Paper RIV-02

Not Enough Time to Catch Extreme Observations? Flag and Report with Macros and Arrays

Jamelle Simmons, California Polytechnic State University

ABSTRACT

Investigator-to-investigator variability of lab measurements can be an issue when all of the results are combined for analyses and ultimate publication or decision-making in areas such as quality control and product release or patient diagnosis. This paper examines methods for comparing multiple researcher measurements as compared to a “gold standard”. It also reports the individuals, and their measurements, that fall outside of acceptable ranges. As an example, histology data was collected in a skeletal tissue mechanics course over three academic quarters where students were required to take multiple bone section measurements and calculate densities; their results were compared with the professor’s standards as an exercise to demonstrate investigator variability. This project combines SAS macros, arrays, and reporting methods to identify the individuals whose measurements fell outside of an acceptable tolerance compared to the professor’s standards. The methods presented serve as means to verify investigator measurement accuracy and to help Principal Investigators or Quality Control Management more easily monitor lab technician or research investigator performance and training needs and ultimately determine if intervention is needed.

INTRODUCTION

When measurements are manually executed by multiple investigators, a number of factors could cause a wide results distribution for a given sample set. In the case of lab histological measurements, differences in results could be due to image clarity, chemical stains that bleed out, level of bone remodeling, investigator experience, calculation errors, and when there are no clear-cut remodeling landmarks it is up to the investigator to make a decision about how to proceed taking measurements. As an exercise in investigator variability, histological images were taken in white field and UV light (where chemical stains could be detected) and students were required to take a number of measurements using Image J, as well as use these results in formulas to calculate additional variables. A total of 142 investigator values were collected over 3 academic quarters on the same histology samples and compared to the professor’s measurements to give students an appreciation of how difficult these measurements could be for researchers in the field.

During each quarter, the professor checks each investigator’s values against their “gold standard” values in Microsoft Excel. This is time consuming with 40+ people in each class and a total of 10 variables to check per investigator as well as having to (1) decide the upper and lower limits of acceptability for each variable reported, (2) generate summary statistics and (3) identify the investigators that achieved results closest to the gold standards. This process could be sped up with the use of macros, arrays, and procedures such as PROC REPORT as well as by adding additional layers of information with data graphics as well as combining all information into a report, for example, PDF format to best organize generated output. Herein, a lab investigator evaluation reporting method is developed through the creation of global macro variables and data tables, the flagging of extreme observations and outliers, and the reporting of investigator success.

METHODS

Each year a sample of 16 images of 8 microscopic bone samples are given to students showing differing levels of osteonal remodeling under bright light and UV light. The students are tasked with using Image J to count/measure osteonal features of the slides to obtain several perimeter, thickness, and distance measurements on a micron scale and use these to calculate averages. These averages are then used to calculate apposition rates, formation rates, activation frequency, lag time, and time to refill bone molecular units (BMUs). As this is a demonstration of how difficult these measurements can be, even for experienced researchers in this field, it was ultimately left up to the students to decide how they would treat measurements situations based on what they had learned previously and their own best judgment. The final results from each student were entered into a Microsoft Excel sheet separated by tabs for each year this assignment was given. SAS 9.3 was then used to read and process all observations into a single document where investigators are grouped based on the accuracy of their histology results.

CREATING GLOBAL MACRO VARIABLES

Prior to working with the data, the first step is to create and store the gold standards for each variable as well as the percent deviation allowable for each variable. Placement of these standards at the beginning of the code allows for ease of use where users can change the standards as needed without having to search through code. The gold standards are established using a series of %LET statements (figure 1A). As long as the %LET statements are outside of DATA steps and %MACRO definitions, they are viewed as global variables. If these statements are used within a %MACRO, they are considered local and cannot be called outside of that macro unless the %GLOBAL statement is used. By using a SAS %MACRO with the %GLOBAL statement to create and store the variables globally (figure 1B), each standard is not limited by the DATA step or %MACRO definition it is called into and any changes made to this macro will extend to all parts of the code where these variables are used.

Users start with %MACRO, followed by the name they wish to give the macro and each variable's parameters. Next, a new variable name is specified for each variable after the %GLOBAL statement. The given parameters must then be assigned to the new global names by using the %LET statement, specifying the new variable names and setting them equal to the previously defined macro parameters; an ampersand (&) must be used in front of each macro variable parameter that is called. Lastly, a %MEND is needed to close out the macro and the macro name must also be specified. Users can change parameters within the macro and these changes are carried throughout the programming code.

A

```
%LET GSr=471; *Avg. resorp. space perimeters (µm);
%LET GSf=305;  *Avg. osteoid seam perimeters (µm);
%LET GTos=6.7;  *Avg. osteoid thicknesses (µm);
%LET GDl=20;    *Avg. distance b/w labels (µm);
%LET GTmw=43.8; *Avg. mean wall thickness (µm);

%LET GMf=2;     *Mineral apposition rate (µm/day);
%LET GTml=3.3;  *Mineral lag time (days);
%LET GTf=21.9;  *Time to refill a BMU (days);
%LET Gfa=0.053;  *Activation freq. (BMUs/mm²/day);
%LET GVf=0.256; *Bone formation rate (mm²/mm²/year);
%LET Gpercent=35; *Max % dev. from gold standard;
```

B

```
%MACRO GLB (GSr=471, GSf=305, GTos=6.7, GDl=20,
           GTmw=43.8, GMf=2, GTml=3.3, GTf=21.9,
           Gfa=0.053, GVf=0.256, Gpercent=35);

%GLOBAL PSr PSf PTos PDl PTmw PMf PTml PTf Pfa PVf
percent;

%LET PSr=&GSr;   %LET PSf=&GSf;   %LET PTos=&GTos;
%LET PDl=&GDl;   %LET PTmw=&GTmw; %LET PMf=&GMf;
%LET PTml=&GTml; %LET PTf=&GTf;   %LET Pfa=&Gfa;
%LET PVf=&GVf;   %LET percent=&Gpercent;

%MEND GLB;
```

Figure 1. (A) Gold standard macro variables, global if outside of DATA steps and %MACRO definitions, for (bone measurement experiment) and (B) Macro to house all gold standards and make them available globally.

