Abstract

Weibull regression is suitable for analyzing survival data in a regression-like format. This alternative to the Cox "proportional hazards" model offers several advantages:

- The analyst can estimate survival probabilities for individuals, together with confidence intervals. These help him interpret and describe results.
- A single parameter describes whether individuals have decreasing, stable, or increasing risk (hazard) functions. This helps test theoretical predictions about rising or falling risks.
- The method is an M-estimate (from robustness theory), which makes available several practical results.

We have written SAS procedures to carry out Weibull regression and calculate the auxiliary statistics required for a satisfying analysis. Several of our procedures lead to novel graphical presentations (e.g. estimates of the combined hazard function of the sample). Our paper will describe these methods and our SAS implementation. The poster session will demonstrate our procedures in action, analyzing a complex breast cancer dataset.

Introduction

Survival analyses are concerned with the occurrence of events over time. For example, cancer studies are frequently concerned with the risk of relapse or death. In demographic studies, the interesting event might be household migration or the birth of a child. Since the study almost inevitably ends before all subjects have been observed, some observations are cut off or censored. The analysis tries to predict relapse from a set of independent variables describing treatment type, initial disease status, demographics, etc.

Weibull regression \[5,8\] uses a regression-like model to predict the time \(T\) to relapse:

\[ \log T = \beta x + e \]

where \(\beta\) is a vector of parameters, \(x\) is a vector of independent variables, and \(e\) is a log-Weibull error term. The error \(e\) can have a scale parameter \(\sigma\) analogous to ordinary regression. This scale parameter describes the entire range of Weibull distributions. We assume that each subject has a time to relapse, although this time may be censored by the end of the study at time \(T_E\). In such a case, we can say only that \(T > T_E\).

A fundamental concept for understanding survival models is the hazard function \(h(t)\). Intuitively, this function describes the risk faced by a patient who has survived up to time \(t\). A declining hazard function means that the patient faces the greatest risk early--more survival implies less risk. Weibull hazard functions are proportional to the power curves \(t^a\), where \(a = (1-\sigma)/\sigma\). This family includes strictly ascending \((\sigma < 1)\), horizontal \((\sigma = 1)\), and strictly descending \((\sigma > 1)\) forms. The exact level of the hazard function depends on the independent variables:

\[ h(t) = \exp(-\beta x/\sigma)) t^a/\sigma. \]

The Cox, or "proportional hazards" model \[4\], implemented in SAS by the supplemental procedures CONREG and PHREG, is an analytic alternative to Weibull regression. In Cox's model, the hazard function is proportional to a function of the parameters, but the underlying hazard is left unspecified:

\[ h(t) = \exp(\beta x) h_0(t). \]

Note that the hazard function has the same exponential relation to the independent variables in the Cox model and Weibull regression. In practice, the parameters for the Weibull model (divided by \(\sigma\)) tend to be very similar to results for the Cox model. Thus tools developed for the Weibull model give useful information to Cox model users.

A Toolkit for Weibull Regression

Our procedure builds on a mixture of classical maximum likelihood theory \[11\], recent results in robust estimation \[1,7\] and regression diagnostics \[2,3,10\], and some new suggestions \[6\].

The first step in constructing a likelihood function is to separate the stochastic model from the deterministic one. The stochastic specification is just the standard two-parameter Weibull distribution. Each individual \(i\) in our sample has superparameters \(\mu_i\) and \(\sigma_i\), with Weibull likelihood:
We call $\mu_i$ and $\sigma_i$ superparameters to distinguish them from the parameters of the deterministic model. The deterministic specification links one (or both) or the superparameters to the independent variables:

$$y_i = \beta x_i$$

and

$$\log y_i = \mu_i$$

The independent variables are not necessarily the same in both models, but we always use the notation $x_i$ to refer to the appropriate set of independent variables.

Usually, no independent variables other than a constant term are used for log $\sigma_i$. In this case, our parameterization of the Weibull regression is identical to Kalbfleisch and Prentice [7]. By specifying independent variables in the log $\sigma_i$ equation, we allow the hazard function (shape) to vary by individual. We find this to be a parsimonious and effective test of the proportional hazards assumption.

