

Per the SASweave documentation, SAS® “code chunks” (portions of SAS® code) must be surrounded by `\begin{SAScode}` and `\end{SAScode}` statements to be executed properly. Text statements, similar to comments used in SAS® programs, should not be surrounded by `\begin{SAScode}` and `\end{SAScode}` statements. Again, the code is case-sensitive: `\begin{SASCode}` or something similar will not work here. In addition, the SASweave statements are also sensitive to spacing: `\begin {SAScode}` will not work.

The first code chunk, which contains code for the PRINT and FORMAT procedures, is shown below:

```
\maketitle

Print a portion of the data and create formats for new variables...

\begin{SAScode}

proc print data = sashelp.shoes (obs = 10) noobs;
   title 'Basic Print of a portion of the data';
run;

title;

proc format;
   value group 0 = 'Men Casual or Dress'
               1 = 'Women Casual or Dress';

   value continent 1 = 'Africa'
                   2 = 'Asia'
                   3 = 'North America'
                   4 = 'Europe'
                   5 = 'South America';
run;

\end{SAScode}

…etc…
```

The output was generated using the `sasweave` command in the UNIX terminal window. Output generated by SASweave included a SASweave .tex file (which can be passed through pdfLaTeX to generate a .pdf file) and a .log file. A portion of the .pdf file generated by pdfLaTeX is shown below:



Results from the Basic Code Example using SASweave are listed below.

Shannon Morrison

August 13, 2012

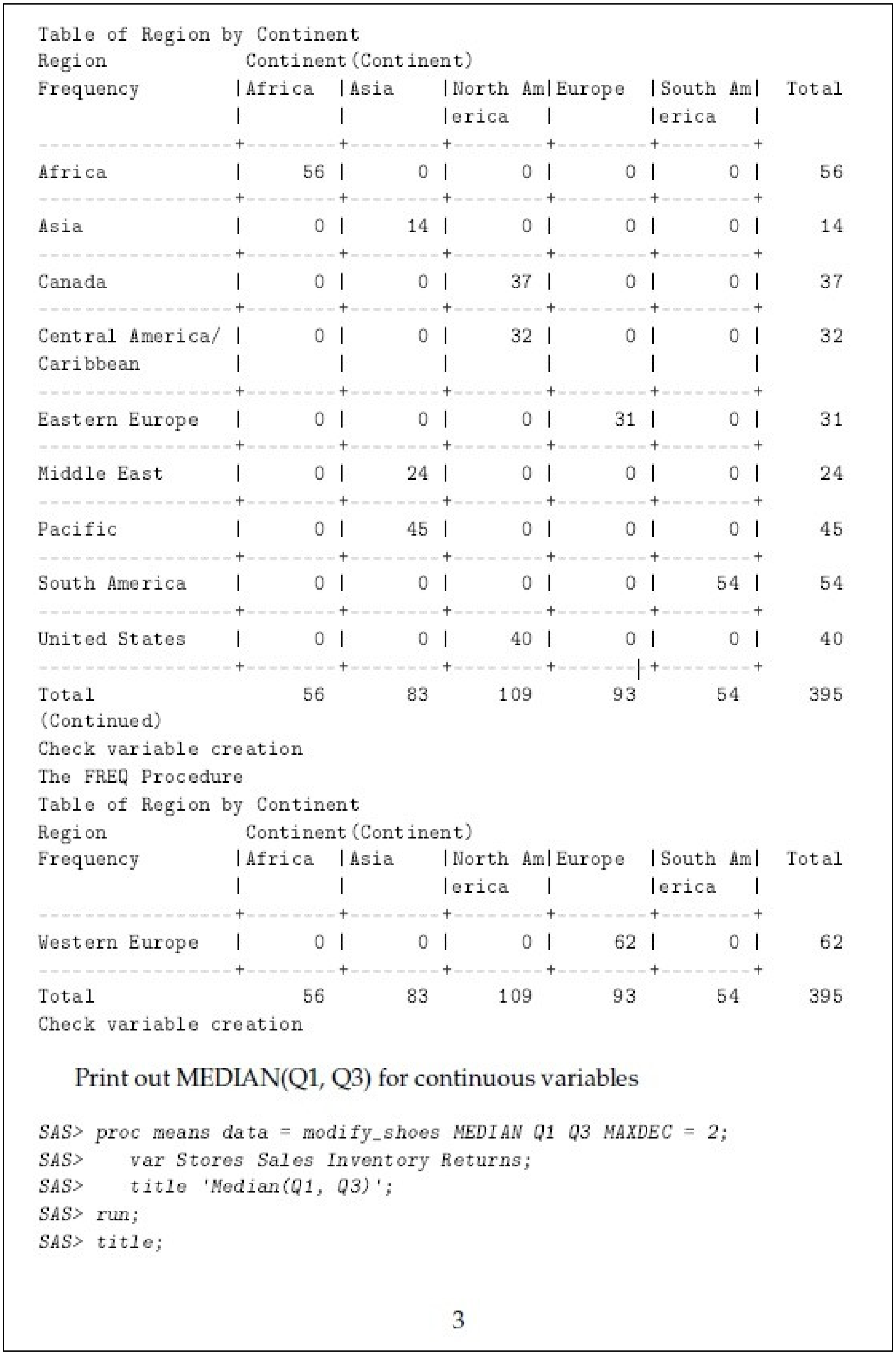
Print a portion of the data and create formats for new variables...

```
SAS> proc print data = sashelp.shoes (obs = 10) noobs;
SAS>    title 'Basic Print of a portion of the data';
SAS> run;
SAS> title;
SAS>
SAS>  proc format;
SAS>    value group 0 = 'Men Casual or Dress'
SAS>                     1 = 'Women Casual or Dress';
SAS>
SAS>    value continent 1 = 'Africa'
SAS>                         2 = 'Asia'
SAS>                         3 = 'North America'
SAS>                         4 = 'Europe'
SAS>                         5 = 'South America';
SAS>
SAS> run;
```

```
Basic Print of a portion of the data
Region Product        Subsidiary  Stores       Sales    Inventory      Returns
Africa Boot           Addis Ababa   12       $29,761     $191,821         $769
Africa Men's Casual   Addis Ababa    4       $67,242     $118,036       $2,284
Africa Men's Dress    Addis Ababa    7       $76,793     $136,273       $2,433
Africa Sandal         Addis Ababa   10       $62,819     $204,284       $1,861
Africa Slipper        Addis Ababa   14       $68,641     $279,795       $1,771
Africa Sport Shoe     Addis Ababa    4        $1,690      $16,634          $79
Africa Women's Casual Addis Ababa    2       $51,541      $98,641         $940
Africa Women's Dress  Addis Ababa   12      $108,942     $311,017       $3,233
Africa Boot           Algiers       21       $21,297      $73,737         $710
Africa Men's Casual   Algiers        4       $63,206     $100,982       $2,221
```

