

TAIL AREA PLOT

PURPOSE

Generates a tail area plot (the empirical survival distribution function against the sorted failure times).

DESCRIPTION

In reliability analysis, many data sets consists of a set of failure times where these failure times are typically truncated at some upper limit value. The cumulative distribution (or CDF) of these failure times is defined as:

$$F(t) = \text{prob}(T < t)$$

where T is the lifetime of a randomly selected unit. That is, the cumulative distribution is the probability that an item fails before time t. The survival distribution function (SDF) is defined as:

$$\begin{aligned} S(t) &= \text{prob}(T > t) \\ &= 1 - F(t) \end{aligned}$$

That is, the survival function is the probability that an item is still surviving at time T.

A tail area plot is a plot of the empirical SDF versus the sorted failure times. The empirical SDF of t is the number of observations greater than or equal to t divided by the number of observations. At each failure time, DATAPLOT calculates the following two points and plots them on the vertical axis:

$$\begin{aligned} y1 &= (N - I + 1)/(N + 1) \\ y2 &= (N - I)/(N + 1) \end{aligned}$$

where N is the number of data points and I is the rank of the failure time. The last failure time only calculates one of these points.

When all of the points are connected, a staircase type plot results. The vertical step is constant for the failure times. The lengths of the horizontal steps are determined by the distances between the sorted failure times. Ties show up as longer vertical strips. The beginning of a horizontal strip represents an estimate of the survival probability at time t_i while the end of the horizontal strip represents an estimate of the survival probability at time t_{i+1} . This estimate is essentially the number of items still surviving divided by the total number of items (plus 1).

SYNTAX

TAIL AREA PLOT <y> <SUBSET/EXCEPT/FOR qualification>
 where <y> is a response variable containing failure times;
 and where the <SUBSET/EXCEPT/FOR qualification> is optional.

EXAMPLES

```
TAIL AREA PLOT Y1
TAIL AREA PLOT Y1 SUBSET TAG > 1
```

NOTE 1

Some analysts prefer to plot the log of the empirical SDF function. Enter the command YLOG ON to do this. In addition, the empirical cumulative hazard function is $-\log(\text{empirical SDF})$. This can be plotted with the following commands:

```
TAIL AREA PLOT Y
LET YNEW = -LOG(YPLOT)
LET XNEW = XPLOT
PLOT YNEW XNEW
```

NOTE 2

If the data are censored, a more complicated estimate is required for the empirical SDF. This estimate is called the Kaplan-Meier or product-limit estimate. DATAPLOT does not generate explicit Kaplan-Meier estimates at this time.

NOTE 3

An estimate of F(t) can be obtained by entering the command LET CDF = 1 - YPLOT after generating the tail area plot.

DEFAULT

None

SYNONYMS

SURVIVAL PLOT

RELATED COMMANDS

LINES	=	Sets the type for plot lines.
CHARACTERS	=	Sets the type for plot characters.
CME PLOT	=	Generates a conditional mean exceedance plot.
WEIBULL PLOT	=	Generates a Weibull plot.
PROBABILITY PLOT	=	Generates a probability plot (over 25 distributions).
PLOT	=	Generates a data or function plot.

REFERENCE

“Statistical Models and Methods for Lifetime Data,” Lawless, Wiley, 1982.

APPLICATIONS

Reliability

IMPLEMENTATION DATE

88/9

PROGRAM

```
SKIP 25
READ HAHN.DAT MILES TAG
LEGEND 1 CUT-OFF IS 135,000 MILES; LEGEND 1 COORDINATES 20 25
TITLE AUTOMATIC
XLABEL MILES WHEN FAILED; Y1LABEL CUMULATIVE PROPORTION SURVIVING
XLIMITS 0 150000
TAIL AREA PLOT MILES
```