CREATION OF DATA TABLE FROM EXCEL SHEETS

Each quarter's data is stored in a separate Microsoft Excel® sheet within the same workbook where each sheet is named hist 2011, hist 2012, and hist 2013. A macro named CALLIT (figure 2) is created with parameters for year and the data set name. In figure 2, the macro (1) points to the Excel Workbook with the individual sheet, (2) changes the year in the FILENAME path with &year used to help specify each worksheet, (3) names the data set in the DATA step with &hist, (4) reads in the investigator variables from each worksheet and labels them, (5) and creates a new column with the year in the resulting SAS data set created from each Excel sheet.

As academic course sizes are similar, a wide cell range has been specified, row 3 column 1 to row 53 column 11 (R3C1:R53C11) within each Excel sheet, where if SAS runs out of data before it reaches the 53rd row, the read-in terminates and the data set is finished. The next portion of the code uses PROC APPEND to create a base file that each SAS data set created will be appended to. This entire procedure is started with the %MACRO definition and ended with %MEND. The macro CALLIT is called a total of 3 times, after the procedure has been defined, to read in each sheet of the Excel file and build the base file. It should be noted that each time the script file is run, the base file must be deleted beforehand or else duplicate data sets will be appended to the existing file instead of overwriting them.

```
%MACRO CALLIT(YEAR=,HIST=);

FILENAME xls DDE
"Excel|C:\Users\Lenovo\Desktop\SAS Flag\[Histology
Assignment.xlsx]hist&year!R3C1:R53C11";

DATA &hist;
INFILE xls NOTAB DLM='09'x DSD MISSOVER;
INPUT Invest Sr Sf Tos Dl Tmw Mf Tml Tf fa Vf;
class_yr="&year";
LABEL Invest='Investigator'
      Sr='Resorption Space Perimeter (Avg, um)'
         Sf='Osteoid Seam Perimeter (Avg, um)'
         Tos='Osteoid Thickness (Avg, um)'
         Dl='Distance b/w Labels (Avg, um)'
         Tmw='Mean Wall Thickness (Avg, um)'
         Mf="Mineral Apposition Rate (um/day)"
         Tml='Mineralization Lag Time (days)'
         Tf='Time to Refill a BMU (days)'
         fa="Activation Frequency (BMUs/mm2/day)"
         Vf="Bone Formation Rate (mm2/mm2/year)"
         class_yr='Year';
RUN;

PROC APPEND BASE=histpath.histcombo DATA=&hist;
RUN;

%MEND CALLIT;

%CALLIT(YEAR=2011,HIST=HIST1);
%CALLIT(YEAR=2012,HIST=HIST2);
%CALLIT(YEAR=2013,HIST=HIST3);
```

Parameters for the macro CALLIT are passed throughout code. Values are specified each time the macro is called

Can specify a cell range larger than actual data so that a range does not need to be specified each time the macro is run

The Excel tabs for the data are labeled hist 2011, hist 2012, and hist 2013. When the macro is invoked, year= will specify the tab to read and hist= will name the data set that is created

When year is specified and invoked by the macro, a new column with the year will be created in the data set

Each time a new hist data set it created, it is appended to a base file where histpath specifies the location to write the file to and histcombo becomes the name of the base file

Signals the end of the macro CALLIT

The macro CALLIT is invoked three times to read in all tabs from the Excel sheet and create separate data set that are appended to the base file each time the macro is run

Figure 2. Macro named CALLIT to read in and create a base file for all investigator data over the three academic quarters.

FLAG EXTREME OBSERVATIONS

Initially, the macro GLB was placed at the top of the script file before any DATA steps and was processed but up to this point has not been called. This macro is now placed inside of the DATA step HISTCOMBO2 where the user-defined gold standards are first needed. A temporary array named GOLDTEMP is created with each of the gold standard values for each variable; this array is used to help calculate the upper and lower limits of acceptability. The temporary arrays GOLDLOW and GOLDHIGH houses the lowest and highest acceptable values per variable respectively. The array STUDENT holds the investigator values for each variable and the array FLAG contains a numeric value that signals when values are either outside or inside acceptable limits for each variable. A DO LOOP is created which generates the highest and lowest possible values based on the set percentage limit in the GLB macro. In this example a value of 35 is used where values 35% below and above the gold standards are held in the arrays GOLDLOW and GOLDHIGH.

If an investigator's reported results fall below the GOLDLOW limit, a flag value of -1 is given. If their reported results fall above the GOLDHIGH limit, a flag of +1 is given; else a flag value of 0 is given. Finally, the variable in_range is created as a counter that tallies the number of investigator values that are not within the upper and lower limits. This variable can be used to separate those investigators who produced acceptable results for all variables from those individuals who had one or more variables where their values were out of range (figure 3). At the end of the data set, the temporary arrays are automatically deleted to reduce the size of the SAS data set and the DROP statement is used to drop unnecessary counters I and J.

```
DATA HISTCOMBO2;
SET HISTPATH.HISTCOMBO;

%GLB;

ARRAY GOLDTEMP (10)
_TEMPORARY_(&PSR, &PSF, &PTOS, &PDL, &PTMW,
            &PMF, &PTML, &PTF, &PFA, &PVF);

ARRAY GOLDLOW (10)_TEMPORARY_;
ARRAY GOLDHIGH (10)_TEMPORARY_;

ARRAY STUDENT (10) Sr Sf Tos Dl Tmw Mf Tml Tf fa Vf;
ARRAY FLAG (10) $25;

DO I=1 TO 10;
GOLDLOW(I)=GOLDTEMP(I)-(GOLDTEMP(I)*(&PERCENT/100));
GOLDHIGH(I)=GOLDTEMP(I)+(GOLDTEMP(I)*(&PERCENT/100));

  IF STUDENT(I) < GOLDLOW(I) THEN FLAG(I)=-1;
  ELSE IF STUDENT(I) > GOLDHIGH(I) THEN FLAG(I)=1;
  ELSE FLAG(I)=0;
END;

IN_RANGE=0;
DO J=1 TO 10;
      IF FLAG(J)^=0 THEN DO;
            IN_RANGE+1;
      END;
END;
DROP I J;
RUN;
```