The title and author (generated by code from the source file) as well as the date (automatically generated) are listed in the above portion of the document. Each line of SAS® code is marked with the indicator `SAS>` by default. Various options exist to hide this indicator or portions of code or output, store graphics files in a separate directory, change font size and color, etc.

Pagination in SASweave could be better. The output listed in the document is pulled directly from the SAS® output, and this includes indications that tables are continued on the next page. The portion of output below from the FREQ procedure is an example of how the pagination in SASweave is not always very smooth. The entire table is included on one page in the .pdf document, yet the table is split by an indication that it will be continued on the next page. This is accurate in the SAS® output, but not in the .pdf report:



```
Table of Region by Continent
Region            Continent(Continent)
Frequency        |Africa  |Asia    |North Am|Europe  |South Am|  Total
                 |        |        |erica   |        |erica   |
-----------------+--------+--------+--------+--------+--------+
Africa           |     56 |      0 |      0 |      0 |      0 |     56
-----------------+--------+--------+--------+--------+--------+
Asia             |      0 |     14 |      0 |      0 |      0 |     14
-----------------+--------+--------+--------+--------+--------+
Canada           |      0 |      0 |     37 |      0 |      0 |     37
-----------------+--------+--------+--------+--------+--------+
Central America/ |      0 |      0 |     32 |      0 |      0 |     32
Caribbean        |        |        |        |        |        |
-----------------+--------+--------+--------+--------+--------+
Eastern Europe   |      0 |      0 |      0 |     31 |      0 |     31
-----------------+--------+--------+--------+--------+--------+
Middle East      |      0 |     24 |      0 |      0 |      0 |     24
-----------------+--------+--------+--------+--------+--------+
Pacific          |      0 |     45 |      0 |      0 |      0 |     45
-----------------+--------+--------+--------+--------+--------+
South America    |      0 |      0 |      0 |      0 |     54 |     54
-----------------+--------+--------+--------+--------+--------+
United States    |      0 |      0 |     40 |      0 |      0 |     40
-----------------+--------+--------+--------+--------+--------+
Total                  56       83      109       93       54      395
(Continued)
Check variable creation
The FREQ Procedure
Table of Region by Continent
Region            Continent(Continent)
Frequency        |Africa  |Asia    |North Am|Europe  |South Am|  Total
                 |        |        |erica   |        |erica   |
-----------------+--------+--------+--------+--------+--------+
Western Europe   |      0 |      0 |      0 |     62 |      0 |     62
-----------------+--------+--------+--------+--------+--------+
Total                  56       83      109       93       54      395
Check variable creation
```

Print out MEDIAN(Q1, Q3) for continuous variables

```
SAS> proc means data = modify_shoes MEDIAN Q1 Q3 MAXDEC = 2;
SAS>    var Stores Sales Inventory Returns;
SAS>    title 'Median(Q1, Q3)';
SAS> run;
SAS> title;
```

