

Applying A Robust Generalized Axis-scaling Algorithm to the Use of SAS/GRAPH®

Dorothy E. Pugh, ClinTrials Research, Inc.

ABSTRACT

Generating high-quality plots quickly is an important task for the clinical trials programmer. Creating reasonable-looking axes for these plots presents a special challenge because of the mutually conflicting needs to 1) make the number of tick marks moderate and consistent, 2) reduce the number of significant figures in the tick mark labels, and 3) minimize wasted space in the resulting plot. A solution is this simple axis-scaling algorithm, an enhancement to the default algorithm of SAS/GRAPH®, achieves an optimal tradeoff among these criteria to produce a fairly consistent (4 to 8) number of evenly-spaced tick marks with labels that are always integers multiplied by powers of five or ten.

INTRODUCTION

In the clinical trials field, there is a continuing demand for the rapid production of presentable graphs. Typically, a single program produces many of these graphs.

Axis-scaling is the determination of the values of the tick marks along a graph axis. A typical graph of experimental data plots the values of "independent variable" values against those of a "dependent variable." The independent variable contains values that are controlled by the experimenter, such as the data collection time points, or the dose of an administered drug; determining the axis-scaling for these involves simply entering these known quantities. The dependent variable, on the other hand, contains the responses of the experimental subjects, which are highly variable. For example, a single clinical trials laboratory data graphing program may produce approximately 30 graphs, each summarizing the data for one lab parameter over all the patients in the trial. Automation of the determination of tick marks for all of these graphs is a practical necessity as a result.

The most basic consideration in determining tick marks is the determination of the maximum and minimum of the values to be plotted. For aesthetic reasons, there should also be some space between the top of the graph and the maximum point plotted, and an analogous

consideration holds for the minimum value. On the other hand, if this space is too large, the data will appear cluttered and possibly be unreadable. There is an optimum number of tick marks: too many make the graph too cluttered, while too few impart too little information. The tick mark labels should have as few non-zero digits as possible for the sake of both appearance and clarity.

Achieving the best tradeoff among all these considerations is further complicated by the fact that the number of tick marks that minimizes the number of digits (ideally, those that are integral multiples of powers of ten) varies greatly. For example, if the maximum value on an axis is 200 and the minimum 0, one additional tick mark, at 100, would minimize the number of non-zero digits, but the number of tick marks would be too small.

THE CODE

```
%macro axiscale(var,ds);

%global min max interval;

proc means data=&ds noprint;
  var &var;
  output out=saswork._temp
    min=min
    max=max;
run;

data _null_;
  set saswork._temp;
  range = max - min;
  eformat = put(range,e7.);
  rexp= 1*substr(left(eformat),5,3);
  mantissa = abs(range/10**rexp);

*****
* Depending on the value of the ratio *
* of the range to the order of      *
* magnitude of the range, determine *
* interval, and, by implication, the *
* maximum and minimum, for this    *
* plot.                             *
*****;
```

```

if mantissa ge 3 then do;

    max = round(max+.5*10**rexp,
                10**rexp);
    min = round(min-.5*10**rexp,
                10**rexp);
    interval = 10**rexp;

end;

else if 1.3 <= mantissa < 3 then do;

    max = round(max+.25*10**rexp,
                .5*10**rexp);
    min = round(min-.25*10**rexp,
                .5*10**rexp);
    interval = .5*10**rexp;

end;

else if mantissa => 1 then do;

    max = round(max+.5*10**(rexp-1),
                10**(rexp-1));
    min = round(min-.5*10**(rexp-1),
                10**(rexp-1));
    interval = 10**(rexp-1);

end;

*****
* Moderate number of tick marks. *
*****

test = round((max-min)/interval);

do while (test ge 11);
    if mod(test,2) ne 0 then
        max = max + interval;

```

```

    interval = 2*interval;
    test = round((max-min)/interval);
end;

if test le 3 then interval = .5*interval;

*****
* Put these values in macro variables. *
*****

call symput('min',left(min));
call symput('max',left(max));
call symput('interval',left(interval));

run;

%mend axiscale;

```

AUTHOR CONTACT INFORMATION

Dorothy E. Pugh
 Senior Programmer/Analyst
 ClinTrials Research, Inc.
 11000 Weston Blvd.
 Cary, NC 27511
 dpugh@clintrialsresearch.com

SAS is a registered trademark or trademark of
 SAS Institute Inc. in the USA and other
 countries. ® indicates USA registration.