The superscore for each individual is just the derivative of the log likelihood with respect to the superparameter:

$$S_i^\mu = \partial (\log L_i) / \partial \mu_i$$

The scores with respect to the parameters are defined in the usual (and parallel way):

$$S_i^\mu = \partial (\log L_i) / \partial \beta_j = x_i S_i^\mu$$

A parallel construction is used for log $\sigma_i$ and its parameters. The overall score vector formed by joining the score vectors for the $\mu$ parameters and the log $\sigma$ parameters is very important in maximum likelihood theory. Call this overall score vector $S_i$.

Two matrices calculated from the scores are very important in classical likelihood theory. These are:

$$R_1 = \sum_j S_j S_j'$$

and

$$R_2 = - \sum_j \partial^2 \log \mathcal{L} / \partial (\text{parameters})$$

$R_2$ is the second derivative matrix of the log likelihood with respect to the parameters. If the model is correct, $R_1$ and $R_2$ are asymptotically equal and may serve as a covariance matrix for the parameters. Inequality of these matrices tests the goodness-of-fit for the Weibull model type.

Huber [7] has shown that the covariance matrix estimate

$$\text{cov} (S_1, S_2) = R_2^{-1} R_1 R_2^{-1}$$

is an asymptotically correct covariance matrix, even if the model is not correct! This covariance matrix also holds up if some of the sample individuals are correlated, as when more than one survival interval is collected for the same sample person. The score cross products must be expanded to:

$$R_1 = \sum_j S_j S_j'$$

where the sum for $j$ includes only those individuals correlated with each observation $i$. The number of such intercorrelated observations should be small. Both of these results are "caveat emptor" in the sense that maximizing the wrong likelihood function often makes the parameter estimates inconsistent or inefficient.

An excellent residual analysis can also be fashioned from the score information. When these methods are applied to a normal distribution likelihood, the parameter estimates are the usual regression parameters, and the supescores are the standardized residuals. In the Weibull regression case, the supescores describe the empirical influence [1] of each observation for estimating the parameters. Censoring does not affect how the numbers are interpreted. (We also implemented Pregibon’s deviance concept [10], which measures how extreme an observation is.)

The effect of including an observation $i$ can be approximated by the one-step Newton-Raphson shift

$$\Delta (S_i, \beta) = R_2^{-1} S_i$$

These effects can be compared to the standard errors of the parameter estimates to show which observations strongly influence particular parameters. In regression, the multivariate "influence" is described by Cook’s distance [3]. The analog for Weibull regression is

$$D_i = d_i/(1-d_i), \quad \text{where } d_i = S_i' R_1^{-1} S_i / p.$$
PROC MLLWEI finds maximum likelihood estimates for Weibull distribution regression-like models. Standard output from the procedure includes model summary statistics, parameter estimates with standard errors, and a collinearity diagnostic. Optional output to SAS datasets includes the parameters, the parameter covariance matrix, and observation-level regression diagnostics. Optional printed output includes iteration summaries and regression diagnostic summaries.

**Standard Input**

A standard call to PROC MLLWEI might look like this:

```sas
PROC MLLWEI DATA=mysasds;
DEFVAR censind logtimel logtime2;
MU x1 x2 x3;
```

**DATA** Name of a SAS dataset containing the observations.

**DEFVAR** A list of the dependent variable information. Each of the variables on this list must appear in the dataset (mysasds).

The first variable on this list indicates the censoring type of each observation. The values are 0=uncensored, 1=right-censored, 2=left-censored, 3=window-censored.

The last two variables describe the survival time (or range of survival times). These must be logarithms of times. In most applications they are the same variable (code the name twice). The second variable on the DEFVAR list (logtimel) is the earliest possible log survival time, and the third variable (logtime2) is the latest possible log survival time. If the data is uncensored (censind=0) or right censored (censind=1), the survival time information is contained in the second variable (logtimel), and the third variable (logtime2) may contain any information. Likewise, if the data is left censored (censind=2) the information is saved in the third variable (logtime2), and the second variable (logtimel) may contain any information.