The Macro containing all of the histologic standards and percent acceptability limits is called

Creating temporary arrays for each standard specified in the macro GLB. Standards are renamed GOLDTEMP1, GOLDTEMP2, etc.

Creating blank temporary arrays for lower and upper acceptability limits that investigator values will be compared to. GOLDLOW represent lowest acceptable values and GOLDHIGH represent highest acceptable values

Existing investigator values for each variable are turned into an array

The loop is carried out ten times, once for each histologic variable

GOLDLOW(I) finds the specified percentage below the standard for each variable. GOLDHIGH(I) finds the specified percentage above each variable. These define the lower and upper limits

For each investigators values per variable, if it falls below the lower limit a flag of -1 is produced. If values fall above the limit a flag of +1 is produced, else a flag of 0 is produced

A counter, in_range, is made where a count is tallied for each investigator values per variable when a value other than 0 is produced.

Figure 3. Flagging observations outside of the pre-set acceptable limits specified in the global macro.

OUTPUT DELIVERY SYSTEM (ODS) PDF

The default for SAS 9.3 sets the output to HTML format which can be changed by either opening additional output locations or closing current destinations and opening others. In figure 4, the LISTING and HTML output destinations are closed and the ODS PDF destination is opened. The basic syntax requires that a file path is specified followed by the name of the document to be created, then .pdf. The LIBNAME specified for this paper was named HISTPATH (figure 2) and was used in the previous code sections; as this path is still in use, the PDF will be written to that path and given the name “histology_class results.” Options for a PDF such as NONUMBER and NODATE removes the default page numbers and date from the PDF which are normally present in SAS output. With the path still open, all procedures created will be saved to the PDF until the user ends the flow of information to the PDF with the syntax ODS PDF CLOSE. New destinations can then be reopened by first specifying ODS followed by the output location.

```
ODS LISTING CLOSE;
ODS HTML CLOSE;
OPTIONS NONUMBER NODATE ORIENTATION=LANDSCAPE;
ODS PDF FILE='histology_class results.pdf';

ODS PROCLABEL="Investigator Summary Statistics by Year";
ODS NOPTITLE;
PROC MEANS DATA=HISTCOMBO2;
VAR Sr Sf Tos Dl Tmw Mf Tml Tf fa Vf;
CLASS class_yr;
TITLE "Investigator Statistics by Year";
FOOTNOTE "Missing Values Not Included";
run;

ODS PROCLABEL="Investigators Outside Gold Standard Ranges";
PROC REPORT DATA=HISTCOMBO2 NOWD STYLE(COLUMN)={JUST=RIGHT CELLWIDTH=.7IN}
      SPLIT='|' CONTENTS="Investigator Outliers";
      WHERE in_range^=0;
COLUMNS class_yr Invest Sr Sf Tos Dl Tmw Mf Tml Tf fa Vf in_range;
DEFINE class_yr / ORDER STYLE(COLUMN)={JUST=RIGHT CELLWIDTH=.5IN};
DEFINE Invest / DISPLAY STYLE(COLUMN)={JUST=CENTER CELLWIDTH=.95IN};
DEFINE Sr / DISPLAY FORMAT=5.0 'Sr (Avg,um)' ;
DEFINE Sf / DISPLAY FORMAT=5.0 'Sf (Avg,um)' ;
DEFINE Tos /DISPLAY FORMAT=5.2 'Tos (Avg,um)';
DEFINE Dl / DISPLAY FORMAT=5.1 'Dl (Avg,um)' ;
DEFINE Tmw /DISPLAY FORMAT=5.2 'Tmw (Avg,um)';
DEFINE Mf /DISPLAY  FORMAT=5.2 'Mf (um/day)' ;
DEFINE Tml /DISPLAY FORMAT=5.1 'Tml (days)'  ;
DEFINE Tf /DISPLAY  FORMAT=5.2 'Tf (days)'   ;
DEFINE fa /DISPLAY  FORMAT=5.3 'fa (BMUs/mm^2/day)' STYLE(COLUMN)={JUST=RIGHT CELLWIDTH=.7IN};
DEFINE Vf / DISPLAY FORMAT=5.3 'Vf (mm^2/mm^2/year)'STYLE(COLUMN)={JUST=RIGHT CELLWIDTH=.7IN};
DEFINE in_range / DISPLAY      'No. Outliers' STYLE(COLUMN)={JUST=CENTER BACKGROUND=LIGHTBLUE};

%MACRO COLOR(STU=, GOLDS=);
COMPUTE &STU;
    IF &STU=. THEN CALL DEFINE(_COL_,"STYLE","STYLE={BACKGROUND=ORANGE}");
    ELSE IF &STU GT &GOLDS+&GOLDS*&PERCENT/100 THEN CALL DEFINE
(_COL_,"STYLE","STYLE={BACKGROUND=YELLOW}");
    ELSE IF &STU LT &GOLDS-&GOLDS*&PERCENT/100 THEN CALL DEFINE
(_COL_,"STYLE","STYLE={BACKGROUND=GREEN}");
ENDCOMP;
%MEND COLOR;

%COLOR(STU=SR, GOLDS=&PSR);   %COLOR(STU=SF, GOLDS=&PSF);   %COLOR(STU=TOS, GOLDS=&PTOS);
%COLOR(STU=DL, GOLDS=&PDL);   %COLOR(STU=TMW, GOLDS=&PTMW); %COLOR(STU=MF, GOLDS=&PMF);
%COLOR(STU=TML, GOLDS=&PTML); %COLOR(STU=TF, GOLDS=&PTF);
%COLOR(STU=FA, GOLDS=&PFA);   %COLOR(STU=VF, GOLDS=&PVF);

TITLE1 bold FONT=times H=12 'Histologic Gold Standards';
TITLE3 FONT=times H=5 "Sr= &PSr, Sf= &PSf, Tos= &PTos, Dl= &PDl, Tmw= &PTmw,
                       Mf= &PMf, Tml= &PTml, Tf= &PTf, fa= &Pfa, Vf= &PVf";
TITLE5 FONT=times H=5 "Orange Cells=Missing, Yellow=Above Standards, Green=Below Standards";

FOOTNOTE1 bold FONT=times H=10 'Histologic Abbreviations';
FOOTNOTE3 FONT=times H=5 'Sr=Resorption Space Perimeter, Sf=Osteoid Seam Perimeter,
                          Tos=Osteoid Thickness';
FOOTNOTE5 FONT=times H=5 'Dl=Distance b/w Labels, Tmw=Mean Wall Thickness, Mf=Mineral
                          Apposition Rate';
FOOTNOTE7 FONT=times H=5 'Tml=Mineralization Lag Time, Tf=Time to Refill a BMU,
                          fa=Activation Frequency, Vf=Bone Formation Rate';
BREAK BEFORE class_yr/ CONTENTS="By Class Year" PAGE;
RUN;
ODS PDF CLOSE;
ODS LISTING;
```

