FROM THE PRESIDENT

Hopefully, the unusually-hot summer in San Diego is over. We are preparing for the last SANDS meeting of the year. We have invited two speakers to present at our December meeting. Kirk Paul Lafler, from Software Intelligence Corporation, will present his popular SAS programming tips and techniques in the afternoon. He will then present a featured presentation on Massive Open Online Content (MOOC) for SAS learners during our evening meeting. Dr. Joey Lin, from San Diego State University, will present the second talk on data analysis in multivariate regression. Both of them have been long-time SANDS supporters, and have presented at SANDS multiple times. I look forward to their presentations!

The SANDS Executive Committee is planning a half-day SANDS-sponsored training to our members. The location will be our usual meeting place at Pfizer. The instructor of the training is Kirk Lafler. We plan to have the training either at the end of January or in early March. Kirk is a well-known SAS book author and trainer. His popular SAS courses and seminars cover a wide range of topics, such as ODS, SQL, Enterprise Guide, Macros, hash programming, etc. Please let us know which topic you would like to learn and if you plan to attend the training.

SANDS is run by volunteers. We always need more people to help support this local SAS user community. If your company can sponsor our meetings or would like to advertise in our newsletter, please contact Vivian Huang or Tyler Smith, or any of our EC members.

I would like to take this chance to thank our sponsors for this meeting: WUSS EC and Pfizer. SANDS meetings are held with no cost to our members because of our generous sponsors. I would also like to thank all of the SANDS EC members for their wonderful work throughout the year.

See you at the meeting!

- Wei Cheng
# MEETING AGENDA - DECEMBER 9, 2015

**Location:** Building CB2, Room 1110 - RSVP Now!

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:30 - 5:00</td>
<td><strong>Workshop:</strong> SAS® Programming Tips and Techniques for Programmers - Kirk Paul Lafler</td>
</tr>
<tr>
<td>5:15 - 6:00</td>
<td>Dinner / Networking</td>
</tr>
<tr>
<td>6:00 - 6:15</td>
<td>Welcome / SANDS Business</td>
</tr>
<tr>
<td>6:15 - 7:15</td>
<td><strong>Featured Presentation:</strong> A Review of &quot;Free&quot; Massive Open Online Content (MOOC) for SAS® Learners - Kirk Paul Lafler</td>
</tr>
<tr>
<td>7:15 - 7:45</td>
<td><strong>Short Talk:</strong> Data Analyses in Multivariate Regression – Dr. Joey Lin</td>
</tr>
<tr>
<td>7:45 - 8:00</td>
<td>Stump the Programmer and Coders’ Corner</td>
</tr>
<tr>
<td>8:00 - 8:15</td>
<td>Door Prizes and Giveaways</td>
</tr>
</tbody>
</table>

---

**Synchr ony Solutions**

The mission of Synchrony Solutions is simple:
To provide your company with timely and effective search solutions that exceed expectations. Our business is about relationships. We build our network, and our reputation, one person at a time, every time. When you have a position that needs to be filled, or you are ready to find your next opportunity, let us do the work for you!

molly@synchrony-solutions.com
Phone: 1-866-272-1228

www.synchrony-solutions.com

---

**Need a Math Tutor?**

**Call/Text Archie at 858-633-8486 or info@askawaytutoring.com**

- Algebra
- Pre-Calculus
- Geometry
- Calculus
- Trigonometry
- SAS
- Discrete Math
- Differential Equations
- Number Theory
- Math Sections of: SAT, GRE, GMAT

Ask A Way Tutoring
**Workshop: SAS® Programming Tips and Techniques for Programmers (Kirk Paul Lafler)**

*Abstract:*

Explore a collection of proven programming tips and techniques found in the base-SAS® Software. Attendees learn "key" system options to improve productivity; performance-tuning techniques for SAS users to apply in their applications; how to remove duplicate observations (or rows) from data sets (or tables) based on the row’s values and/or keys; SQL procedure and table options to influence the optimizer to use specific join algorithms; how to access variable attributes in a table with a single statement; use table integrity constraints to prevent data issues; and use user-defined tools with SAS metadata Dictionary tables and SASHELP views.

