ABSTRACT
This paper uses SAS® software to create a graphical representation of a 95% confidence interval for the mean. A reference scale is printed at the top then confidence intervals of appropriate length are printed on separate lines with optional labels attached.

INTRODUCTION
Statistical information is absorbed more easily if displayed in a graph. The technical calculations for confidence intervals are not difficult but listing sets of numbers e.g. (12.56, 26.94) (8.50, 14.75) (26.65, 40.23) to show the relationship or relative size of these findings is not as informative as seeing them printed out. The SAS code in this paper shows 95% confidence interval data in an easy to read format without the use of any special graphics or printers. Only minor changes to the SAS code are needed to accommodate different data sets.

DESCRIPTION AND OPERATION
First, the SAS code will be explained step-by-step for the main program (Appendix A). Then changes that make up the modified SAS code (Appendix B) will be described.

To calculate a 95% confidence interval for the mean, SAS Communications, Vol XIII No. 4 suggests using PROC REG. Specifying only a dependent variable and using output options produces a confidence interval based on the input values of the dependent variable. Since displaying one confidence interval would not be very interesting, the addition of a BY variable allows for multiple confidence intervals per data set.

Normally, PROC REG would be used in a regression situation and the confidence intervals would be calculated for the mean predicted value of the dependent variable. Do not let the use of PROC REG confuse you into thinking about only regression.

STEP 1: Calculation of Confidence Intervals
The SAS data set containing the variables of interest should be sorted according to the grouping or BY variable. Using this data as input to the regression procedure, the SAS system will calculate 95% confidence intervals for the mean values. Using PROC REG write a MODEL statement specifying only a dependent variable. Using the OUTPUT OUT = option, a BY statement and L95M and U95M statistics the SAS System will put the 95% confidence intervals for the mean (dependent variable) into an output data set. This new data set contains 1 observation for every observation in the input data set. When the confidence intervals are actually printed only 1 observation per BY group is needed since all observations within a BY group have the same limits.

Calculation of the interval is:
sample mean ± t/2 * standard error of the mean.

Note: degrees of freedom (n - 1).

STEP 2: Find Range of Intervals
In order to produce a reference scale the lower and upper limits for the confidence interval data must be identified. PROC MEANS run on the output data set from PROC REG will give a new output data set with only 1 observation (do not use the BY statement). Find the maximum and minimum values for the upper bound, lower bound and the BY variable; specify min= and max= as output options.

A data step is used to calculate the range of values in the confidence intervals. The maximum value of the upper limit minus the minimum value of the lower limit is the range. If the range is small for the page size, the program allows for half of the standard scale to be increased to fill the original size page. The value of FACTOR = .5 sets each dash in the reference scale equal to .5. If the range is large the scale (1 printing character = n numeric units) is increased by an integer factor n. Thus the scale increments which correspond to the values of the variable FACTOR can be 0.5, 1, 2, ..., n. In order to have these values available for the actual printing, a DO loop was set up to produce observations with values of the BY variable and the scaling factor (values of FACTOR).

STEP 3: Calculate Printing Configurations
A _NULL_ data step merges the data sets of confidence intervals (by BY variable) from PROC REG with observations produced by the preceding data step containing the scaling factor. Using a FILE PRINT option and PUT statements the reference scale is printed to the listing file with confidence intervals following. The length of a confidence interval is
calculated by taking the integer value of the upper limit minus the lower limit divided by the scaling factor and then subtracting 2. The statement is written as: \[ \text{LENGTH} = \text{ROUND}((\text{UPPER} - \text{LOWER}) / \text{FACTOR}) - 2 \]

The 2 is subtracted because of the vertical bars printing characters which will be the ends of the graphical representation (confidence interval). The ROUND math function rounds a value to the nearest roundoff unit. The length is rounded to an integer value that will be translated to printing characters (vertical bars and horizontal dashes). A word of caution. When you have a small confidence interval and a value of FACTOR greater than one, you could encounter a situation where LENGTH would have a value less than 1 (due to subtracting 2). When the value of LENGTH is negative no dashes will be printed; this an indication that the interval is very small.

**STEP 4: Print Reference Scale and Confidence Intervals**

The reference scale is designated by: \[ X = \text{REPEAT}(-)--,n) \]

Starting the confidence interval at the correct position is accomplished by rounding off the lower limit of the interval to an integer value dividing by the scaling factor and subtracting one from that value. At this position a vertical bar is printed then horizontal dashes are printed. The number of dashes printed corresponds to the length of the interval. All the confidence interval printing is done by the last line of the program which is a PUT statement.

The steps described above are given in Appendix A and will work for any data set that has confidence intervals starting around zero. If the confidence intervals are large or small the addition of the scaling factor takes care of this problem, but what if the intervals are not close to zero. The reference scale would need to have a different starting point. To accomplish this the I loop values can be changed and a variable ORIG created to show the relationship of printing position to actual numeric values. The SAS statement \( \text{PUT @I INDEX @} \) is changed to \( \text{PUT @ORIG INDEX @} \). The other change is in specifying the starting position for the confidence interval, you need to subtract the starting value from the variable START. All these modifications are incorporated in the SAS code shown in Appendix B. The only revisions needed to the code given in appendices for any SAS data set of similar form are to:

1. Pick Appendix A or Appendix B.
2. Change the BY variable name and labels.
3. Change the width of the printout according to paper size (now set for standard size paper). Lines requiring change are designated in Appendix A with # page size in the comments to the right.
   a. Change I values (add or truncate number of values).
   b. Change factor divisor (now set at 80 printing units).
   c. Increase or decrease number of repeats that draw the reference scale.