Figure 4. Generate and store summary statistics and flagged observations in PDF format

PROC MEANS IN THE ODS PDF PATH

With the ODS PDF path still open, the created data set HISTOCOMBO2 is used in a PROC MEANS step where only the histological variables specified by the VAR statement are called. The summary statistics (N, Mean, Std. Dev., Min, Max) are separated into groups using the CLASS Statement for the years 2011-2013; also, the variable labels, that were created when the data set was initially made, are displayed in the output. This PROC MEANS procedure does not include missing values from investigators in its summary statistics. By default, the title "Proc Means Procedure" appears in the PDF output but this can be turned off with the option ODS NOPTITLE (figure 4). To name the bookmark associated with this procedure in the PDF, the option ODS PROCLABEL can be used.

PROC REPORT PDF TABLE FORMATTING

PROC REPORT gives users the flexibility of table formatting, variable manipulation, text usage, and the ability to highlight important observations. With the HISTCOMBO2 data set (figure 4), the NOWD option runs the reporting procedure and sends it directly to the PDF output destination (here, the PDF). By specifying,

STYLE (COLUMN)={JUST=RIGHT CELLWIDTH=0.7IN},

all columns are given the column width of 0.7 inches and the text within these columns will be right justified unless additional column options are specified (figure 5). The SPLIT option has a **" / "** default where if used in text, places information that follows it on a new line in the column. For example, with the variable VF, the text used is 'Vf (mm^2/mm^2/year). With the current default delimiter, this would place "Vf (mm^2)" on one line, followed by mm^2 on the next line, and year on the last line (figure 5). To correct for this, the delimiter was changed to **" | "**. By using the WHERE statement, only those investigators who had at least one measurement/calculation outside of the 35% limit were used in this section of the report. Next, only the variables that are desired in the report are listed next to the COLUMNS statement.

Histologic Gold Standards

Sr= 471, Sf= 305, Tos= 6.7, Dl= 20, Tmw= 43.8, Mf= 2, Tml= 3.3, Tf= 21.9, fa= 0.053, Vf= 0.256

Orange Cells=Missing, Yellow=Above Standards, Green=Below Standards

Year	Investigator	Sr (Avg, um)	Sf (Avg, um)	Tos (Avg, um)	Dl (Avg, um)	Tmw (Avg, um)	Mf (um/ day)	Tml (days)	Tf (days)	fa (BMUs/ mm^2/ day)	Vf (mm^2/ mm^2/ year)	No. Outliers
2011	1	.	481	4.35	14.9	46.68	1.49	2.9	31.31	0.038	0.314	4
	2	414	267	10.28	20.4	38.08	2.04	5.0	18.70	0.059	0.219	2
	3	549	519	6.38	18.8	52.80	1.88	3.4	28.11	0.041	0.410	2
	4	505	.	10.78	17.3	29.49	1.73	4.5	17.00	0.085	0.358	5
	6	415	160	3.10	19.0	56.09	1.90	1.6	29.50	0.051	0.457	4
	7	457	239	2.07	17.9	47.40	1.79	11.6	26.48	0.047	0.195	2
	10	421	256	3.25	13.7	12.32	1.37	2.3	9.01	0.139	0.160	5
	12	461	234	6.41	18.1	64.64	1.81	3.6	35.80	0.034	0.185	3
	13	500	225	9.20	15.0	38.00	1.50	6.3	26.20	0.052	0.162	3
	14	245	208	7.60	16.4	33.00	1.64	4.6	20.10	0.087	0.218	3
	15	434	229	16.20	18.5	44.10	1.85	8.8	23.90	0.052	0.193	2
	16	446	304	9.26	13.1	19.48	1.31	7.1	14.88	0.077	0.167	4
	18	.	208	8.30	18.4	68.90	1.84	4.5	37.40	0.029	0.154	6
	19	496	256	6.03	16.9	136.0	1.69	3.6	80.60	0.019	0.237	3
	20	.	203	6.90	18.2	23.89	1.82	3.8	13.10	0.012	0.219	4
	21	495	192	4.50	11.0	13.60	1.10	4.1	12.36	0.015	0.439	7
	22	494	320	10.60	18.9	56.90	1.89	5.6	31.10	0.042	0.276	3