**Featured Presentation: A Review of "Free" Massive Open Online Content (MOOC) for SAS® Learners (Kirk Paul Lafler)**

*Abstract:*

Leading online providers are now offering SAS® users with “free” access to content for learning how to use and program in SAS. This content is available to anyone in the form of massive open online content (or courses) (MOOC). Not only is all the content offered for “free”, but it is designed with the distance learner in mind, empowering users to learn using a flexible and self-directed approach. As noted on [wikipedia.org](http://wikipedia.org), “a MOOC is an online course or content aimed at unlimited participation and made available in an open access forum using the web.” This presentation illustrates how anyone can access a wealth of learning technologies including comprehensive student notes, instructor lesson plans, hands-on exercises, PowerPoints, audio, webinars, and videos.

*Bio: Kirk Paul Lafler* has been using SAS since 1979 and is consultant and founder of Software Intelligence Corporation. He is a SAS Certified Professional, provider of IT consulting services, trainer to SAS users around the world, mentor, and [sasCommunity.org](http://sasCommunity.org) emeritus Advisory Board member. As the author of six books including Google® Search Complete! (Odyssey Press. 2014) and PROC SQL: Beyond the Basics Using SAS, Second Edition (SAS Press. 2013); Kirk has written more than five hundred papers and articles; been an Invited speaker and trainer at five hundred-plus SAS International, regional, special-interest, local, and in-house user group conferences and meetings; and is the recipient of 23 “Best” contributed paper, hands-on workshop (HOW), and poster awards.

**Short Talk: Data Analyses in Multivariate Regression (Joey Lin)**

*Abstract:*

Regression analysis is one of the most used statistical methodologies. It can be used to describe or predict causal relationships between a response variable and a set of predictor variables. An extension of univariate (one response variable) regression analysis is multivariate regression analysis. Multivariate regression extends regression analysis into more than one response variable. In a multiple linear regression model, linear association between the response variable and the predictors are assumed. In addition, the random errors are assumed to be uncorrelated and normally distributed with a constant variance. In conducting a multivariate regression analysis, the assumptions are similar to multiple linear regression models but extended to multivariate aspect. In this paper, we introduce SAS® codes for data analyses related to multivariate regression using PROC REG. The mtest statement in PROC REG is a key component for multivariate regression analysis. An example is introduced to demonstrate the usage of the mtest statement.

*Bio: Dr. Joey Lin* is the Associate Chair of Statistics in the Department of Mathematics & Statistics and also the co-director of the Statistical Consulting Center at San Diego State University. He has been using SAS for more than 25 years. He teaches SAS programming and data management course at SDSU regularly and has inspired students to become SAS programmers. He also uses SAS as his primary computational tool for his research.
TIPS FOR USING A DATA WAREHOUSE

In our last episode, we took a turn from our usual path and discussed some techniques I use for data analytics related to auditing business transactions. This time, back on our usual path, we will explore a simplified way to subset a SAS data set using the SELECT statement. You may remember, we did something similar in the April/May 2015 edition using IF/THEN/ELSE statements.

Let's consider a situation where the data set in our data warehouse contains rows for several (finite and known) departments within the company. For an analysis we want to perform, we need to create a separate SAS data set for each unique department. So, we need to read the SAS data set into a SAS Data Step and subset the input data set based on the value of the department (DEPT) variable. We will accomplish this using the SELECT statement, the syntax of which is as follows:

```
SELECT <(select-expression)>;
WHEN-1 (when-expression-1 <..., when-expression-n>) statement;
<... WHEN-n (when-expression-1 <..., when-expression-n>) statement;>
<OTHERWISE statement;>
END;
```

The select-expression specifies any SAS expression that evaluates to a single value. So, we will set our select-expression to use the DEPT variable. Consider the following very simple SAS Data Step where the existing data set in our data warehouse is In.BigData:

```
data out.dept01 out.dept02 out.dept03 out.junk;
set in.BigData;
SELECT (DEPT);
WHEN ("01") OUTPUT out.dept01;
WHEN ("02") OUTPUT out.dept02;
WHEN ("03") OUTPUT out.dept03;
OTHERWISE OUTPUT out.junk;
end;
run;
```

