The SAS LEFTTRUNC Macro

Ellen Hertzmark and Donna Spiegelman

August 26, 2013

Abstract

The %LEFTTRUNC macro makes publication-quality Kaplan-Meier-type curves using left-truncated data for a whole sample or for subgroups/strata.

Keywords: SAS, macro, cumulative incidence plot, survival plot, Kaplan-Meier plot, left-truncated data

Contents

1 Description **2**

2 Invocation and Details **2**

3 Examples **7**

3.1 Example 1. Mortality in children started on antiretroviral therapy 7

3.2 Example 2. Including covariates, but not using ADJDAT 8

3.3 Example 3. Give a dataset of values to use as ADJDAT 9

3.4 Example 4. Do not use separate models for different values of the exposure 10

3.5 Example 5. Making a text file of the plotting points 11

4 Importing the Graph into a MS/WORD Document **12**

5 Warnings **13**

6 Frequently asked Questions **13**

6.1 Q: I ran LEFTTRUNC with no STRATA parameter and the default PICTNAME, and the pictname contains _MVAR_ where the STRATA variable name should be. . 13

6.2 Q: All I see is shades of gray. 13

6.3 Q: One of my curves doesn't show. 13

6.4 . Q: I am making 2 curves and want the confidence bands for both to show. Can LEFTTRUNC do this? 13

7 Credits **13**

8 See Also **14**

1 Description

%LEFTTRUNC is a SAS macro that uses the `BASELINE` command in `PROC PHREG` to make publication-quality Kaplan-Meier-type curves using left-truncated data. This is useful for plotting survival (or incidence) curves where the natural time scale is age, for instance. It can also accommodate covariates by running proc phreg with the baseline option. The macro is written to accept the normal Channing dataset, in which a person's time is broken into small pieces, and the age at the beginning of the time period and the follow-up time within the time period are known. The defaults are set to conform to the requirements of JAMA (cumulative incidence, solid lines of different colors, table of number-at-risk for selected times), but the user can control color, line type, and cumulative incidence vs. survival. The user can also control font and font size, as well as whether the label of the vertical axis prints horizontally or vertically. Curves for a whole sample show the 95% confidence bounds, while those for two or more subgroups show only the point estimates. A text file suitable for importing to PC graphics programs can be made.

2 Invocation and Details

In order to run this macro, your program must know where to look for it. You can tell SAS where to look for macros by using the options

```
mautosource sasautos= <directories where macros are located>}.
```

For example, an options statement might be

```
options nocenter ps=78 ls=125 replace
mautosource
sasautos=('/usr/local/channing/sasautos',
   '/proj/nhsass/nhsas00/nhstools/sasautos);
```

This will allow you to use all the SAS read macros for the data sets (`/proj/nhsass/nhsas00/nhstools/sasautos`), as well as other public SAS macros, such as %PM, %INDIC3, %EXCLUDE, and %MPHREG9 (`/usr/local/channing/sasautos`).

NOTE: With this and all other macros, DO NOT include optional parameters in your macro call unless you want to give them non-default values. For example, giving

```
strata=,
```

will override the default and cause problems for the running of the macro.