Histologic Abbreviations

Sr=Resorption Space Perimeter, Sf=Osteoid Seam Perimeter, Tos=Osteoid Thickness

Dl=Distance b/w Labels, Tmw=Mean Wall Thickness, Mf=Mineral Apposition Rate

Tml=Mineralization Lag Time, Tf=Time to Refill a BMU, fa=Activation Frequency, Vf=Bone Formation Rate

Figure 5. Formatting and highlighting in PROC REPORT

If users wish to format and manipulate a variable, they can define the variable followed by its specifications. For example, by default, the academic year would appear next to each investigator number and could be distracting. By specifying the ORDER option (figure 4), only the first observation per new academic year will show the year that the data was collected. All other investigators under the same year would have an empty cell for their academic year variable (figure 5).

As significant figures may be an issue with numerous investigators following different guidelines for rounding, a format can be applied to variables in the PROC REPORT options by using FORMAT=. When using the FORMAT option, the width and decimal place must be given. For the variable MF, the FORMAT 5.2 is given and an investigator value of 1.4909 would be cut down to 1.49 giving a total width of 4 carried out to the second decimal place (figure 5).

The IN_RANGE variable that was originally created to help separate those investigators can now be used as a summed field to let users know how many variables fell out of range per investigator. To help this column stand out in the PDF, the following syntax was used,

STYLE (COLUMN)={JUST=CENTER BACKGROUND=LIGHTBLUE} and its results can be seen in figure 5.

FLAGGING VALUES WITH MACROS (PROC REPORT)

Every investigators recorded values can fall into four possible categories per variable (missing, below range, within range, or above range). With a total of 10 variables, this would make for a lot of code but macros can be used to reduce typing. In the PROC REPORT step in figure 4, the macro COLOR was created with the parameters (STU= , GOLDS=) where STU is set equal to the investigators recorded value and GOLDS is set equal to the gold standard. To flag values in the PROC REPORT step the COMPUTE option and the variable we wish to use (figure 4) can be specified in an IF/ELSE-IF series. For values that are missing per investigator, those cells are color coded orange with the following syntax,

IF &STU=. THEN CALL DEFINE (_COL_,"STYLE","STYLE={BACKGROUND=ORANGE}").

Investigator values greater than the upper limit of the gold standard are color coded yellow with the following syntax,

ELSE IF &STU GT &GOLDS + &GOLDS * &PERCENT/100 THEN CALL DEFINE (_COL_,"STYLE","STYLE={BACKGROUND=YELLOW}").

The syntax is similar for coloring cells less than the lower gold standard limit (figure 4). To close the procedure, ENDCOMP is used when the computation is finished, and the macro is ended. Now, the macro COLOR can be invoked ten times by adjusting the parameters rather than using copy/paste and typing the changes into each repeated procedure which could lead to errors. This macro will produce the colored output of figure 5 for each histologic variable.

ADDITIONAL OPTIONS FOR PDF (PROC REPORT)

With the code in figure 4 creating pages of output, it is important to label the procedure bookmarks in the PDF file for easier navigation. By using ODS PROCLABEL, the highest level of the bookmark is labeled and the second level of the bookmark can be labeled using the CONTENTS option. To clean up the PDF, page breaks are added when SAS reaches a new academic year in the data set. With the following syntax (figure 4),

BREAK BEFORE class_yr/ CONTENTS="By Class Year" PAGE,

each page break occurs before each new class year and the CONTENT option names the 3rd level of the bookmark. For the PROC REPORT procedure, TITLE and FOOTNOTE options are used instead of the ODS PDF TEXT= option. By specifying odd numbered titles and footnotes (figure 4), spaces between lines in the resulting PDF are created (figure 5). Additional formatting options are used to control the font size (FONT), boldness of the font (BOLD), and height (H) of the text which appear on each page of the procedure by using syntax similar to the following,

TITLE1 BOLD FONT=times H=12 'Histologic Gold Standards'.

SAS® ODS GRAPHICS DESIGNER

In addition to tables, it may be of use to include graphics which display summary data and groupings by year for reporting purposes. With ODS Graphics Designer, code can be generated which can later be used in a SAS script for multiple variables. Using the MF (Mineral Apposition Rate) variable as an example it will be shown how to format figures to generate code for inclusion in a SAS script. After opening the ODS Graphics Designer, users can select the Histogram icon followed by the library the data is stored in and the data set to be used. For this paper, the Work Directory is used. Next, by selecting Insert>Row at the top of the window, a blank object is generated below the histogram. Users can then drag and drop the Box (H) object into the blank object and select the Y variable as MF and the X variable as Class_yr in the Assign Data box (figure 6). A normal curve is selected in the left pane and dragged onto the histogram and title text is added. Users can also specify additional options as well as add/remove objects and options as they see fit. To see the code generated from formatting, users select View>Code (figure 7) and the displayed code can be copied and pasted into a SAS script.