Let's get some background on the SELECT statement from the SAS Institute (Sas(r) 9.2 language, 2011):

The SELECT statement begins a SELECT group. SELECT groups contain WHEN statements that identify SAS statements that are executed when a particular condition is true. Use at least one WHEN statement in a SELECT group. An optional OTHERWISE statement specifies a statement to be executed if no WHEN condition is met. An END statement ends a SELECT group.

If the select-expression is present, SAS evaluates the select-expression and when-expression. SAS compares the two for equality and returns a value of true or false. If the comparison is true, statement is executed. If the comparison is false, execution proceeds either to the next when-expression in the current WHEN statement, or to the next WHEN statement if no more expressions are present. If no WHEN statements remain, execution proceeds to the OTHERWISE statement, if one is present.

So, what's going on here? Our input SAS data set has a department variable that should contain one of three values: 01, 02, or 03. The SELECT statement tells the Data Step to pick a row and look in the DEPT variable and if it sees "01" then write that row out to output data set out.dept01; but if not, then go to the next WHEN statement and look in the DEPT variable and if it sees "02" then write that row out to output data set out.dept02, and so forth. Because we know our data, there should not be any values in DEPT other than 01, 02, and 03. But, just in case because we are not sloppy programmers, we use the OTHERWISE statement to tell the Data Step what to do with a bad row: in this case, write it to a junk data set so we can view the bad data later.

As you know, we can accomplish the same thing with a series of IF/THEN/ELSE statements. So, which is better? From the research and experimenting I have done, both methods seem to perform about the same. But, the SELECT statement with the WHEN statement is very simple and clean programming.

Thanks for reading.

Curtis Smith
DoD Project Manager
ca.smith86@att.net

KIRK’S KORNER: UNDERSTANDING DATA REMERGING IN PROC SQL

SAS® users are accustomed to seeing many informative notes displayed on the SAS Log. From the number of observations (or rows) read from a SAS dataset (or table) to the number of observations written to a SAS dataset, the amount of CPU, I/O and memory resources used by a SAS program to the many other informative messages, users are accustomed to seeing details about their running programs. While these types of messages provide useful information about the processing performed; not all messages are as easy to detect, understand, and/or solve.

Apart from the many informational messages produced from a successfully running program, users often act surprised when the Log contains messages, warnings and/or errors about problems found in their programs. In fact, the first sign exhibited by many is one of disbelief followed by a, “this can’t be actually happening” moment, as if their code were a masterpiece by Michelangelo. But, the statistics along with the realities of programming and debugging say otherwise. As a result, users often become perplexed, confused and frustrated, producing a less than productive programming experience.

Because warnings and errors can, it seems, appear out of nowhere, the savvy SAS user knows that the most important rule to remember when debugging is to check the SAS Log immediately after each run. While this approach often helps to detect and understand the causes of warnings and/or errors, it does little to diminish the “punch to the gut” feeling that a SAS user gets when seeing those pesky warning and error messages appear on the Log in the first place. The one thing that all SAS users learn about the debugging process is that it can be a daunting, and sometimes humbling, experience.

This tip presents details about the not-so-obvious note, “The query requires remerging summary statistics back with the original data,” that is produced in PROC SQL queries. While not a warning or error, this note does serve to alert users about the possibility of a query expending additional computing resources to complete a particular task. In my explanation of the causes and solutions to this note, the Movies table is used, shown below, consisting of six columns: title, length, category, year, studio, and rating. Title, category, studio, and rating are defined as character while length and year are defined as numeric.