Below, any values given to the right of the "=" are the defaults.

```
==================
REQUIRED PARAMETERS
==================
data=           Name of data set to use ,
                REQUIRED
tbeg=           Name of the variable for time/age at the beginning of the time interval
                REQUIRED
tend=           Name of the variable for time/age at the end of the time interval
                This is the earliest of
                  TBEG for the next time interval
                  some censoring event
                NOTE:  You must give one of TEND or TIME.
                If you give both, the macro will use TEND.
time=           Name of the variable for survival time within the time period,
                This is the same as for any other proportional hazards model, i.e.
                time from tbeg to the earliest of
                  TBEG for the next time interval
                  some censoring event
                NOTE:  you must give at least one of TIME and TEND.
                If you give both, the macro will use TEND.
event=          Name of the censoring/event variable.
                This variable should be coded so that
                0 means censored and 1 means failure.
                REQUIRED
============================================
OPTIONAL PARAMETERS RELATING TO DATA, MODEL, OR OUTPUT
============================================
where=          A subsetting clause, in the form of a 'where' or 'if' clause.
                NOTE:  Use 'eq', 'ne', 'gt', etc., rather than '=', '^=', '>'.
                OPTIONAL
strata=_mvar_   Name of the strata variable, if you want 2 or more curves.
                This is the parameter to use to compare multiple levels of a variable.
                _mvar_ is a 'junk' variable the macro adds to the working dataset
                for the convenience of the macro programmer.
                OPTIONAL
adj=            List of model variables, if you are modelling
                For example, you may want to control for sex, birth weight,
                or other variables.
adjdat=_basel_  The name of the dataset you want to use for the covariates
                for the curve estimation, if you have model variables.
                If you do not give a value, the macro will make a dataset
                called _basel_, in which all the model variables have their
                mean values.
                In general, we suggest using mean values for continuous
                variables and the indicator for the median or the mode
                for a set of indicators.
                Thus, mean of total cholesterol as a continuous variable,
                quintile 3 for a set of quintiles, most common marital
                status for marital status.
sepmodels=T     Whether you want separate models for the strata (i.e.you want
```

```
                to allow the covariates to have different effect estimates in the
                different strata)
modprint=T      Whether you want the macro to print the model coefficients.
                OPTIONAL
calcmeth=CH     What method you want to use to estimate the survival.
                OPTIONS:  CH, EMP, NELSON (all Breslow estimator),
                          PL, KM (Kaplan-Meier).
                If you have time-varying covariates, you should use the
                Breslow estimator.
timelist=       A list of times for which you want the survival table.
                OPTIONAL
small=.000001   A small number added to TEND (or TBEG + TIME) when TEND = TBEGl to keep the
                observation in the risk set.
                This number should be smaller than the smallest interesting unit of time.
                OPTIONAL
================================================
BASIC PARAMETERS RELATING TO THE GRAPHIC OUTPUT
================================================
plot=2          The type of output you want for your graph:
                  0 = no plot
                  1 = proc plot
                  2 = proc gplot (the default)
                  4 = text file only
                OPTIONAL
outplot=PS      If PLOT=2, the type of format you wish to make the plot in:
                    PS = postscript (the default)
                    JPEG = JPEG
                    HTML = HTML
                    CGM = computer graphics metafile
                    PDF = PDF (Adobe)
                JPEG, PDF, and HTML can be imported into MS/WORD.
                OPTIONAL
                NOTE:  Postscript (PS) does not produce different
                colors (except shades of gray).
                If you need to see the colors, JPEG and PDF work fine.
                You can view the JPEG file using Netscape or StarOffice.
                You can view a PDF using acroread (but it's slow).
                Otherwise, email the file to yourself as an attachment and
                view it on the PC.
plotdata=DATA.TIME.TBEG.EVENT.STRATA.txt
         If PLOT=4, the name of the text file to which the plotting
         points will be output.
         The file will be 'pipe' (|) delimited, and will be in the
         following order:
         STRATA  TIME  SURVIVAL  SLCL  SUCL  INC  ILCL  IUCL,
         where 'LCL' means lower 95% confidence limit, and 'UCL' means
         upper 95% confidence limit.
         OPTIONAL
pictname=DATA.TIME.TBEG.EVENT.STRATA.OUTPLOT
         If PLOT=2, the name of the graphics file.
```

```
         OPTIONAL, but we strongly suggest that you use mnemonic names.
pictdirec=  The directory in which the graph is to be stored.
           If you are running on the Channing system and want the
           graph to be in the same directory as the program, leave this
           parameter blank.  If you want the graph somewhere else,
           give the full path name of the directory, INCLUDING THE
           FINAL SLASH (/).
           If you are running on a PC, you should give the full