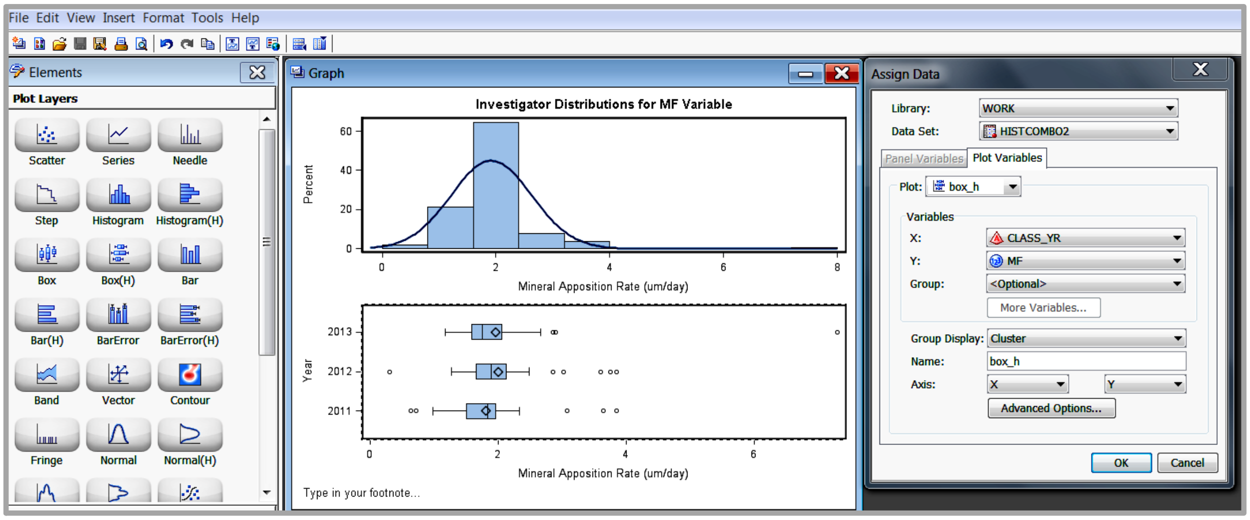

Figure 6. ODS Graphics Designer formatting

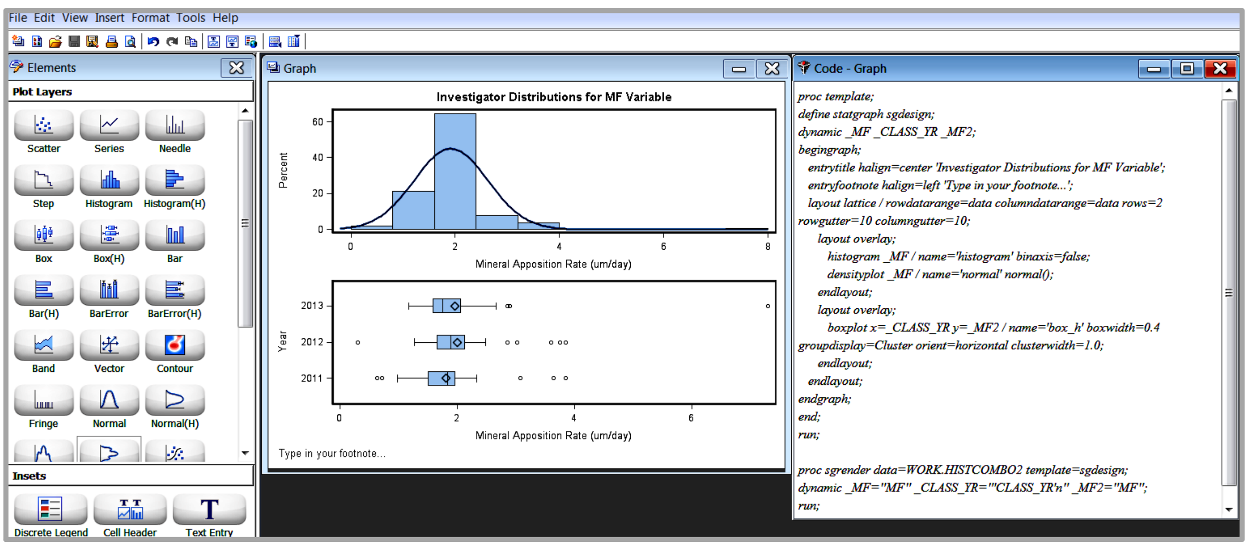

Figure 7. ODS Graphics Designer code viewing

ODS GRAPHICS MACRO REPORTING

Again, as there are a number of variables that should be included in the report, a macro can be used to reduce the amount of coding or else the steps in figures 6 and 7 will need to be repeated by hand increasing the chance of mistakes. Instead, the ODS Graphics Designer code can be copied and placed inside a macro where the names of the variables are swapped out with macro parameter names. The macro can then be called ten times, substituting variables names each time where graphics are produced with the same formatting. To add to the existing PDF created in figure 4, the macro PLOTS (figure 8) can be inserted before the ODS PDF CLOSE statement which will generate plots such as that seen in figure 9. The macro parameters VARB= and VARB2= in figure 8 refer to the values collected from the investigators and the parameter DAT= refers to the gold standard value for the variable graphics. It should be noted that some features of SAS® ODS Graphics Designer do not carry over to PDF format, such as some color styles.

```
%MACRO PLOTS(VARB=, VARB2=, DAT=);

ODS PROCLABEL="Investigator Distributions for &varb Variable (Gold Standard=&dat)";
PROC TEMPLATE;
DEFINE STATGRAPH GRAPH;
DYNAMIC _&VARB _&VARB2 _CLASS_YR2;
BEGINGRAPH / designwidth=736 designheight=578;

entrytitle halign=center "Investigator Distributions for &varb Variable (Gold Standard=&dat)" /
textattrs=(family='Times New Roman' );

   LAYOUT LATTICE / rowdatarange=data columndatarange=union rows=2 rowgutter=10 columngutter=10;

     LAYOUT OVERLAY / wallcolor=CXFAF5E5 xaxisopts=( display=(TICKS LINE ) tickvalueattrs=(family='Times