**REFERENCES**


*SAS and SAS Communications are registered trademarks of SAS Institute, Inc., Cary, NC, USA.*
EXAMPLE 1

CONFIDENCE INTERVALS FOR MEAN VALUE BY MONTH

\[
\begin{array}{cccccccccc}
1 & 5 & 10 & 20 & 30 & 40 & 50 & 60 & 70 & 80 \\
\end{array}
\]

\begin{itemize}
\item \text{MONTH}=1
\item \text{MONTH}=2
\item \text{MONTH}=3
\item \text{MONTH}=4
\item \text{MONTH}=5
\item \text{MONTH}=6
\item \text{MONTH}=7
\item \text{MONTH}=8
\item \text{MONTH}=9
\item \text{MONTH}=10
\item \text{MONTH}=11
\item \text{MONTH}=12
\end{itemize}

EXAMPLE 2

CONFIDENCE INTERVALS FOR MEAN VALUE BY MONTH

\[
\begin{array}{cccccccccc}
20 & 30 & 40 & 50 & 60 & 70 & 80 & 90 & 100 \\
\end{array}
\]

\begin{itemize}
\item \text{MONTH}=1
\item \text{MONTH}=2
\item \text{MONTH}=3
\item \text{MONTH}=4
\item \text{MONTH}=5
\item \text{MONTH}=6
\item \text{MONTH}=7
\item \text{MONTH}=8
\item \text{MONTH}=9
\item \text{MONTH}=10
\item \text{MONTH}=11
\item \text{MONTH}=12
\end{itemize}
EXAMPLE 3

CONFIDENCE INTERVALS FOR MEAN VALUE BY MONTH

<table>
<thead>
<tr>
<th>0.5</th>
<th>2.5</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MONTH=1
MONTH=2
MONTH=3
MONTH=4
MONTH=5
MONTH=6
MONTH=7
MONTH=8
MONTH=9
MONTH=10
MONTH=11
MONTH=12

EXAMPLE 4

CONFIDENCE INTERVALS FOR MEAN VALUE BY MONTH

<table>
<thead>
<tr>
<th>30</th>
<th>60</th>
<th>90</th>
<th>120</th>
<th>150</th>
<th>180</th>
<th>210</th>
<th>240</th>
<th>270</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MONTH=1
MONTH=2
MONTH=3
MONTH=4
MONTH=5
MONTH=6
MONTH=7
MONTH=8
MONTH=9
MONTH=10 *
MONTH=11 *
MONTH=12 *

* very small confidence intervals will not have dashes printed

The author can be contacted at:
University of Central Florida
Department of Statistics
Orlando, Florida 32816-0370
(407) 275-2740
DATA WATER;
INPUT MONTH VALUE @@;
CARDS;
1 10 1 15 1 19 1 20 1 13 1 30 1 40 1 19
2 5 2 10 2 10 2 15 2 17 2 12 2 14
3 5 3 15 3 25 3 15 3 14 3 12 3 15
4 48 4 35 4 29 4 25 4 26 4 25 4 45 4 21
5 2 5 10 5 14 5 12 5 20 5 22 5 18
6 2 6 5 6 10 6 13 6 19 6 15 6 12 6 8 6 9 6 14
7 13 7 15 7 19 7 20 7 9 7 27 7 30 7 11
8 5 8 10 8 11 8 13 8 15 8 17 8 9 8 11 8 12
9 65 9 65 9 68 9 78 9 77 9 67 9 65 9 63
10 36 10 43 10 38 10 45 10 33 10 44
11 45 11 45 11 40 11 36 11 43 11 36 11 45 11 30
12 50 12 40 12 44 12 40 12 39 12 50 12 44 12 43
;
PROC SORT;
BY MONTH;
/* 95% CONFIDENCE INTERVALS FOR THE MEAN VALUE BY MONTH ARE */
/* CALCULATED FOR EVERY OBSERVATION IN THE DATA SET WATER. */
/* A PROC REG OUTPUT OPTION REQUESTS THE CONFIDENCE INTERVALS TO */
/* BE OUTPUT AS DATA SET CI. */
PROC REG DATA=WATER NOPRINT;
MODEL VALUE=;
OUTPUT OUT= CI L95M=LCWERU95M=LJPPER;
BY MONTH;
/* OUTPUT MAXIMUM AND MINIMUM VALUES FOR THE CONFIDENCE INTERVAL DATA SET ALONG WITH THE NUMBER OF MONTHS TO DATA SET (MEAN). */
PROC MEANS DATA=CI NOPRINT;
VAR LOWER UPPER MONTH:
OUTPUT OUT= MEAN MIN=MINUPPR MAXUPPR MAXMONTH;
/* USE THE MAXIMUM VALUE AND THE MINIMUM VALUE FROM CONFIDENCE */
/* INTERVAL DATA IN CALCULATING THE RANGE OF THE DATA. THIS */
/* INFORMATION WILL BE USED TO IDENTIFY THE APPROPRIATE REFERENCE */
/* SCALE. THE RANGE IS THEN CHECKED TO SEE IF IT IS SMALL AND */
/* THIS SHOULD HAVE THE REFERENCE SCALE REDEFINED. */
/* VARIABLE FACTOR REPRESENTS THE RATIO OF NUMERICAL UNITS TO */
/* PRINTING UNITS. THE VALUE OF FACTOR IS FOUND BY TAKING THE */
/* DATA RANGE AND DIVIDING BY THE NUMBER OF PRINTING FIELDS THAT */
/* FIT ON THE PAGE WIDTH. THIS CODE USES 80, STANDARD SIZE */
/* PAPER (8 1/2 BY 11 INCH) PICA CONTAINS 10 PRINTING FIELDS PER */
/* INCH (E.G. FACTOR = 1 HAS 1 NUMERICAL UNIT = 1 PRINTING UNIT). */
/* THIS EACH DASH (-) IS EQUAL TO A NUMERICAL DISTANCE OF 1. */
/* FACTOR INFORMATION IS ASSEMBLED IN A DATA SET (MEANINFO) THAT */
/* WILL BE MERGED WITH THE ORIGINAL CONFIDENCE INTERVAL DATA (CI). */
DATA MEANINFO (KEEP = MONTH FACTOR);
SET MEAN;
RANGE=MAXUPPR - MINLOWER;
IF RANGE < 40 THEN FACTOR = .5; /* PAGE SIZE */
ELSE FACTOR = CEIL(RANGE/80); /* PAGE SIZE */
DO 1 = 1 TO MAXMONTH;
MOMTH = 1;
OUTPUT;
END;
/* PRINT REFERENCE SCALE */