           path name of the directory, INCLUDING THE FINAL
           BACKSLASH (\).
pwhich=inc  Whether you want to plot survival or cumulative incidence.
            Options are surv and inc (the default).
            OPTIONAL
vlabel    Label of the vertical axis.
          If you do not give a value for this parameter, the macro will
          check PWHICH and label the axis either
          Fraction not Failed    or
          Cumulative Incidence
          It is always better to bee specific.
          ``Cumulative Incidence of Death,'' rather than ``Cumulative Incidence.''
          OPTIONAL
axordv=0 to 1 by .1  Limits and major tick marks for the vertical axis, in the form
          LOW to HIGH by INCREMENT.
          OPTIONAL
tlabel=Time  Label of the horizontal (time) axis.
             It is always better to make TLABEL as specific as possible,
             For example, ``Years since diagnosis of Breast Cancer,'' rather
             than ``Time (years)''.
             OPTIONAL
axordt    Limits and major tick marks for the time (horizontal) axis.
          If you do not give this parameter, the macro will find the highest
          value of TIME and make the axis
          0 to MAXTIME by (maxtime/10).
          This is unlikely to come out looking nice.
          In general, however, you want around 10 major tick marks.
          NOTE also that people tend to think in years.  If your time
          scale is in months, make the tick marks correspond to years and
          ordinary fractions of years (e.g. 'by 3' or 'by 6', not 'by 5')
          OPTIONAL
==================
OTHER GRAPH OPTIONS
==================
landscape=F  Whether you want the plot to be in landscape, rather than
             portrait.
font=swiss  Name of font to use for the graph.
            Other fonts for which the macro has been tested are cent (century)
            and zapf.
            NOTE:  The font should exist in both regular and BOLD forms.
            OPTIONAL
header1   Top title on graph (if any).
```

```
          OPTIONAL
header2   Second title on graph (if any).
          OPTIONAL
header3   Third title on graph (if any).
          OPTIONAL
hsize1=2  Print size for header1.
hsize2=1.7  Print size for header2.
hsize3=1.5  Print size for header3.
            You may wish to vary these print sizes depending on the font, the
            length of the header, and OUTPLOT (Things come out differently in JPEG
            from PS).
fontmult=1  a multiplier for the standard character size everywhere but
            the headers, to make the characters the size you want them.
            Since the actual size of the characters varies depending on the font
            and on the output device, this allows the user to customize the
            output.  This will require some experimentation.
href      List of values of TEND for which you want vertical lines.
          OPTIONAL
vref      List of values of survival or incidence (depending on pwhich)
          for which you want horizontal lines.
          This may be helpful for estimating when survival gets
          below a specified level (or cumulative incidence gets
          above a specified level).
          OPTIONAL
nolegend=F  Whether you want to prevent the graph from having a legend.
          OPTIONAL
legloc=center bottom outside Location of legend.
                             center bottom outside: prints at the bottom
                             outside the graph.
                             other options are any combination of
                             left/center/right  with  top/middle/bottom.
                             if you say 'inside' instead of 'outside,'
                             the legend will be inside the axes.
                             Usually, left bottom inside is good for survival curves,
                             and right bottom inside is good for incidence curves.
                             OPTIONAL
legacross=1  Number of columns in legend.
             OPTIONAL
legframe=F  Whether you want a frame for the legend (sometimes useful
            if you put the legend inside the axes).
            OPTIONAL
leglabel  A name for the STRATA variable suitable for the title of the
          legend.
          OPTIONAL
leglab1  A description of the first level of the STRATA variable
         for the legend.
         OPTIONAL
.
.
.
```

```
leglab7  A description of the 7th level of the STRATA variable
color1   The color for the first survival/incidence curve.
         default=black
color2   default=red
color3   default=tan
color4   default=lib (light blue)
color5   default=violet
color6   default=gold
color7   default=pink
linetype1 default=1 (solid)
 .
 .
linetype7 default=1 (solid)
         for the legend.
         If you are not using color, a reasonable list of linetypes is
         1, 4, 3, 2, 35, 37, 43 .
         You can also use gray.
         A diagram of all the line types is shown in the SAS/Graph manual,
         version 8, page 249.
         OPTIONAL
linewidth=6  The width of the incidence or survival lines.
         For any given value, the actual width will depend on
         the value of OUTPLOT, so some experimentation may be
         necessary.
         Note that you want reasonably thick lines so they will be
         visible when the graph is shown in reduced form in a paper.
extend   whether the final values of the survival (or incidence) curve(s)
         should be extended beyond the last event.
         default=F
         OPTIONAL
showci=T  Whether to show the confidence band of the
          incidence or survival curve.
          NOTE:  This option only applies when there is one
          curve.  The macro automatically changes showci to
          F if STRATA is given.
vlabelstyle=V  Whether you want the label for the vertical axis to print parallel to the ax
               or horizontally (H).
```