**MU** This is a list of independent variables for the \( \mu \) superparameter. These variables may be given in any order.

**Optional Input**

Here is an example of a call to PROC MLLWEI using all of the optional parameters:

```sas
PROC MLLWEI DATA=mysasds START=startds OUTEST=estds OUTPASS=passds OUTCOV=covds OUTDEV=devs;
OUTLIER SCORES INFLU NOCONST PRINT=none NKTEHR=mm;
DEFVAR censind logtimel logtime2;
MU x1 x2 x3;
INSIGMA x1;
GROUP groupvar;
```

The optional parameters are defined as follows:

**START** A dataset containing starting values of the parameters. This is compatible with OUTEST output from PROC SYSREG or PROC MLLWEI (see below). Variable names for startds would correspond to those in mysasds. The observations stand for the superparameters. In startds, the variables _TYPE_ and INTERCEP stand for the name of the superparameter and the coefficient of the intercept term respectively. Here is an example of how starting values would be set up from scratch:

```sas
PROC MIWEI DATA=STARTDS;
MU: MODEL LOGTIME1 = X1 X2 X3;
```

PROC SYREG DATA=MYSASDS OUTEST=STARTDS;
MU: MODEL LOGTIME1 = X1 X2 X3;

If a starting value is not found, the PROC MLLWEI assumes it is zero. Thus INSIGMA in the regression example above starts with the INTERCEP and X1 parameters at zero.

**OUTEST** This is a SAS dataset which contains the parameter values and some other useful data after convergence has occurred. It may be used to supply starting values in a future run.

**OUTPASS** This SAS dataset contains the results for each iteration. Except for the additional observation, it has exactly the same information as the OUTEST dataset.

**OUTCOV** This SAS dataset is filled with Huber's estimate of the parameter covariance matrix. It is stored in the format of a correlation matrix with the variable names relabeled. The extra variable_names _SPNAME and _NAME give the superparameter and parameter names corresponding to the internal variable name given in _IGNN_. The internal variable names are unique, but
OUTDEV
This SAS dataset corresponds to the DATA (mysasds), observation for observation. Its contents depend on which observation-level diagnostics (OUTLIER, SCORES, or INFLU) are specified.

OUTLIER
This option causes the signed square-root of the deviance for MU to be placed in the OUTDEV dataset (devds) variable SRDEVOOI. This concept is defined by Pregibon [ ]. It may (caveat emptor) be interpreted like a standard normal variate, extreme values indicating lack of fit.

SCORES
This option writes out the superscores in variable names MU and LNSIGMA. These may be plotted against predicted values for the superparameters or against independent variables, or may be regressed on additional candidates for independent variables in scouting work.

INFLU
This option writes Cook-like influence diagnostics into the OUTDEV dataset. The variable names are FINFLU, FINFLU2, and CRIT_ for the value of the likelihood at that observation.

NOCONST
This option deletes the intercept term from all superparameter linear combinations. Individual intercept terms may be introduced explicitly as independent variables.

PRINT
This optional value (default 10) controls the amount of diagnostic printing. A zero value suppresses just about everything except the final report.

MXPASS
This optional value (default 30) sets the number of iterations permitted before the procedure gives up. CAUTION: if the likelihood is maximized for an infinite value of the parameters, convergence is slow and the program usually gives up.

LNSIGMA
A variable list giving independent variables for the LNSIGMA superparameter.

GROUP
A variable list with exactly one variable that defines grouping in the observations. Grouping affects the covariance matrix of the parameters but not the parameter values themselves.

Notes
Missing values are deleted from the dataset listwise.

PROC EST computes linear combinations of the coefficients, according to data vectors, and produces results in the set of superparameters. It is a general procedure and can be used with any multilinear work. PROC EST is presently being expanded to calculate the variances and covariances of the predicted parameters. The output of this procedure is a data set containing the newly derived variables and the original data values. The EST procedure produces no printed output.