DATA _NULL_
MERGE MEANINFO CI /* MERGE FACTOR & CONFIDENCE INTERVAL DATA SETS */
BY MONTH; /* NOTE: ONLY ONE OBSERVATION PER BY GROUP IS NEEDED */
FILE PRINT NOTITLES; /* PUTS OUTPUT TO THE LISTING FILE */
IF FIRST.MONTH; /* FIRST OBSERVATION OF A BY GROUP */
IF MONTH = 1 THEN DO;
   PUT ' CONFIDENCE INTERVALS FOR MEAN VALUE BY MONTH'
   PUT;
   PUT;
DO I = 1,5,10,20,30,60,70,80;
   /* PAGE SIZE */
INDEX = I* FACTOR;
PUT @INDEX @; /* TRAILING @ HOLDS LINE */
END; /* END DO LOOP */
X = REPEAT('----',15); /* 15 IS THE NUMBER OF REPEATS - PAGE SIZE */
PUT X; /* ACTUAL PRINTING OCCURS */
END; /* END IF LOOP */

PRINT CONFIDENCE INTERVALS */

LENGTH = ROUND(((UPPER - LOWER) / FACTOR),1) - 2;
START = ROUND((LOWER / FACTOR),1) - 1; /* STARTING POINT OF INTERVAL */
END = '1';
MID = REPEAT('---',LENGTH); /* LENGTH + 1 DASHES ARE PRINTED */
M = -1; /* M IS USED TO POSITION PRINT */
PUT; /* BLANK LINE */
PUT +START END +M MID +M END +2 'MONTH=' MONTH;

/* APPENDIX B */

DATA _NULL_
MERGE MEANINFO CI;
BY MONTH;
FILE PRINT NOTITLES; /* PUTS OUTPUT TO THE LISTING FILE */
IF FIRST.MONTH; /* FIRST OBSERVATION OF A BY GROUP */
IF MONTH = 1 THEN DO;
   PUT ' CONFIDENCE INTERVALS FOR MEAN VALUE BY MONTH'
   PUT;
   PUT;
DO I = 20,30,60,70,80,90,100;
   /* MODIFICATION */
INDEX = I* FACTOR;
ORIG = I - 19; /* MODIFICATION */
PUT @ORIG INDEX @; /* TRAILING @ HOLDS LINE - MODIFICATION */
END; /* END IF LOOP */
X = REPEAT('----',15); /* 15 IS THE NUMBER OF REPEATS */
PUT X;
END; /* END IF LOOP */

PRINT CONFIDENCE INTERVALS */

LENGTH = ROUND(((UPPER - LOWER) / FACTOR),1) - 2;
START = ROUND((LOWER / FACTOR),1) - 21; /* STARTING POINT OF INTERVAL */
END = '1'; /* START HAS MODIFICATION */
MID = REPEAT('---',LENGTH); /* LENGTH + 1 DASHES ARE PRINTED */
M = -1; /* M IS USED TO POSITION PRINT */
PUT; /* BLANK LINE */
PUT +START END +M MID +M END +2 'MONTH=' MONTH;