3 Examples

All the examples are from studies of HIV in Tanzania.

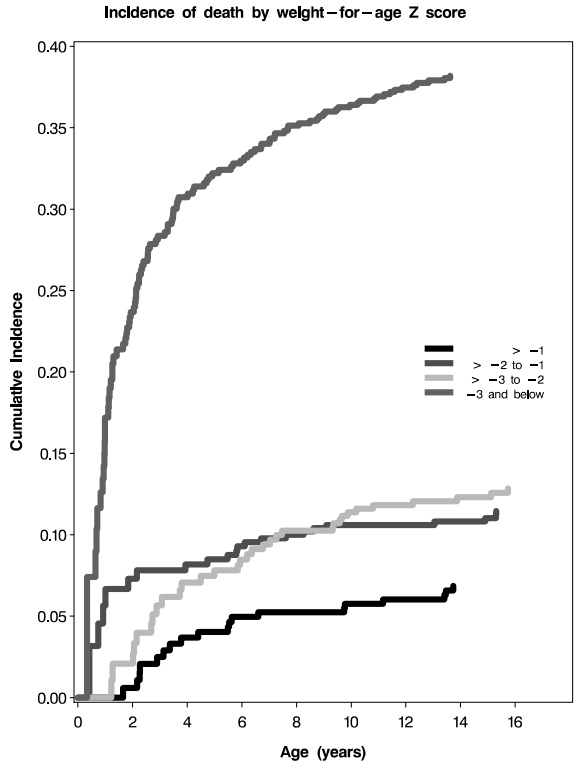
3.1 Example 1. Mortality in children started on antiretroviral therapy

In this example, children were started on antiretroviral therapy at different ages, and we think that age, not time since initiation of antiretroviral therapy, is the relevant time metameter. The *STRATA* variable, which is the exposure, is weight-for-age z-scores, in categories (`WAZCAT`).

The macro call is

```
/* no adjusting variables */
%lefttrunc(data=arvbaseped2, time=timefu, event=arvdeath, tbeg=agearvbase,
tlabel=Age (years),  axordt=0 to 16 by 2, axordv=0 to 0.4 by .05,
pictname=dthplot.bywaz.nomod.ps, pwhich=inc,
notes=notes,
legacross=1, legloc=right middle inside,
leglabel=Weight-for-age z-score,
leglab1=> -1, leglab2=> -2 to -1, leglab3=> -3 to -2, leglab4=-3 and below,
strata=wazcat, header1=Incidence of death by weight-for-age Z score);
```

The graph is



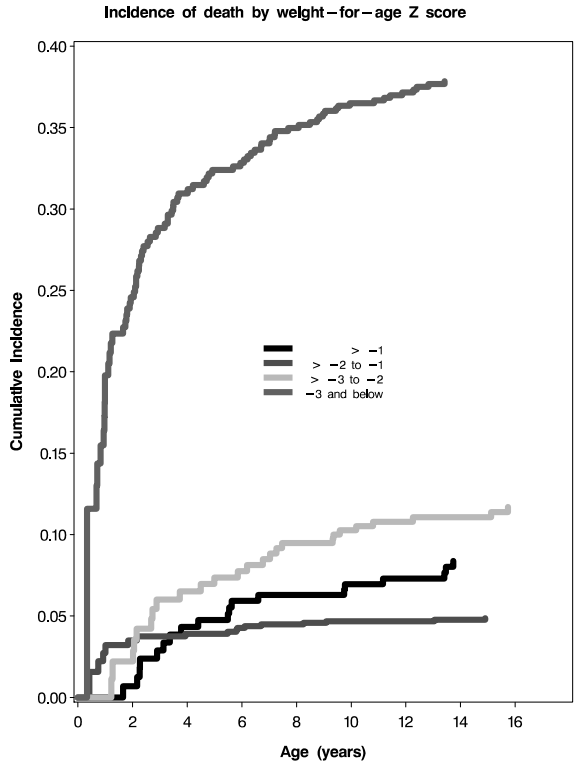
3.2 Example 2. Including covariates, but not using ADJDAT