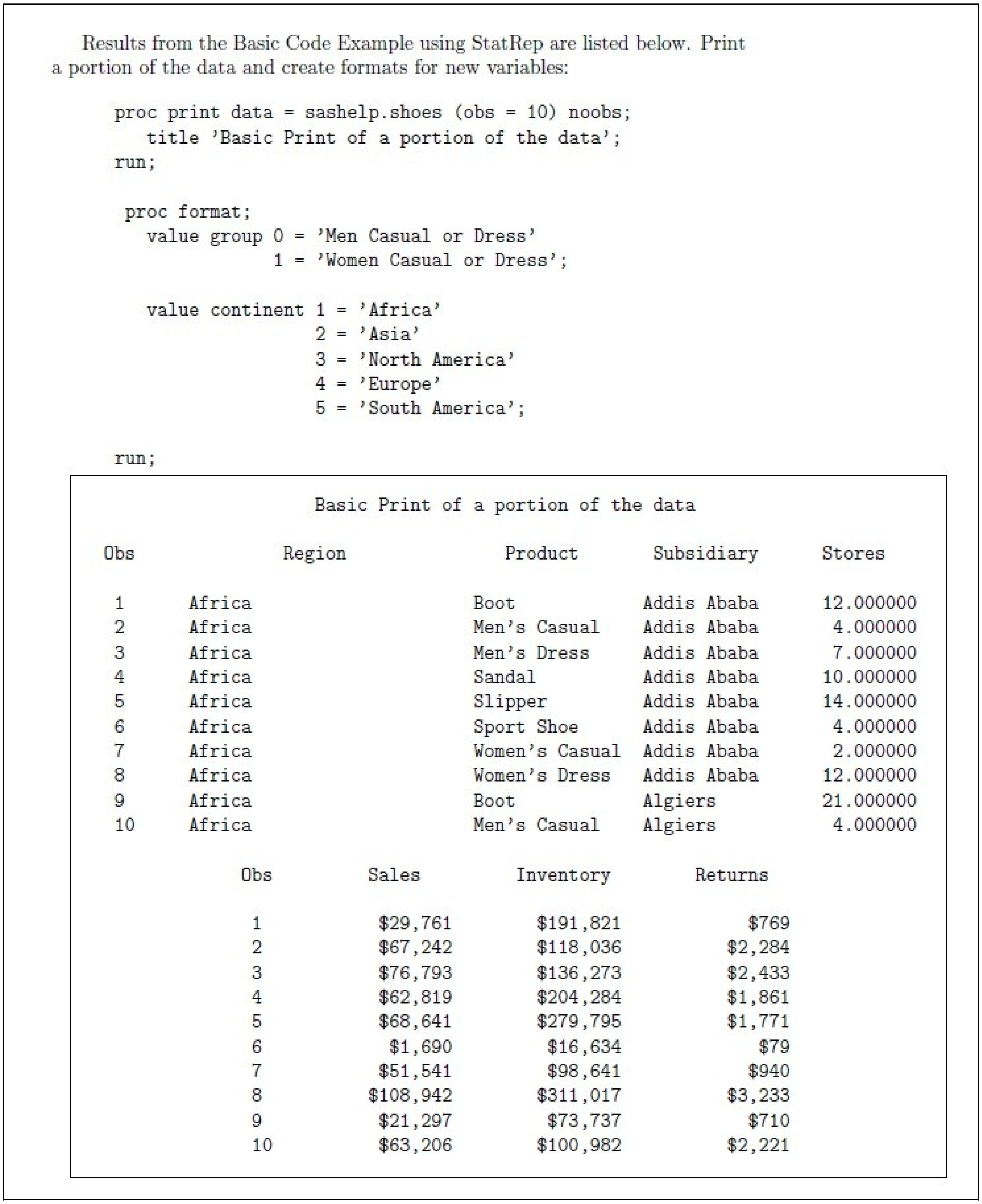
StatRep

Setting up code to run in StatRep is similar to what is involved with SASweave, with a few exceptions. Per StatRep documentation, code must be identified as being in the `Datastep` (reading in and manipulating data) or `Sascode` (generating output) environments, and output can be tagged as being a `Listing` (displaying tabular output or notes) or a `Graphic` (graphical output). Unlike SASweave, which automatically prints all output in the .pdf file unless it is instructed to hide portions, StatRep will not print any output to the .pdf report by default. The user must specify what they wish to have printed, creating a bit more work for the user, but not much.

The same code shown for SASweave in the above example is shown below, this time for StatRep.

```
\documentclass{article}
\usepackage{statrep}
\begin{document}

Results from the Basic Code Example using StatRep are listed below.

Print a portion of the data and create formats for new variables:

\begin{Sascode}[store=printOut]

proc print data = sashelp.shoes (obs = 10) noobs;
   title 'Basic Print of a portion of the data';
run;

proc format;
   value group 0 = 'Men Casual or Dress'
               1 = 'Women Casual or Dress';

   value continent 1 = 'Africa'
                   2 = 'Asia'
                   3 = 'North America'
                   4 = 'Europe'
                   5 = 'South America';
run;

\end{Sascode}

\Listing[store = printOut]{printOut}
```

The user must specify that is to use the StatRep LaTeX package in order for the tool to work properly. Note that there is a bit more work involved in setting up the StatRep code compared to setting up the SASweave code. In the above example, the output from the PRINT Procedure needed to be stored and then specified as a `Listing` in order to actually print. Specifying the PROC PRINT output as a Listing also ensures that it will print to a `lst` folder that is automatically created by StatRep within the main project folder. The output for the above portion of code begins on the next page:

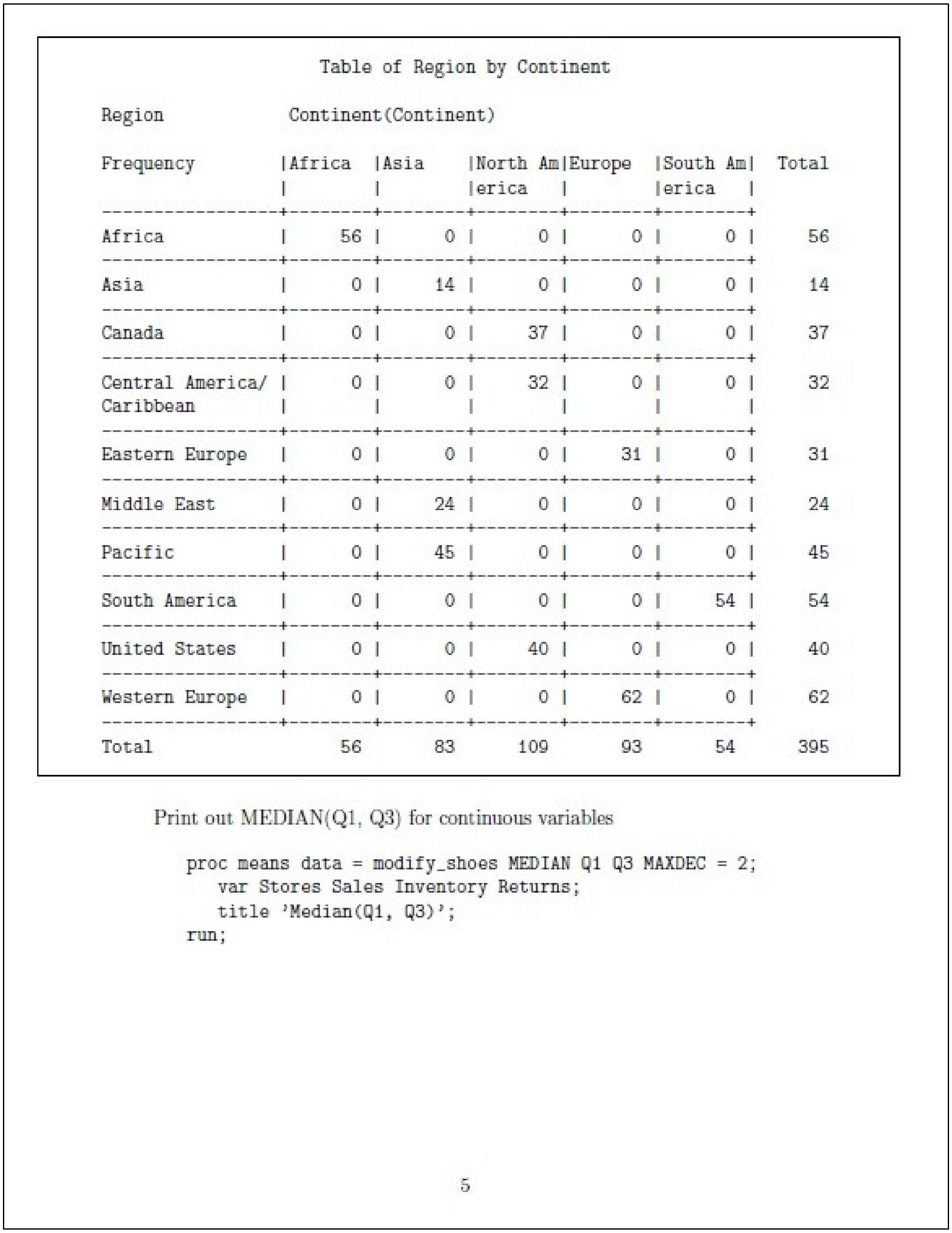


Results from the Basic Code Example using StatRep are listed below. Print a portion of the data and create formats for new variables:

```
proc print data = sashelp.shoes (obs = 10) noobs;
   title 'Basic Print of a portion of the data';
run;

 proc format;
   value group 0 = 'Men Casual or Dress'
               1 = 'Women Casual or Dress';

   value continent 1 = 'Africa'
                   2 = 'Asia'
                   3 = 'North America'
                   4 = 'Europe'
                   5 = 'South America';

run;
```

```
                      Basic Print of a portion of the data

 Obs             Region                  Product       Subsidiary      Stores

  1     Africa                        Boot           Addis Ababa     12.000000
  2     Africa                        Men's Casual   Addis Ababa      4.000000
  3     Africa                        Men's Dress    Addis Ababa      7.000000
  4     Africa                        Sandal         Addis Ababa     10.000000
  5     Africa                        Slipper        Addis Ababa     14.000000
  6     Africa                        Sport Shoe     Addis Ababa      4.000000
  7     Africa                        Women's Casual Addis Ababa      2.000000
  8     Africa                        Women's Dress  Addis Ababa     12.000000
  9     Africa                        Boot           Algiers         21.000000
  10    Africa                        Men's Casual   Algiers          4.000000

              Obs         Sales         Inventory        Returns

               1         $29,761        $191,821            $769
               2         $67,242        $118,036          $2,284
               3         $76,793        $136,273          $2,433
               4         $62,819        $204,284          $1,861
               5         $68,641        $279,795          $1,771
               6          $1,690         $16,634             $79
               7         $51,541         $98,641            $940
               8        $108,942        $311,017          $3,233
               9         $21,297         $73,737            $710
               10        $63,206        $100,982          $2,221
```

The output from running PROC PRINT actually begins on a new page, presumably because starting it on the same page of the .pdf document would have split the listing across two pages. Above it was moved right under the StatRep code for the purposes of this paper. In general, StatRep output seems nicer than SASweave output: code stands out and results are nicely centered and enclosed with a black border, making it easy to tell the code from the output itself. Like SASweave, comment font is different from code and output font, making the two easy to distinguish. The author and title commands don't work here like they do in SASweave, and similar coding to use in StatRep was not found. Pagination seems to work better in StatRep: the "continued on next page" indication that was in the SASweave output for one of the PROC FREQ tables disappears in StatRep, and the table is on one page:



```
                    Table of Region by Continent

Region           Continent(Continent)

Frequency        |Africa  |Asia    |North Am|Europe  |South Am|  Total
                 |        |        |erica   |        |erica   |
-----------------+--------+--------+--------+--------+--------+
Africa           |     56 |      0 |      0 |      0 |      0 |     56
-----------------+--------+--------+--------+--------+--------+
Asia             |      0 |     14 |      0 |      0 |      0 |     14
-----------------+--------+--------+--------+--------+--------+
Canada           |      0 |      0 |     37 |      0 |      0 |     37
-----------------+--------+--------+--------+--------+--------+
Central America/ |      0 |      0 |     32 |      0 |      0 |     32
Caribbean        |        |        |        |        |        |
-----------------+--------+--------+--------+--------+--------+
Eastern Europe   |      0 |      0 |      0 |     31 |      0 |     31
-----------------+--------+--------+--------+--------+--------+
Middle East      |      0 |     24 |      0 |      0 |      0 |     24
-----------------+--------+--------+--------+--------+--------+
Pacific          |      0 |     45 |      0 |      0 |      0 |     45
-----------------+--------+--------+--------+--------+--------+
South America    |      0 |      0 |      0 |      0 |     54 |     54
-----------------+--------+--------+--------+--------+--------+
United States    |      0 |      0 |     40 |      0 |      0 |     40
-----------------+--------+--------+--------+--------+--------+
Western Europe   |      0 |      0 |      0 |     62 |      0 |     62
-----------------+--------+--------+--------+--------+--------+
Total                  56       83      109       93       54      395
```

Print out MEDIAN(Q1, Q3) for continuous variables

```
proc means data = modify_shoes MEDIAN Q1 Q3 MAXDEC = 2;
   var Stores Sales Inventory Returns;
   title 'Median(Q1, Q3)';
run;
```

Note that using StatRep requires “SAS 9.2 or later, pdfLaTeX 1.30 or later, and the StatRep package itself” (Arnold and Kuhfeld 2012). The package will not run using earlier versions of SAS®.