Standard Call
A standard call to PROC EST might look like this:

```
PROC EST DATA=myasds EST=param OUT=predict;
VAR var1 var2 .... varn;
```

DATA
The name of a SAS dataset containing the observations for which the superparameters are calculated. If it is omitted, the last data set created will be used.

EST
The name of a SAS data set containing the coefficients. This data must include a MODEL_ variable which contains the names of the parameters to be estimated.

OUT
The name of the output dataset to be created by EST. The dataset includes all the variables included in the input dataset and the parameters calculated. Omitting OUT_ is equivalent to specifying OUT_=_DATA_ ;

VAR
The list of variables for which linear combinations will be calculated. Each variable on this list must appear in both the DATA and EST datasets. If this list is omitted, linear combinations will be calculated for all numeric variables on the EST dataset excluding the variables named _CRIT_ and _PASS_.

NOCON
This option deletes the intercept term from all super-parameter linear combinations.

PROC HAZARD calculates hazard and survival functions from the output of censoring and/or censoring-truncation models. The procedure requires two input datasets: one with the population superparameter values, and the second with the time values. The output of this procedure is a data set containing the value of the hazard and the survival functions for each time value along with all the variables contained in the original dataset (the dataset referenced by DATA= ). The HAZARD procedure produces no printed output.
A standard call to PROC HAZARD might look like this:

```sas
PROC HAZARD DATA=mysasds PARM=param
   DIST='WEIBULL' OUT=predict;
   VAR timel time2;
```

**DATA**
The name of a SAS dataset containing the time values for which the hazard and survival functions are calculated. If it is omitted, the last data set created will be used.

**PARM**
The name of the SAS data set containing the population superparameter values. This dataset must include the variables MU and LNSIGMA.

**OUT**
The name of the output SAS dataset to be created by HAZARD. The dataset includes all the variables in the input dataset and the variables containing the values for the hazard and survival functions (named HAZ1, HAZ2, ..., SURV1, SURV2, ... respectively). Omitting OUT= is equivalent to specifying OUT=DATA;

**VAR**
Hazard and survival functions will be calculated for the values of these variables. Unlike PROC MLWEI, the time value is not transformed to logarithms. If this list is omitted, calculations will be done for all numeric variables on the dataset.

**DIST**
This parameter provides the distribution used in the calculations of the survival and hazard functions. Presently, only WEIBULL and LOGISTIC distributions are implemented.