The covariates are baseline CD4 count (with a missing indicator), sex, and HIV stage. A separate model will be fitted for each level of the exposure (*WAZCAT*).
The macro call is

```
/* not using adjdat */
%lefttrunc(data=arvbaseped2, time=timefu, event=arvdeath, tbeg=agearvbase,
adj= bcd4 bcd4miss msex &stg_ ,
tlabel=Age (years),  axordt=0 to 16 by 2, axordv=0 to 0.4 by .05,
sepmodels=T, modprint=t,
pictname=dthplot.noadjdat.ps, pwhich=inc,
legloc=center middle inside,
```

```
leglabel=Weight-for-age z-score,
leglab1=> -1, leglab2=> -2 to -1, leglab3=> -3 to -2, leglab4=-3 and below,
strata=wazcat, header1=Incidence of death by weight-for-age Z score);
```

The graph is



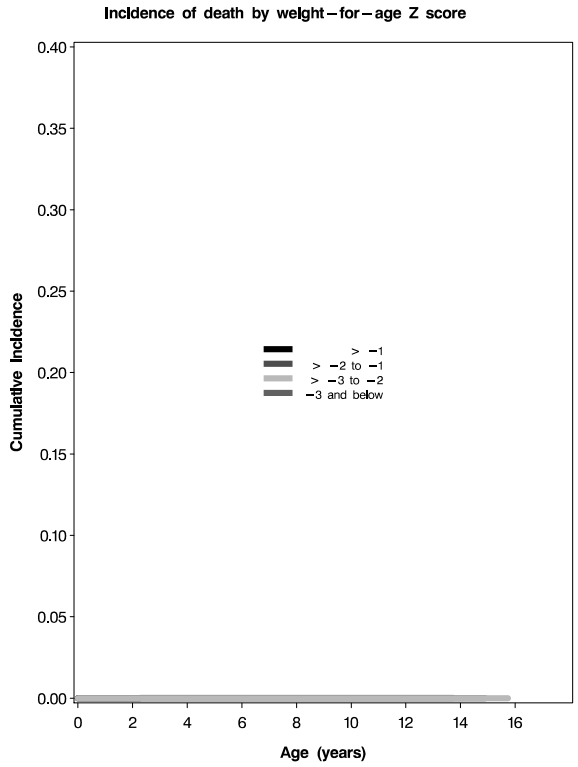
3.3 Example 3. Give a dataset of values to use as ADJDAT

In this example we have the same covariates as in Example 1, but are fitting one stratified model. This forces the coefficients for the covariates to be the same for all levels of weight-for-age z-score category.

The macro call is

```
/* give dataset of values to use for baseline statement */
%lefttrunc(data=arvbaseped2, time=timefu, event=arvdeath, tbeg=agearvbase,
adj=bcd4 bcd4miss msex &stg_ ,
tlabel=Age (years),  axordt=0 to 16 by 2, axordv=0 to 0.4 by .05,
sepmodels=T, modprint=t,
adjdat=cvf,
pictname=dthplot.withadjdat.ps, pwhich=inc,
legloc=center middle inside,
leglabel=Weight-for-age z-score,
leglab1=> -1, leglab2=> -2 to -1, leglab3=> -3 to -2, leglab4=-3 and below,
strata=wazcat, header1=Incidence of death by weight-for-age Z score);
```