MOVIES Table

<table>
<thead>
<tr>
<th>Title</th>
<th>Length</th>
<th>Category</th>
<th>Year</th>
<th>Studio</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brave Heart</td>
<td>177</td>
<td>Action Adventure</td>
<td>1995</td>
<td>Paramount Pictures</td>
<td>R</td>
</tr>
<tr>
<td>Casablanca</td>
<td>103</td>
<td>Drama</td>
<td>1942</td>
<td>MGM / UA</td>
<td>PG</td>
</tr>
<tr>
<td>Christmas Vacation</td>
<td>97</td>
<td>Comedy</td>
<td>1989</td>
<td>Warner Brothers</td>
<td>PG-13</td>
</tr>
<tr>
<td>Coming to America</td>
<td>116</td>
<td>Comedy</td>
<td>1988</td>
<td>Paramount Pictures</td>
<td>R</td>
</tr>
<tr>
<td>Dracula</td>
<td>130</td>
<td>Horror</td>
<td>1993</td>
<td>Columbia TriStar</td>
<td>R</td>
</tr>
<tr>
<td>Dresses to Kill</td>
<td>105</td>
<td>Drama Mysteries</td>
<td>1980</td>
<td>Ritz Pictures</td>
<td></td>
</tr>
<tr>
<td>Forrest Gump</td>
<td>142</td>
<td>Drama</td>
<td>1994</td>
<td>Paramount Pictures</td>
<td>PG-13</td>
</tr>
<tr>
<td>Ghost</td>
<td>127</td>
<td>Drama Romance</td>
<td>1950</td>
<td>Paramount Pictures</td>
<td>PG-13</td>
</tr>
<tr>
<td>Jaws</td>
<td>125</td>
<td>Action</td>
<td>1975</td>
<td>Universal Studios</td>
<td>PG</td>
</tr>
<tr>
<td>Jurassic Park</td>
<td>127</td>
<td>Action</td>
<td>1993</td>
<td>Columbia TriStar</td>
<td>R</td>
</tr>
<tr>
<td>Lethal Weapon</td>
<td>110</td>
<td>Action Copp &amp; Robber</td>
<td>1987</td>
<td>Warner Brothers</td>
<td>R</td>
</tr>
<tr>
<td>Michael</td>
<td>105</td>
<td>Drama</td>
<td>1997</td>
<td>Warner Brothers</td>
<td>PG-13</td>
</tr>
<tr>
<td>National Lampoon’s Vacation</td>
<td>98</td>
<td>Comedy</td>
<td>1983</td>
<td>Warner Brothers</td>
<td>PG-13</td>
</tr>
<tr>
<td>Poltergeist</td>
<td>115</td>
<td>Horror</td>
<td>1982</td>
<td>MGM / UA</td>
<td>PG</td>
</tr>
<tr>
<td>Rocky</td>
<td>120</td>
<td>Action Adventure</td>
<td>1976</td>
<td>MGM / UA</td>
<td>PG</td>
</tr>
<tr>
<td>Scarface</td>
<td>170</td>
<td>Action Copp &amp; Robber</td>
<td>1983</td>
<td>Universal Studios</td>
<td></td>
</tr>
<tr>
<td>Silence of the Lambs</td>
<td>118</td>
<td>Drama Suspense</td>
<td>1991</td>
<td>Orion</td>
<td>R</td>
</tr>
<tr>
<td>Star Wars</td>
<td>124</td>
<td>Action Sci-Fi</td>
<td>1977</td>
<td>Lucas Films Ltd</td>
<td>PG</td>
</tr>
<tr>
<td>The Hunt for Red October</td>
<td>135</td>
<td>Action Adventure</td>
<td>1989</td>
<td>Paramount Pictures</td>
<td>PG</td>
</tr>
<tr>
<td>The Wizard of Oz</td>
<td>101</td>
<td>Adventure</td>
<td>1938</td>
<td>MGM / UA</td>
<td>G</td>
</tr>
<tr>
<td>Titanic</td>
<td>194</td>
<td>Drama Romance</td>
<td>1997</td>
<td>Paramount Pictures</td>
<td>PG-13</td>
</tr>
</tbody>
</table>
Understanding Data Remerging

Data remerging involves the process of making two passes through the dataset (table). As might be imagined, additional processing time (or resources), such as CPU, are required when data remerging occurs. Remerging is performed whenever one of the following conditions occurs:

1. The SELECT statement specifies one or more columns that don’t appear in a GROUP BY-clause;
2. The SELECT statement specifies one or more columns that don’t appear in a HAVING-clause;
3. The values returned by a summary function are specified in a calculation.