In-house macros

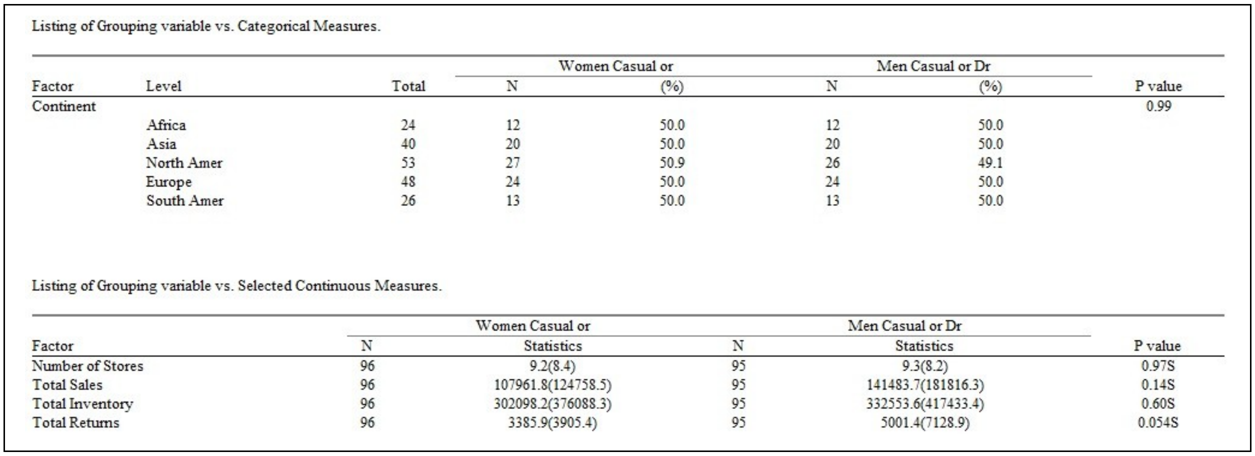
In addition to generating results from the basic code above, more complex, in-house macros were also run using both tools. The macros `%cattable` and `%conttable` were called using a `SASAUTOS` system option (see SAS® documentation on `SASAUTOS` for more information):

```
options sasautos = ('/home/mkarafa/GlobalIncludes/MyMacros',
                    "&AutosBase./sasautos",
                    "&AutosBase./SAS9.1BaseMacros",
                    "&AutosBase./SAS8BaseMacros",
                    sasautos
                   )
        implmac mautosource mrecall
        ;

%cattable(ds=modify_shoes, GroupVar = Product_group, vlist = Continent, owrite=T,
MainTitle = -1);

%conttable(ds = modify_shoes, GroupVar = Product_group, vlist = Stores Sales Inventory
Returns, owrite = F, MainTitle = -1);
```

As written above, the macro generates two forms of output. The listing from the SAS® output window is shown below:

```
      Listing of Grouping variable vs. Categorical Measures.

                             Women Casual or      Men Casual or Dr         P
   Factor   Level      Total         N        (%)          N         (%)  value

  Continent                .         .          .          .           . 0.99
            Africa       24         12       50.0         12        50.0
            Asia         40         20       50.0         20        50.0
            North Amer   53         27       50.9         26        49.1
            Europe       48         24       50.0         24        50.0
            South Amer   26         13       50.0         13        50.0

         Listing of Grouping variable vs. Selected Continuous Measures.

                         Women Casual or            Men Casual or Dr
      Effect        N    Statistics           N    Statistics            Pvalue

 Number of Stores   96   9.2(8.4)             95   9.3(8.2)              0.97S
 Total Sales        96   107961.8(124758.5)   95   141483.7(181816.3)    0.14S
 Total Inventory    96   302098.2(376088.3)   95   332553.6(417433.4)    0.60S
 Total Returns      96   3385.9(3905.4)       95   5001.4(7128.9)        0.054S
```

In this instance, an external .html file containing tables more closely formatted for inclusion in a memo is also produced:



Listing of Grouping variable vs. Categorical Measures.

			Women Casual or		Men Casual or Dr		
Factor	Level	Total	N	(%)	N	(%)	P value
Continent							0.99
	Africa	24	12	50.0	12	50.0	
	Asia	40	20	50.0	20	50.0	
	North Amer	53	27	50.9	26	49.1	
	Europe	48	24	50.0	24	50.0	
	South Amer	26	13	50.0	13	50.0	

Listing of Grouping variable vs. Selected Continuous Measures.