The graph is

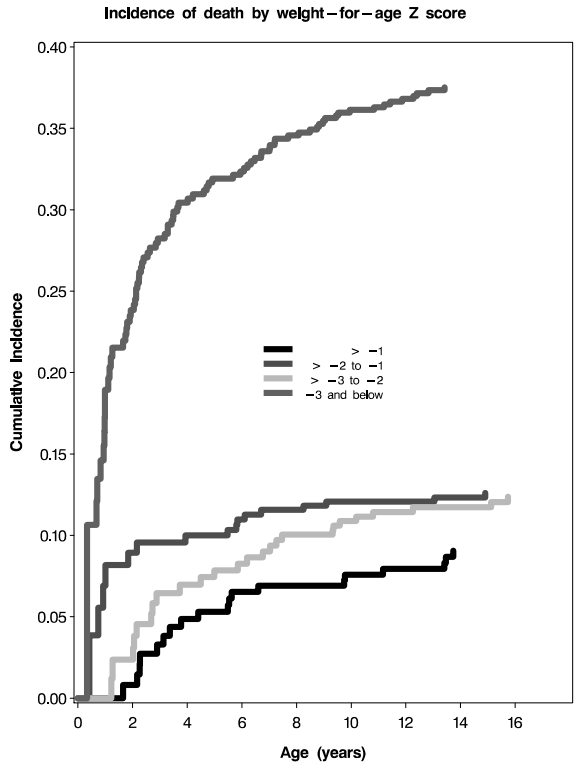


3.4 Example 4. Do not use separate models for different values of the exposure

Not using separate models forces the coefficients of the covariates to be the same for all values of the exposure.

```
/* don't use adjdat */
/* sepmodels=F */
%lefttrunc(data=arvbaseped2, time=timefu, event=arvdeath, tbeg=agearvbase,
adj= bcd4 bcd4miss msex &stg_ ,
tlabel=Age (years),  axordt=0 to 16 by 2, axordv=0 to 0.4 by .05,
sepmodels=f, modprint=t,
pictname=dthplot.nosepmod.ps, pwhich=inc,
legloc=center middle inside,
leglabel=Weight-for-age z-score,
leglab1=> -1, leglab2=> -2 to -1, leglab3=> -3 to -2, leglab4=-3 and below,
strata=wazcat, header1=Incidence of death by weight-for-age Z score);
```

The graph is



3.5 Example 5. Making a text file of the plotting points

This example is the same as that of Example 4, except that instead of making a graph, we make a pipe-delimited text file that can be imported into other graphics programs.

The macro call is

```
/* don't use adjdat */
/* sepmodels=F */
%lefttrunc(data=arvbaseped2, time=timefu, event=arvdeath, tbeg=agearvbase,
adj= bcd4 bcd4miss msex &stg_ ,
tlabel=Age (years),  axordt=0 to 16 by 2, axordv=0 to 0.4 by .05,
sepmodels=f, modprint=t,
plot=4, plotdata=noadjdat.nosepmod.txt,
strata=wazcat, header1=Incidence of death by weight-for-age Z score);
```

