

Tips and Tricks for Introductory Workshops in SAS® for Health Professionals

Jason Brinkley, Ph.D. American Institutes for Research

ABSTRACT

It can sometimes be the case that general health professionals need some basic SAS training in order to effectively create simple reports and manipulate incoming data. The presenter will share his experiences in leading SAS Workshops Series in a university setting across the course of several years. Heading a team of university faculty members, the presenter has designed, implemented, and refined short term SAS overview training for general health professionals. While multiple topics have been discussed in these workshops, some have fared better with a general health professional audience than others. Topics will include tips on introducing code based work to individuals with no previous experience, workshop format, good practices on instruction and delivery, and introducing SAS macros in an example based manner.

INTRODUCTION

SAS Software continues to be among the most widely utilized and highest valued software for importing, managing, and analyzing data from many different structures. As the rest of the software world has turned to more graphical based output; SAS and other more recent software innovations to the world of data (R, Python, SQL, etc.) continue the tradition of coding for data management and analytics. This need to learn coding can be a severe gap in training of individuals to work with data in SAS. SAS professionals tend to be a mixed group, with many individuals having some background in an area that allows for easier training to create code based solutions.

In 2009, when the main author was affiliated with an academic biostatistics unit; he was tasked with developing a series of workshops to introduce SAS coding for a general academic health audience. The background of participants was not in mathematics, computer science, statistics or any code-friendly science; instead attendees were health professionals with representation mostly from the attached medical school and department of public health. There was mixed experience in working with statistical software and very few with existing SAS coding experience. Several of the masters of public health students (MPH) had received some SAS training in their Epidemiology coursework, but the mix was somewhat uneven which led to the design of the series being discussed here.

The SAS Code Workshop series ran yearly from 2009 to 2014 with a major overhaul in 2013. The purpose here is to discuss both the early and overhauled version of the training, the catalyst for the overhaul, and lessons learned. Participant feedback from the 2014 series illustrated that the overhaul was a success and the redesigned workshop was a lot more effective.

EARLY ITERATIONS (2009-2012)

The original workshop was designed as a four-part series; each workshop was approximately 2 hours. Listed below is a general breakdown of the four parts.

Session 1 (Data Management): Topics were –

- A brief history of SAS Software
- Discussion of the layout of SAS (Output, Log, and Editor windows)
- Introduction of the Data step for inputting data
- Importing data, stacking, and merging was discussed
- The Output Delivery System (ODS) was discussed
- Procs discussed: Print, Sort, Means, Freq

Session 2 (Analyzing Data): Topics were –

- Importing data
- Univariate and bivariate statistical analysis
- Procs discussed: univariate, boxplot, means, ttest, glm, npar1way, gplot, gchart, freq

Session 3 (Linear and Logistic Regression): Topics were –

- Focus was on regression
- Procs discussed: gplot, reg, corr, glm, logistic
- Statistical discussions regarding model fitting and assessment

Session 4 (Repeated Measures and Mixed Models): Topics were –

- MANOVA, Repeated Measures, and Linear Mixed Models
- Procs discussed: Proc Mixed

An examination of the early set of topics reveals that they were highly focused on statistical analyses. Discussion with stakeholders (senior faculty and experienced university researchers) early on suggested that the goal for the participants was to learn how to analyze data in SAS. Therefore biostatistics faculty took the analytic perspective and focused efforts toward producing analytic output. The first workshop was designed to be hands on and force participants to type and manipulate data in SAS. Later workshops were posted with pre-written SAS code for individuals to follow along. The primary reference point for data and examples was Cody (2005).

RESTRUCTURING

The need for restructuring came predominately from participant feedback; after two years of implementation it became clear that the early discussions with stakeholders did not match up with the actual needs of participants. The emphasis on analysis was too much; participants with limited statistical and coding training just could not internalize all of the information being given to them. Feedback suggested that the workshops were too long and there was a need to have more opportunities for practice.