### REFERENCES


DATA CANCER1 (KEEP=CENSORED POST HOM MARG MARGNA AGE HISTO MOTHER NODES LOGTIME COUNT (KEEP=NUM));
INFILE CANCER OBS=309 END=EOF;
INPUT ( CODE POST SELF FIRM NIP MARG1 HOM HISTO AGE MOTHER PARA AGE1 YOUNGST ESTRO PROGES BREAST CCC XRAY OPATH PSIZE ALNAV ALNS ALNR NODES MENO METS ) (10*RBB. / 10*RBB. / 8*RBB.);
******,************,*'r**r*:*************-!ri:*~h'r*~'r***,'r*****'1";*SELF = LESION DISCOVERED BY SELF (AS OPPOSED TO DOCTOR)
** FIRM = CONSISTENCY OF LESION
** NIP = NIPPLE INVOLVEMENT (SURROGATE FOR TUMOR UNLIKELY TO INVADE CHEST WALL)
** MARG = MARGIN FROM TUMOR TO FACIA (THE LARGER THE BETTER).
** MARGNA = MARGIN NOT MEASURED (VERY LARGE).
** AGE = AGE OF PATIENT IN YEARS.
** HISTO = SPECIALIZED HISTOLOGY.
** HOM = HOMOLATERAL LYMPH NODES CLINICALLY INVOLVED.
** MOTHER = MATERNAL HISTORY OF BREAST CANCER.
** NODES = NUMBER OF LYMPH NODES INVOLVED (MUST BE 1-3).
** THE PURPOSE OF THE ANALYSIS IS TO PREDICT RISKS FOR NEW PATIENTS.
** PATIENTS WITH A HIGH RISK OF RECURRANCE MIGHT BE GIVEN ADJUVANT CHEMOTHERAPY WHILE THOSE WITH LOW RISK MIGHT BETTER AVOID THIS TREATMENT. A SECONDARY PURPOSE IS TO GIVE THE PHYSICIAN AN INDICATION OF THE AMOUNT OF INFORMATION IN THE PREDICTION.
** THE DATA WAS COLLECTED BY S. J. KISTER, MD (COLUMBIA UNIVERSITY) AND PUBLISHED IN CANCER CHEMOTHERAPY AND PHARMACOLOGY 2: 147-158 (1979);
LOGTIME = LOG(POST);
MARG = (MARG1=1)
MARGNA = MARG1>0 AND MARG1<1;
CENSORED = 1 - METS;
NUM= _N_;********************************************************************;
** PREVIOUS RESEARCH ESTABLISHED THAT NO PATIENTS WITH SELF=O, FIRM=O, OR NIP=1 RELAPSE. ACCORDINGLY, WE DELETE;
IF SELF=O OR FIRM=O OR NIP=1 THEN DELETE;
OUTPUT CANCER1;
IF EOF THEN OUTPUT COUNT;
PROC MLLWEI DATA=CANCER1 OUTTEST=START OUTDEV=DEVSS SCORES INFLU;
DEPVAR CENSORED LOGTIME LOGTIME;
PROC MLLWEI DATA=CANCER1 START=START OUTDEV=DEVSS OUTTEST=ESTS OUTCDEV=COV SCORES INFLU;
DEPVAR CENSORED LOGTIME LOGTIME;
MU HOM MARG MARGNA AGE HISTO MOTHER;
PROC EST OUT=PRDNS1 EST=ESTS(TYPE=EST) DATA=CANCER1;
PROC PRINT DATA=PRDNS1;
DATA CANCER2;
MERGE CANCER1 DEVS;
DROP _FIP __SV18 __SV2S; /* DROP UNNEEDED VARIABLES */
PROC PLOT; PLOT MU=AGE;
* CREATION OF SCHEMATIC PLOTS;

PROC SORT DATA=CANCER OUT=SPLOT; BY MARG;
PROC SPLIT MAX=3.5 MIN=-2.0 DATA=SPLOT;
VAR NU; CLASSES MARG;
PROC SORT DATA=CANCER OUT=SPLIT; BY NODES;
PROC SPLIT MAX=3.5 MIN=-2.0 DATA=SPLIT;
VAR NU; CLASSES NODES;
PROC SORT DATA=CANCER OUT=SPLIT; BY MOTHER;
PROC SPLIT MAX=3.5 MIN=-2.0 DATA=SPLIT;
VAR NU; CLASSES MOTHER;

* CHECK FOR DEPARTURE FROM THE PROPORTIONAL HAZARDS ASSUMPTION;
PROC MLNWELI DATA=CANCER OUTEST=ESTS;
DEPVAR CENSORED LOGTIME LOGTIME;
MU HOM MARG AGE HISTO MOTHER;
SIGMA HOM MARG AGE HISTO MOTHER;
PROC MLNWELI DATA=CANCER OUTEST=ESTS;
DEPVAR CENSORED LOGTIME LOGTIME;
MU HOM MARG AGE HISTO MOTHER;
SIGMA HOM MARG AGE HISTO MOTHER;
PROC EST OUT=PRINS EST=ESTS DATA=CANCER;

**************************************************************************************************;
* CALCULATE THE SURVIVAL CURVE AND THE HAZARD FUNCTION FOR THE
* THE TOTAL POPULATION;
**************************************************************************************************;

PROC HAZARD DATA=CANCER PARM=PRINS DIST='WEIBULL' OUT=HAZARD;
VAR POST;
PROC SORT DATA=HAZARD; BY POST;

* CALCULATE THE HAZARD FUNCTION FOR A SINGLE INDIVIDUAL;
DATA INDIV;
SET PRINS1;
IF POST=1.40;
PROC HAZARD DATA=CANCER PARM=INDIV DIST='WEIBULL' OUT=HAZARD2;
VAR POST;