Some representative lines of the output text file are

```
1 | 0 | 1 | . | . | 0 | . | .
1 | 1.6527217015 | 0.9919679793 | 0.97635394 | 1 | 0.0080320207 | 0 | 0.02364606
1 | 2.1626892574 | 0.9854291518 | 0.9653696583 | 1 | 0.0145708482 | 0 | 0.0346303417
1 | 2.2584354722 | 0.9790904752 | 0.9556458268 | 1 | 0.0209095248 | 0 | 0.0443541732
1 | 2.266222062 | 0.9727250537 | 0.9463301566 | 0.9998561533 | 0.0272749463 | 0.0001438467
1 | 2.896557318 | 0.9670306467 | 0.938527221 | 0.9963997322 | 0.0329693533 | 0.0036002678 |
.
.
.
```

```
2 | 0 | 1 | . | . | 0 | . | .
2 | 0.4219718818 | 0.9614039664 | 0.8896192145 | 1 | 0.0385960336 | 0 | 0.1103807855
2 | 0.7452415285 | 0.944430397 | 0.8670798373 | 1 | 0.055569603 | 0 | 0.1329201627
2 | 0.9208002884 | 0.9309179079 | 0.8505352706 | 1 | 0.0690820921 | 0 | 0.1494647294
.
.
.
3 | 0 | 1 | . | . | 0 | . | .
3 | 1.2196467195 | 0.9879196085 | 0.9646489847 | 1 | 0.0120803915 | 0 | 0.0353510153
3 | 1.2716834895 | 0.9763412184 | 0.9444365777 | 1 | 0.0236587816 | 0 | 0.0555634223
3 | 2.0197368421 | 0.9690712589 | 0.9345002291 | 1 | 0.0309287411 | 0 | 0.0654997709
3 | 2.0523431867 | 0.9617963235 | 0.9247923808 | 1 | 0.0382036765 | 0 | 0.0752076192
.
.
.
4 | 0 | 1 | . | . | 0 | . | .
4 | 0.3288211968 | 0.8935379571 | 0.7164309209 | 1 | 0.1064620429 | 0 | 0.2835690791
4 | 0.6794700793 | 0.879027448 | 0.7031700552 | 1 | 0.120972552 | 0 | 0.2968299448
4 | 0.7096611391 | 0.8653212527 | 0.6907456547 | 1 | 0.1346787473 | 0 | 0.3092543453
.
.
.
```

The variables are in the order strata (i.e. exposure), time, p(survival), lower cl survival, uppercl survival, incidence, lower cl of incidence, upper cl of incidence.

4 Importing the Graph into a MS/WORD Document

Below are the steps for importing an encapsulated postscript file into a MS-WORD document. A parallel procedure works for a JPEG file.

```
1.  E-mail the file to yourself as an attachment, and download to your PC.
2.  Open your WORD document.
3.  The sequence of keys (at least in Windows XP and its version of WORD) is
        insert
        picture
        from file
        <locate file>
        convert file (this is a window that WORD gives you)
               encapsulated postscript
        Note:  this last step is unnecessary if the extension of the file
        (the part after the last dot) is 'eps' rather than 'ps'.
        MS/WORD will happily deal with JPEG, HTML, and CGM.
```

NOTE: Conversion from encapsulated postscript may not be installed on your computer, but it is available for Windows 95 and beyond. NOTE: When I did the above procedure with a file I made using %LEFTTRUNC, the picture on my Windows screen was fuzzy. When printed, it was crisp.

If you are really having trouble, consider using one of the other formats (HTML, JPEG, CGM, PDF).

5 Warnings

1. The %LEFTTRUNC macro will not plot values for missing STRATA. If you want to treat STRATA=. as just another value, give it a non-missing value.

2. Usually program titles (i.e. title1 'ellens work'; title2 'very silly';) carry over to graphics. To prevent unwanted titles from appearing in graphs, the macro deletes titles. We are sorry for the inconvenience. Titles from the graphs will also continue to later output, unless you restate the titles you want.

6 Frequently asked Questions

6.1 Q: I ran LEFTTRUNC with no STRATA parameter and the default PICTNAME, and the pictname contains _MVAR_ where the STRATA variable name should be.

A: The default value of STRATA is _MVAR_, which, for the convenience of the programmer is set to 1 for every observation in the original data set.

6.2 Q: All I see is shades of gray.

A: Try JPEG to see the colors.

6.3 Q: One of my curves doesn't show.

A: If you are looking at a postscript file in ghostview, the 'tan' curve is very light colored and looks as if it is not there (though if you look closely, you may be able to see it).

6.4 . Q: I am making 2 curves and want the confidence bands for both to show. Can LEFTTRUNC do this?

A: Not directly. What you can do is make each of the curves separately using plot=4, then combine and do the proc gplot yourself. Unless the curves are very different, we do not recommend showing both confidence bands.

7 Credits

Written by Ellen Hertzmark and Donna Spiegelman for the Channing Laboratory. Questions can be directed to Ellen Hertzmark,
`stleh@channing.harvard.edu`, (617) 432-4597 or 432-1200.

8 See Also

Other Channing graphics macros are %LGTPHCURV8, %MIXCURV8, %GLMCURV8.