      New Roman' )) yaxisopts=( labelattrs=(family='Times New Roman' size=12 style=NORMAL
      weight=BOLD) tickvalueattrs=(family='Times New Roman' size=10 ));
        HISTOGRAM _&varb / name='histogram' binaxis=false outlineattrs=(thickness=2 );
        DENSITYPLOT _&varb / name='normal' normal() lineattrs=(pattern=LONGDASH thickness=2 );
            entry halign=right 'Histogram of All Years' / valign=top
            textattrs=(family='Times New Roman' size=12 style=normal weight=bold );
      ENDLAYOUT;

      LAYOUT OVERLAY / wallcolor=CXFAF5E5 walldisplay=(OUTLINE FILL)
            xaxisopts=(labelattrs=(family='Times New Roman' size=12 style=NORMAL weight=BOLD )
            tickvalueattrs=(family='Times New Roman' )) yaxisopts=( labelattrs=(family='Times New Roman'
            size=12 style=NORMAL weight=BOLD )tickvalueattrs=(family='Times New Roman' size=10 ));
      BOXPLOT x=_CLASS_YR2 y=_&varb2 / name='box_h' boxwidth=0.4 groupdisplay=Cluster
            orient=horizontal clusterwidth=1.0 outlineattrs=(thickness=2 )
            medianattrs=(thickness=2 ) whiskerattrs=(thickness=2 ) meanattrs=(weight=bold )
            outlierattrs=(color=CXFF0000 size=9 weight=bold );
      endlayout;
      columnaxes;
         columnaxis / labelattrs=(family='Times New Roman' );
      endcolumnaxes;
   ENDLAYOUT;

ENDGRAPH;
END;
RUN;

PROC SGRENDER DATA=WORK.HISTCOMBO2 TEMPLATE=GRAPH;
DYNAMIC _&VARB="&VARB" _&VARB2="&VARB" _CLASS_YR2="'CLASS_YR'N";
RUN;

%MEND PLOTS;

%PLOTS(VARB=SR, VARB2=SR2, DAT=&PSR);    %PLOTS(VARB=SF, VARB2=SF2, DAT=&PSF);
%PLOTS(VARB=TOS, VARB2=TOS2, DAT=&PTOS); %PLOTS(VARB=DL, VARB2=DL2, DAT=&PDL);
%PLOTS(VARB=TMW, VARB2=TMW2, DAT=&PTMW); %PLOTS(VARB=MF, VARB2=MF2, DAT=&PMF);
%PLOTS(VARB=TML, VARB2=TML2, DAT=&PTML); %PLOTS(VARB=TF, VARB2=TF2, DAT=&PTF);
%PLOTS(VARB=FA, VARB2=FA2, DAT=&PFA);    %PLOTS(VARB=VF, VARB2=VF2, DAT=&PVF);
```