	Women Casual or		Men Casual or Dr		
Factor	N	Statistics	N	Statistics	P value
Number of Stores	96	9.2(8.4)	95	9.3(8.2)	0.97S
Total Sales	96	107961.8(124758.5)	95	141483.7(181816.3)	0.14S
Total Inventory	96	302098.2(376088.3)	95	332553.6(417433.4)	0.60S
Total Returns	96	3385.9(3905.4)	95	5001.4(7128.9)	0.054S

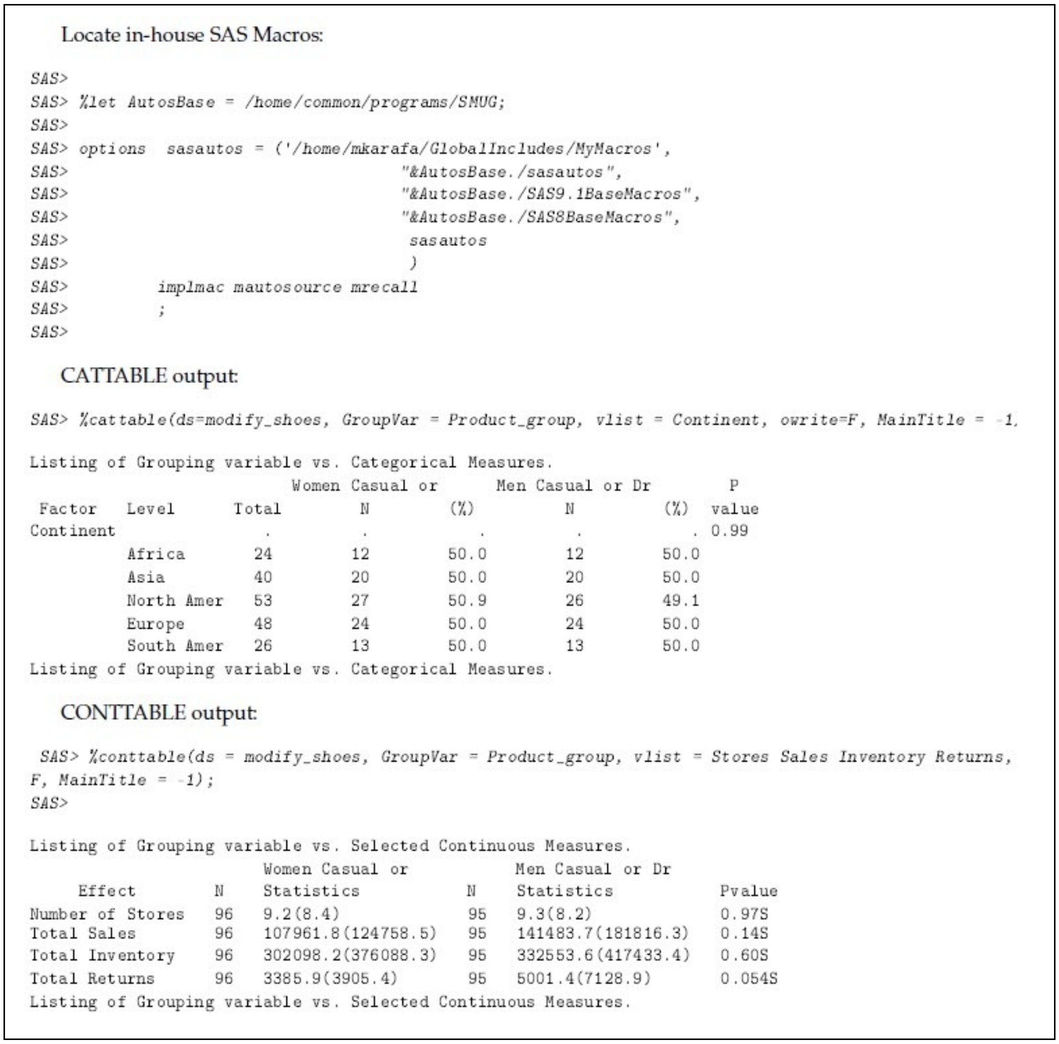
NOTE: The authors acknowledge that calling an outside macro and running it in SASweave and StatRep defeats the purpose of producing a reproducible document as the macro code is not visible…the macros were run purely out of curiosity.

SASweave

As with the basic code example above, preparing the SASweave source file is relatively straightforward when not using advanced SASweave options. Here is the portion of the source file with the code for using our advanced, in-house macros. Because this portion of the code is at the end of the document, an `\end{document}` statement is included:

```
Locate in-house SAS Macros:

\begin{SAScode}
   %let AutosBase = /home/common/programs/SMUG;
   options sasautos = ('/home/mkarafa/GlobalIncludes/MyMacros',
                       "&AutosBase./sasautos",
                       "&AutosBase./SAS9.1BaseMacros",
                       "&AutosBase./SAS8BaseMacros",
                       sasautos
                       )
   implmac mautosource mrecall
   ;
\end{SAScode}

CATTABLE output:

\begin{SAScode}

%cattable(ds=modify_shoes, GroupVar = Product_group, vlist = Continent, owrite=T,
MainTitle = -1);

\end{SAScode}

CONTTABLE output:

\begin{SAScode}

%conttable(ds = modify_shoes, GroupVar = Product_group, vlist = Stores Sales Inventory
Returns, owrite = F, MainTitle = -1);

\end{SAScode}

\end{document}
```

The above code ran fine in SASweave. Output is listed below (snipped from the output document above that contains the SASweave basic code results):



Locate in-house SAS Macros:

```
SAS>
SAS> %let AutosBase = /home/common/programs/SMUG;
SAS>
SAS> options  sasautos = ('/home/mkarafa/GlobalIncludes/MyMacros',
SAS>                               "&AutosBase./sasautos",
SAS>                               "&AutosBase./SAS9.1BaseMacros",
SAS>                               "&AutosBase./SAS8BaseMacros",
SAS>                                sasautos
SAS>                                )
SAS>         implmac mautosource mrecall
SAS>         ;
SAS>
```