During the remerging process, the following occurs while making two passes through the dataset (table):

1. Summary functions are computed with values returned;
2. The conditions specified in the GROUP BY-clause are performed;
3. Rows and columns that are displayed in the output are retrieved;
4. The results from the summary functions are used to compute arithmetic expressions in which the summary function is involved with.

Let’s examine how the data remerging process works for the following query.

SQL Code

```
proc sql;
select
    title,
    rating,
    MIN(length) as Min_Length,
    length
from movies
where rating IN ("PG", "PG-13")
group by rating;
quit;
```

Analysis

PROC SQL computes and returns the value of the MIN function in the first pass of the dataset (table), and groups the results according to the GROUP BY-clause. In the second pass of the dataset (table), PROC SQL retrieves the rows and columns that are displayed in the output.

Log

```
proc sql;
select
    title,
    rating,
    MIN(length) as Min_Length,
    length
from movies
where rating IN ("PG", "PG-13")
group by rating;
quit;
```

NOTE: The query requires remerging summary statistics back with the original data.

Avoiding Data Remerging

At times it is difficult, if not unavoidable, to prevent PROC SQL from performing data remerging. But, one way to prevent PROC SQL from performing the data remerging process, and to avoid the additional costs associated with making two passes of the data, is to rewrite the query so the SELECT statement specifies only the column(s) specified in the GROUP BY-clause. The following query illustrates a SELECT statement that limits the columns specified to those identified in the GROUP BY-clause.
**SQL Code**

```
proc sql;
select
  rating,
  MIN(length) as Min_Length
from movies
where rating IN ("PG", "PG-13")
group by rating;
quit;
```

Another approach to avoid the data remerge process could be to remove the summary function from the SELECT statement altogether and to replace the GROUP BY-clause with an ORDER BY-clause, as follows.

**SQL Code**

```
proc sql;
select
  title,
  rating,
  length
from movies
where rating IN ("PG", "PG-13")
order by rating;
quit;
```


Kirk Paul Lafler  
Senior SAS Consultant, Application Developer, Data Scientist, Trainer and Author  
Software Intelligence Corporation  
E-mail: KirkLafler@cs.com  
LinkedIn: [http://www.linkedin.com/in/KirkPaulLafler](http://www.linkedin.com/in/KirkPaulLafler)  
Twitter: @sasNerd

---

**SANDS Membership**

Membership in the San Diego SAS Users Group, Inc. (SANDS) provides an opportunity for SAS users to meet and keep updated on SAS issues. Seminars and presentations allow SAS users to gain knowledge with little or no cost. It is also an opportunity for consultants and prospective employees to meet possible employers, and companies to show their SAS-related products and host a meeting. We welcome those interested in giving presentations and writing articles for the newsletter. Presently, there is no charge to attend our meetings or to receive our newsletter; small fees may be asked for some special functions or workshops.

The SANDS membership list is not given out to others and is used only to disseminate SANDS related business, such as newsletters or special announcements. Those wishing to sell their products or services have the opportunity to advertise in the newsletter at reasonable rates.

To become a member, please email the following to [Cathy Liu](mailto:CathyLiu): Name, Title, Organization, Telephone, E-mail, Fax, Address, Work or Home indicator, Computer Platform, SAS version, and SAS related areas of interest.
Stump the Programmer #74
by Art Carpenter

Picture Formats and Decimal Values

Picture formats display numeric values using a template that you supply. In the picture format TWODEC, shown here, the template specifies two decimal places. However, the numbers applied to the format range from .001 to 1234. What value will be displayed for each of the numbers in the variable X?

```
proc format;
picture twodec
  other = 000.99;
run;
data x;
do x = .001,.012,.123,1.234, 12.34, 123.4, 1234;
  output;
end;
format x twodec.;
run;
proc print data=x;
run;
```

Data Explorations

Data Management & Analyses for Biological, Ecological & Health Sciences.