Figure 8. ODS Graphics code turned into a macro

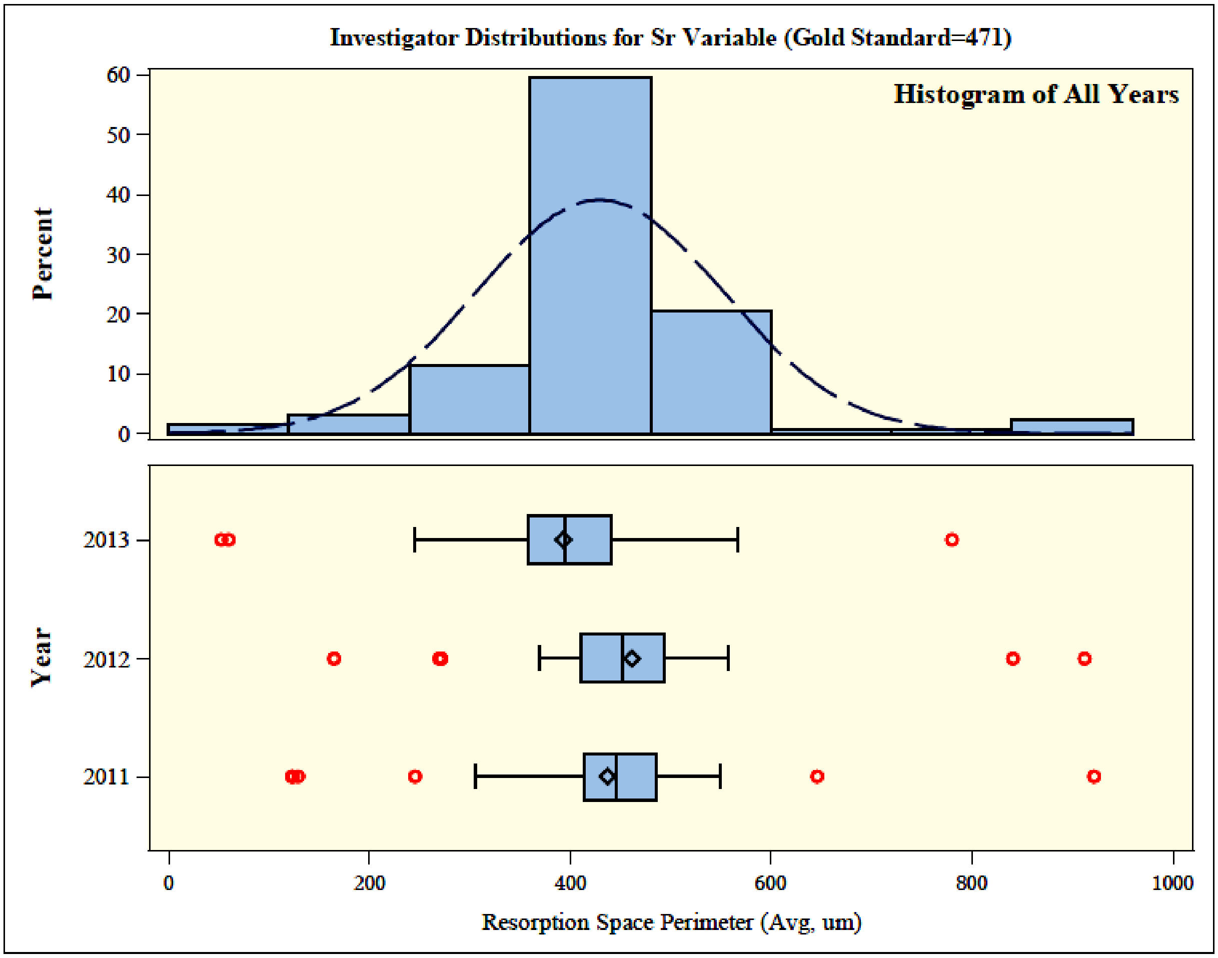

Figure 9. ODS Graphics macro results in PDF format

CONCLUSION

Over the course of this paper, it has been shown that macros have the ability to reduce the amount of coding required when the same formatting and options are applied to numerous variables or when copy/paste is not efficient and could lead to coding errors. With the help of arrays, temporary and permanent variables are able to assist in identifying and labeling investigator values that lie outside the acceptable limits for each variable. In addition, it has been shown that reporting procedures such as ODS PDF, PROC REPORT, and SAS® ODS Graphics Designer are handy tools useful in creating reports which flag questionable investigator values, graphically display data as a whole and by groups, as well as bookmark output in a PDF. Each one of the topics covered in this paper has many additional options that could not be explored in depth here, but there are numerous resources available to help users tailor their code to fit their needs.

FUTURE WORK

The data collection, listed in the methods section, required each investigator to take multiple measurements per variable and create averages. There were no clear-cut guidelines as to how many measurements should be taken to obtain averages for each variable, leaving it up to each investigator to set their own rules. It is also possible that in collecting measurements for the averages, some individual investigators may have had several measurements that spanned a wide range while others had several measurements that were more consistent. When averaged, both approaches could produce similar results making it harder to distinguish between investigators with good techniques and those that need additional training. In a future paper, it may be possible to explore the option of taking every investigators raw data and identify extreme measurements before taking averages and include this information into the final report. This approach will provide Principal Investigators or Quality Control Management the ability to spot investigators and technicians who would have otherwise slipped under the radar.

REFERENCES

Boberg, Wendy. 2008. "Proc Report in Color… What's Your Style." *Proceedings of the 2008 SAS Global Forum*. San Antonio, Texas. Warren E. Stinson. Available at http://www2.sas.com/proceedings/forum2008/224-2008.pdf

Specify the color and the font of a cell based on the cell's value. (2005, July 11). Retrieved from http://support.sas.com/kb/23/353.html

Heath, Dan. 2009. "Secrets of SG Procedures." *Proceedings of the 2009 SAS Global Forum*. Washington, DC. Lori Griffin. Available at http://support.sas.com/resources/papers/proceedings09/324-2009.pdf

Sridharma, Selvaratnam. 2010. "Introduction to Statistical Graphics Procedures." *Proceedings of the 2010 Northeast SAS Users Group*. Baltimore, Maryland. Lisa Eckler. Available at http://www.nesug.org/Proceedings/nesug10/gr/gr07.pdf

Hong, Barry. (2011). *ODS Graphics Designer (Creating Templates for Batchable Graphs)* [PowerPoint Slides]. Retrieved from http://www.sas.com/offices/NA/canada/downloads/presentations/GHSUG-Fall-11/Hong-ODS.pdf

Lund, Pete. (2006). "PDF Can be Pretty Darn Fancy – Tips and Tricks for the ODS PDF Destination." *Proceedings of the 2006 SAS Users Group International*. San Francisco, California. Kimberly J. LeBouton. Available at http://www2.sas.com/proceedings/sugi31/092-31.pdf

Girgis, H., Lilly, E., and Company. (2006). "Custom Template for Reporting Interim Analyses Using Proc Report and Output Delivery System (ODS)." *Proceedings of the 2006 SAS Users Group International*. San Francisco, California. Kimberly J. LeBouton. Available at http://www2.sas.com/proceedings/sugi31/052-31.pdf

"Report Procedure." SAS Institute Inc. 2013. Available at http://support.sas.com/documentation/cdl/en/proc/65145/HTML/default/viewer.htm

"Proc Report Statement". SAS Institute Inc. 2013. Available at http://support.sas.com/documentation/cdl/en/proc/61895/HTML/default/viewer.htm

"Using Style Elements in the REPORT and TABULATE Procedures." SAS Institute Inc. 2008. Available at http://support.sas.com/resources/papers/stylesinprocs.pdf

Delaney, Kevin P. (2003). "ODS PDF: It's Not Just for Printing Any More." *Proceedings of the 2003 SAS Users Group International.* Seattle, Washington. Art Carpenter. Available at http://www2.sas.com/proceedings/sugi28/146-28.pdf

Sadof, Michael G. 1997. "Macros from Beginning to Mend A Simple and Practical Approach to the SAS® Macro Facility." *Proceedings of the 1997 Northeast SAS Users Group*. Baltimore, Maryland. Gerry Pauline and Helen-Jean Talbott. Available at http://www.nesug.org/Proceedings/nesug97/begtut/sadof.pdf

ACKNOWLEDGMENTS

Thanks to Dr. Scott Hazelwood for providing me with the histology course data for this paper and thanks to Rebecca Ottesen, Dr. Stephanie Sullivan, and Demetrice Baskerville for help with resources, proofing, and comments.

CONTACT INFORMATION

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

Name: Jamelle Simmons
Enterprise: Virginia Polytechnic Institute and State University
City, State ZIP: Blacksburg, VA 24060
E-mail: jmsimm13@vt.edu