* CALCULATE KAPLAN-MEIER SURVIVAL CURVE;
DATA SURV;
MERGE CANCER HAZARD COUNT HAZARD2(KEEP=HAZ1 RENAME=(HAZ1=HAZIDV));
RETAIN KM 1 PATRISK;
IF _N_ =1 THEN PATRISK=NUM;
ELSE PATRISK=PATRISK-1;
IF CENSORED=1 THEN PROB=1;
ELSE PROB=(PATRISK-1)/PATRISK;
KM=KM*PROB;
PROC PRINT;
VAR POST SURV1 KM HAZ1 HAZIDV ;
PROC PLOT;
PLOT (HAZ1 HAZIDV)*POST/OVERLAY ;
PLOT (KM SURV1)*POST/OVERLAY ;
BEGINNING ML ROUTINE
REQUESTED MEMORY OBTAINED. REQUEST = 4836
ML ROUTINES BEGINNING ITERATIONS

ITER = 0, CRITERION = -3016.612, PARMS = 0.000E+00 0.000E+00
ITER = 1, CRITERION = -630.260, PARMS = 9.326E-01 1.947E-01
ITER = 2, CRITERION = -175.268, PARMS = 2.168E+00 6.680E-01
ITER = 3, CRITERION = -92.086, PARMS = 3.568E+00 8.405E-01
ITER = 4, CRITERION = -82.270, PARMS = 3.948E+00 6.616E-01
ITER = 5, CRITERION = -92.086, PARMS = 3.568E+00 8.405E-01
ITER = 6, CRITERION = -58.085, PARMS = 4.103E+00 5.162E-01
ITER = 7, CRITERION = -62.755, PARMS = 4.542E+00 -1.179E-02
ITER = 8, CRITERION = -60.867, PARMS = 4.827E+00 -2.219E-01
ITER = 9, CRITERION = -60.446, PARMS = 4.846E+00 -1.196E-01
ITER = 10, CRITERION = -60.639, PARMS = 4.846E+00 -1.153E-01
ITER = 11, CRITERION = -60.639, PARMS = 4.846E+00 -1.153E-01

FINAL REPORT FOR MLWEI - LOG WEIBULL / EXTREME VALUE ML

GOODNESS OF FIT (LL) . . . . -6.0639E+01
TRACE R*INV(R2) . . . . 1.7321
TRIAL T-STAT FOR ABOVE . . . . -0.7577
LOG DET R*MNR(R2) . . . . -0.3395
TRIAL T-STAT FOR ABOVE . . . . -1.8920
NUMBER OF OBSERVATIONS . . 64
PASSES REQUIRED . . . . 11
DATA READING PASSES . . . . 13
CLEAN MAXIMUM . . . . YES

DEPENDENT VARIABLES
CENSORING: 0=NO 1=RI 2=L 3=WINDOW = CENSORED
LEFTMOST LOG (TIME) = LOGTIME
RIGHTMOST LOG (TIME) = LOGTIME

SUPER PARAMETER MU = LOCATION (IN LOGS)

VARIABLE COEFFICIENT SD(COFVF) T CO-LINEAR
INTERCEP 4.8463E+00 2.350E-01 20.62 0.000

SUPER PARAMETER SIGMA = LOG SCALE

VARIABLE COEFFICIENT SD(COFVF) T CO-LINEAR
INTERCEP -1.1532E-01 1.545E-01 -0.74 0.000

INFLUENTIAL OBSERVATIONS ARE LISTED BELOW

OBSERVATION 1 HAS F (INFLUENCE) 0.284
OBSERVATION 1 VARIABLE MU /INTERCEP AFFECTS T BY ABOUT 0.55; FACTOR = 1

REPORT ON THE MOST INFLUENTIAL OBS FOR EACH VAR
OBSERVATION 1 HAS F (INFLUENCE) 0.284
OBSERVATION 64 VARIABLE MU /INTERCEP AFFECTS T BY ABOUT 0.28; FACTOR = 1
OBSERVATION 1 VARIABLE SIGMA /INTERCEP AFFECTS T BY ABOUT 0.25; FACTOR = 1
### Statistical Analysis System