**SAS® Programming Services:** Extensive experience with SAS macros, SAS/STAT®, SAS/GRAPH®, SAS/AF®, SAS/FSP®, & SAS Component Language.

Statistical and data analysis support for clinical trial studies.
Data entry and reporting systems. Application Development.
Database conversion services. SAS Training services.
Experimental Design and Data Management.

Data Explorations is a SAS Alliance Affiliate Member™ with SAS Advanced Certified Professionals.

Art Carpenter & Richard Smith - Partners
(907) 865-9167 Art
(760) 613-5373 Richard

Email: Art: Art@DataXplorations.com
      Richard: RSmith@DataXplorations.com

Web: http://www.DataXplorations.com

SANDS Sponsorship

We would like to thank Pfizer for having once again generously provided our meeting room. This month, WUSS EC and Pfizer are sponsoring our meeting dinner and activities. It is our sponsors who make it possible for us to hold our meetings free of charge so, at the event, please do let sponsor representatives know how much you appreciate their support. If your company would like to sponsor SANDS activities, please contact the SANDS sponsorship coordinators, Tyler Smith and Vivian Huang.
Map to Pfizer La Jolla Campus

Directions to the Campus Buildings

Please refer all Contractors and Visitors that will be visiting
CB1-CB6, CB10 or The Pfizer Incubator to the CB2 Visitors Check-in Center.
A photo ID is required for visitor check in.

Groups of 15 or more will check-in at the lobbies of the buildings visited.
Visitors should arrive 15 to 20 minutes early to allow enough time for check-in.

CB2 Visitors Center
10770 Science Center Drive, San Diego, CA 92121
From San Diego International Airport.
Follow signs to I-5 North.
Proceed on I-5 North for approx. 10 miles.
Exit I-5 at Genesee Ave. and make a left at the stop light.
Make a right at the second stop light which will be Science Center Drive.
CB2 Visitors Center is located at the end of the cul-de-sac.
Visitor Parking is to the left. Please follow Visitors Signs.
ELECTED OFFICERS

President • Wei Cheng
Vice-President • Archie Medrano
Secretary/Treasurer • Song Lin

Advertising Coordinators • Tyler Smith & Vivian Huang

<table>
<thead>
<tr>
<th>Type</th>
<th>Dimensions (w x h)</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Card</td>
<td>3.5” x 2”</td>
<td>$25</td>
</tr>
<tr>
<td>Quarter Page</td>
<td>3.5” x 4.5”</td>
<td>$50</td>
</tr>
<tr>
<td>Half Page</td>
<td>7” x 4.5”</td>
<td>$125</td>
</tr>
<tr>
<td>Full Page</td>
<td>7” x 9”</td>
<td>$225</td>
</tr>
</tbody>
</table>

Note: Prices may change without notice. Additional fees may be charged if modifications are needed. For verification of current prices and/or to send ads in PDF format, please contact Tyler Smith and/or Vivian Huang.

Catering Coordinator • Vivian Huang
Facilities Coordinators • Ren-Yu Tzeng & Yiyun Tang
Job Bank Coordinator • Jia Hu
Membership Coordinators • Cathy Liu & Chieko Seto
SAS Institute Liaison • Andrea Littleton
Speaker/Program Coordinator • Kirk Paul Lafler
Sponsorship Coordinators • Vivian Huang & Tyler Smith
Members At-Large • Scott Leslie, Charlie Shipp

Newsletter Editor/Webmaster • Archie Medrano

The SANDS newsletter is published quarterly and distributed to members by e-mail. The SANDS EC team appreciates the efforts of all contributors, without whom this newsletter would not be possible. We are always looking for volunteers to contribute original content to the quarterly newsletter. If you have ideas or suggestions for the newsletter, please share them with us. We would love to hear them! Newsletter text not otherwise indicated was written or edited by the Editor. Please send any questions or comments to the Editor.

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.

Websites: sandsug.org  sascommunity.org  LinkedIn  Facebook  Twitter