CATTABLE output:

```
SAS> %cattable(ds=modify_shoes, GroupVar = Product_group, vlist = Continent, owrite=F, MainTitle = -1,

Listing of Grouping variable vs. Categorical Measures.
                          Women Casual or      Men Casual or Dr        P
 Factor   Level      Total        N        (%)          N         (%)  value
Continent                .        .          .          .           .  0.99
          Africa       24        12        50.0        12        50.0
          Asia         40        20        50.0        20        50.0
          North Amer   53        27        50.9        26        49.1
          Europe       48        24        50.0        24        50.0
          South Amer   26        13        50.0        13        50.0
Listing of Grouping variable vs. Categorical Measures.
```

CONTTABLE output:

```
 SAS> %conttable(ds = modify_shoes, GroupVar = Product_group, vlist = Stores Sales Inventory Returns,
F, MainTitle = -1);
SAS>

Listing of Grouping variable vs. Selected Continuous Measures.
                        Women Casual or          Men Casual or Dr
     Effect       N     Statistics          N    Statistics          Pvalue
Number of Stores  96    9.2(8.4)            95   9.3(8.2)            0.97S
Total Sales       96    107961.8(124758.5)  95   141483.7(181816.3)  0.14S
Total Inventory   96    302098.2(376088.3)  95   332553.6(417433.4)  0.60S
Total Returns     96    3385.9(3905.4)      95   5001.4(7128.9)      0.054S
Listing of Grouping variable vs. Selected Continuous Measures.
```

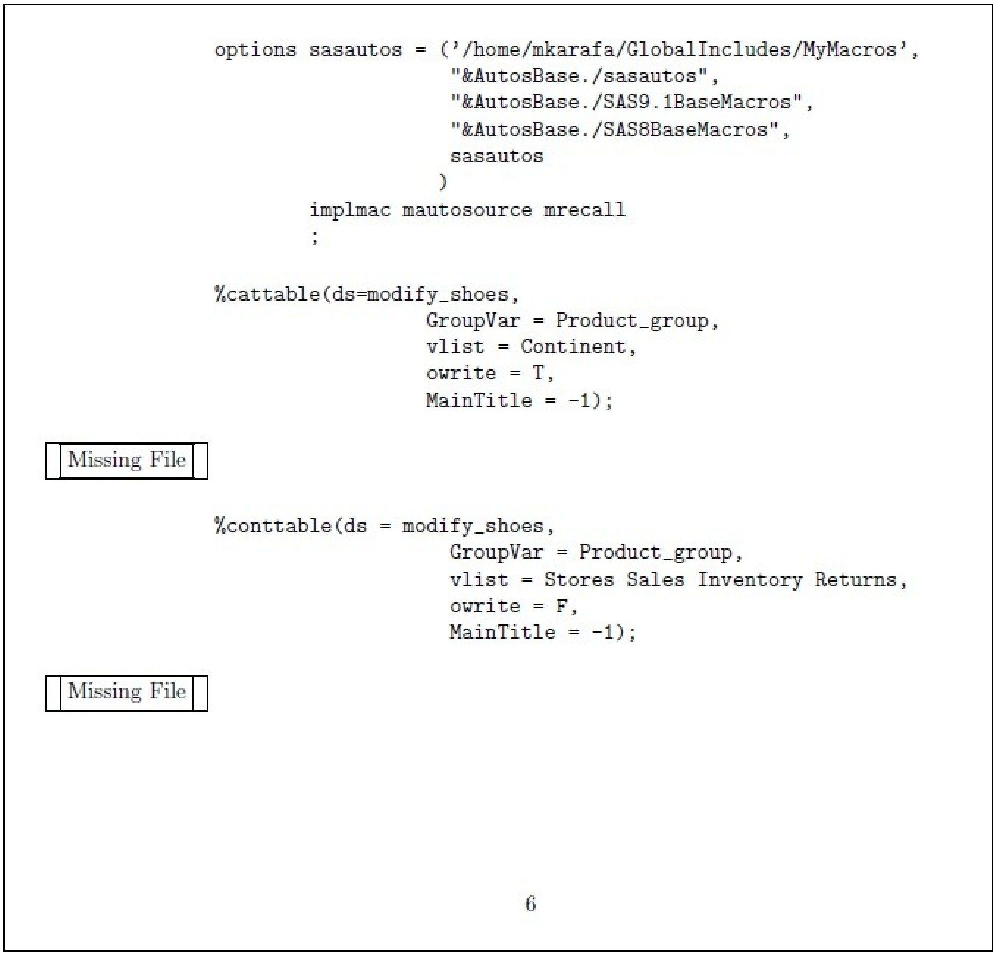
The output for both macros appears as expected, with SASweave using the listing generated in the SAS® output window. As mentioned earlier, the macro outputs temporary HTML files that also contain the results. These files were generated as expected and were saved in the folder along with the .pdf and .log files as indicated in the `%cattable` and `%conttable` macro calls.

One thing that is NOT generated from SASweave by default is a flat SAS® file containing the code run in the project. An additional step is needed to generate the flat SAS® file: the same source file must be run using SAStangle. SAStangle is included in the SASweave package download, and its sole purpose is to generate a flat SAS® file from the text source file. Like the SASweave file, SAStangle had to be opened and properly configured to work in the UNIX environment. Once configured, the `sastangle` command was submitted in the UNIX terminal window, generating the SAS® flat file. This file looks like a normal SAS® file, except that there are comments inserted that identify the "code chunks" in the program. SAStangle ran as expected.