REDESIGNED WORKSHOP (2013-2014)

The redesigned workshops focused less on analyses and more on data management, formatting, and non-analytic aspects of working with data. The workshops themselves were changed from 2 hours to 1.5 hours and each workshop was designed to allow time for participants to implement all methods and have additional practice. The analytic content was not taken far beyond what would be found in an introductory statistics course and more time was spent on discussing the nuts and bolts of working with data in SAS. The number of sessions was expanded from 4 to 6 so the overall training time went from 8 hours to 9 hours total. Listed below is a general breakdown of the workshops six parts.

Session 1 (Data Importing/Creation): Topics were –

- A brief history of SAS Software
- Discussion of the layout of SAS (Output, Log, and Editor windows)
- Introduction of the Data step for inputting data
- Importing data, stacking, and merging was discussed
- Creating new variables
- Output, Keep and Drop statements
- Proc means was briefly discussed

Session 2 (Data Management): Topics were –

- Proc Import, Print, and Contents were discussed to evaluate data structure
- Data step manipulations were discussed to fix import issues
- Proc Freq and Univariate were discussed from both analytic and data quality perspectives
- Proc Format, Labels, and Titles were discussed
- Proc Export and Libname statements were discussed
- Proc Report and ODS were discussed

Session 3 (Messy Data): Topics were –

- Proc Transpose was discussed
- Proc Contents was rediscussed
- Data formats and conversions were discussed
- Significant time was spent on date/time variables and various quality issues

Session 4 (Creating Visuals): Topics were –

- Creating quality visuals was the focus
- Proc Gchar, Gplot and Boxplot were discussed
- Proc Sgplot, Sgscatter, Sgchart, and Sgpanel were discussed
- The symbol statement and various options were discussed

Session 5 (Univariate and Bivariate Analysis): Topics were –

- Univariate and bivariate statistical analysis was focus
- Proc Freq, Gchart, Gplot, Means, Corr, ttest, Reg, and Anova were discussed
- Simple examples were used to illustrate procedure use

Session 6 (Macros): Topics were –

- The SAS Macro language was the focal point
- Participants utilized an existing macro
- Participants created a fresh macro to automate repetitive code

In comparing the previous series to the redesigned series, the reduction in statistics related content is apparent. Much of the series focus shifted away from analysis and more on working with data. Users did get a session on visualization and a session on basic statistical analysis. The addition of formats, time/date data, and macros came specifically from participant feedback forms. The revitalized series received many compliments from participants and two users from the 2009 series returned for the 2014 series to find that the restructured series was much better for them. The macro session was especially well received with illustrations from a commonly used CDC macro that uses growth charts in children to assign BMI percentiles.

LESSONS LEARNED

After five years of running the series there were many lessons learned. The original series was designed by biostatisticians as a mechanism for training junior analysts. This came from discussions with university stakeholders; however it was clear later on that these individuals were providing more self-serving advice. They were less concerned with preparing the next generation of health researchers and more concerned with having students/junior colleagues who were capable of helping senior faculty perform analyses that they themselves were incapable of doing.

One challenge was in finding data to illustrate the many concepts discussed in the revitalized series. For individuals moving forward with the creation of similar workshop series the best lesson learned from here is that SAS is a diverse software with many different aspects for working with data. While analyses are important, it may be more useful for new users to understand the challenges of working with data and manipulating that data into the proper format for analyses. Training that is designed to deliver information in shorter bursts (60 minutes or less) along with opportunities for practice also seem to be better received.

REFERENCES

Cody, RP. Smith JK. 2005. Applied Statistics and the SAS Programming Language 5th Edition. Cary, NC: SAS Press.

“A SAS Program for the 2000 CDC Growth Charts (ages 0 to <20 years)”. Center for Disease Control. 8/1/15. Available at <http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm>.

ACKNOWLEDGMENTS

Special thanks to all of the individuals who have contributed to the workshop series over the years. Especially to the instructors Drs. Xiangming Fang, Andrada Ivanescu, Kevin O’Brien, and Qiang Wu. Additional thanks goes to Drs. Suzanne Lea and Jimmy Efrid on workshop content feedback.

CONTACT INFORMATION

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

Name: Jason Brinkley
Enterprise: American Institutes for Research
Address: 100 Europa Drive Suite 315
City, State ZIP: Chapel Hill, NC 27517
E-mail: jbrinkley@air.org

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.