StatRep

Preparing the StatRep source file to work with the in-house macros was challenging. Ultimately it was discovered that the output from the in-house macros could not come out as desired without making some major modifications to the in-house macro code itself. In particular, it seems that a complex macro that takes control of the ODS LISTING destination will not work well with StatRep, whose macros also control the ODS LISTING destination, turning it off and on at various points. $\LaTeX$ would not allow the .html output to be inserted into the document directly, and whole some alternative work-arounds were suggested to get the output into the document (for example, including the .csv file that resulted – however this produced a table with no spacing that was hard to read), the goal here was to see how it performed out of the box. It seems that some more complicated macros cannot be run using StatRep without some tweaking. From our testing this is especially true if the macro being used controls the ODS LISTING destination by turning it off and on. Since StatRep's output macros also affect the LISTNG, the two pieces step on each other, giving unpredictable output. Thus more complex macros may need to be written with running them in

StatRep in mind. In this case, the tables were output to .html files that were saved into the folder along with the .log and .pdf files as expected, but the tables never did get inserted into the .pdf document properly:



```
                options sasautos = ('/home/mkarafa/GlobalIncludes/MyMacros',
                                    "&AutosBase./sasautos",
                                    "&AutosBase./SAS9.1BaseMacros",
                                    "&AutosBase./SAS8BaseMacros",
                                    sasautos
                                   )
                        implmac mautosource mrecall
                        ;

                %cattable(ds=modify_shoes,
                                  GroupVar = Product_group,
                                  vlist = Continent,
                                  owrite = T,
                                  MainTitle = -1);
```

Missing File

```
                %conttable(ds = modify_shoes,
                                    GroupVar = Product_group,
                                    vlist = Stores Sales Inventory Returns,
                                    owrite = F,
                                    MainTitle = -1);
```

Missing File

6

Note that the simple macro using the FREQ procedure to check creation of new categorical variables `Product_Group` and `Continent` worked fine in StatRep. More complex macros do not seem to work in StatRep without revision of the code beforehand.

Unlike SASweave, StatRep generated a nice SAS® flat file from the .tex code automatically. This file was saved into the folder with the .pdf and .log files, and "_SR" was added to the file name. It is different from a regular SAS® file in that it has code to call the StatRep Preamble Program (an automatically generated program that calls the StatRep macros program) and it also uses the StatRep macros to output listings to the .lst folder using the `%output` macro.

Conclusion

It seems that both methods are useful for the same reason: a user ends up with an attractive set of documentation (.log and .pdf files with memo built in, as well as a flat SAS® file). However both require a possibly lengthy set-up as well as the time and desire to learn a new coding language to full take advantage of the packages (something that is not touched on in this paper).

The set-up process for SASweave can be difficult. It requires the user download a new program and configure their UNIX/online environment accordingly (something that was necessary for the authors and took a significant amount of time). The SASweave website seems to indicate that the most current version of SASweave is from August 2007; it is not known if further updates or coming or if support is available. SASweave processed all code and output it as expected, but parts of the output were hard to read and could have looked more polished. A flat SAS® file can be produced, but this requires running a separate tool, SAStangle, that must also be configured to the user's environment.

StatRep seems like a safer bet for most users and is preferred over SASweave by the authors. The biggest advantage of StatRep over SASweave is that it is actively supported and maintained by the SAS® Institute – it will continue to receive updates and help is just an email away. The user experience with StatRep is similar to SAS® -

no additional programs are needed, except for a typesetting program and a full LaTeX install. Other than these two considerations, set-up time is minimal and only requires a few steps that are outlined in the included README document. The output looks nicer and is easier to read compared to the SASweave output, and a flat SAS® file is automatically generated. Everything is stored in a common folder, making project documentation easier to locate. However StatRep seems to get tripped up when using complex macros that use the ODS LISTING destination. It is difficult to make StatRep work with these types of macros without editing the macro code itself or designing some method to incorporate the output into the resulting document (i.e. using something like an html-to-LaTeX filter).

Additionally, members of our department believe that sending a PDF file to our investigators (doctors, nurses, residents, fellows, etc.) would not be a welcome change. Team members usually send memos containing analysis results to investigators, who then copy and paste portions of the results directly into manuscripts that are submitted to journals. Investigators can still do that with .pdf documents, however they often lose table formatting. For our group, customer demand may outweigh the reproducibility gains when it comes to use of SASweave or StatRep.

References

(1) Arnold, Tim and Kuhfeld, Warren F. "Using SAS® and LaTeX to Create Documents with Reproducible Results." SAS® Global Forum. http://support.sas.com/resources/papers/proceedings12/324-2012.pdf (April 2012).

(2) Karafa, Matthew T. "Illustrating generation of MSWord tables via HTML with the %cattable and %conttable macros." MidWest SAS Users Group. <http://www.mwsug.org/proceedings/2006/stats/MWSUG-2006-SD02.pdf> (October 2006).

(3) Leisch, Friedrich. Sweave: Dynamic generation of statistical reports using literate data analysis. In Wolfgang Härdle and Bernd Rönz, editors, *Compstat 2002 – Proceedings in Computational Statistics,* pages 575-580. Physica Verlag, Heidelberg, 2002. ISBN 3-7908-1517-9.

(4) Lenth, Russell V. and Højsgaard Søren. 2007. "SASweave: Literate Programming Using SAS," *Journal of Statistical Software*, 19(8).

Acknowledgements

The authors wish to acknowledge Tim Arnold and Warren Kuhfeld for their quick feedback and willingness to help us better understand the StatRep package. Thank you for your help!

Contact Information

Your comments and questions are valued and encouraged. Contact the author at:

Name: Shannon Morrison, M.S.
Enterprise: Cleveland Clinic Foundation
Address: 9500 Euclid Avenue / JJN3-01
City, State ZIP: Cleveland, OH 44195
Phone: (216) 636-5413
Fax: (216) 444-8021
E-mail: morriss2@ccf.org
Web: http://lerner.ccf.org/qhs