**BEGINNING MLE ROUTINE**

- **REQUESTED MEMORY OBTAINED. REQUEST =**

**MLE ROUTINES LOOKING FOR STARTING VALUES**

<table>
<thead>
<tr>
<th>Iter</th>
<th>Criterion</th>
<th>PARMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-60.639</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-56.672</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-53.213</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-47.611</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-45.878</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-45.672</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-45.665</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-45.650</td>
<td></td>
</tr>
</tbody>
</table>

**MLE ROUTINES BEGINNING ITERATIONS**

<table>
<thead>
<tr>
<th>Iter</th>
<th>Parms</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>4.346E+00</td>
<td>0.000E+00</td>
<td>0.000E+00</td>
<td>0.000E+00</td>
<td>-1.153E-01</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>4.567E+00</td>
<td>-1.452E+01</td>
<td>-1.384E-01</td>
<td>1.025E-01</td>
<td>3.592E-00</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>4.546E+00</td>
<td>-4.105E+00</td>
<td>-6.832E-01</td>
<td>2.219E-01</td>
<td>7.530E-03</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>4.550E+00</td>
<td>-4.171E+00</td>
<td>-1.162E+00</td>
<td>4.545E-01</td>
<td>1.286E-02</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>3.999E+00</td>
<td>-1.538E+00</td>
<td>-1.647E+00</td>
<td>3.955E+00</td>
<td>3.286E-02</td>
</tr>
</tbody>
</table>

**FINAL REPORT FOR MLLWEI / LOG WEIBULL / EXTREME VALUE HL**

<table>
<thead>
<tr>
<th>Goodness of Fit (LL)</th>
<th>-4.5665E+01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial T-Stat for Above</td>
<td>-0.1395</td>
</tr>
<tr>
<td>Log Det R1*INV(R2)</td>
<td>-0.9054</td>
</tr>
<tr>
<td>Trial T-Stat for Above</td>
<td>-0.5742</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>64</td>
</tr>
<tr>
<td>Data Reading Passes</td>
<td>11</td>
</tr>
<tr>
<td>Clean maximum</td>
<td>YES</td>
</tr>
</tbody>
</table>

#### Different Variables

**Super Parameter \( \mu \) = Location (in Logs)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SD (Coeff)</th>
<th>T</th>
<th>Co-linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.945E+00</td>
<td>6.93E-01</td>
<td>5.69</td>
<td>0.060</td>
</tr>
<tr>
<td>Bone</td>
<td>-1.184E+00</td>
<td>3.68E-01</td>
<td>-3.22</td>
<td>0.334</td>
</tr>
<tr>
<td>Marq</td>
<td>-1.615E+00</td>
<td>3.74E-01</td>
<td>-4.31</td>
<td>0.446</td>
</tr>
<tr>
<td>Margna</td>
<td>1.054E+00</td>
<td>8.81E-01</td>
<td>1.20</td>
<td>0.062</td>
</tr>
<tr>
<td>Age</td>
<td>2.409E+02</td>
<td>1.32E+02</td>
<td>1.82</td>
<td>0.353</td>
</tr>
<tr>
<td>Bisto</td>
<td>7.277E+01</td>
<td>4.07E+01</td>
<td>1.76</td>
<td>0.150</td>
</tr>
<tr>
<td>Other</td>
<td>-1.471E+00</td>
<td>5.92E-01</td>
<td>-2.49</td>
<td>0.514</td>
</tr>
</tbody>
</table>

#### Super Parameter \( \sigma \) = Log Scale

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SD (Coeff)</th>
<th>T</th>
<th>Co-linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.670E+00</td>
<td>1.64E+01</td>
<td>-2.23</td>
<td>0.000</td>
</tr>
</tbody>
</table>

### Influential Observations are listed below

<table>
<thead>
<tr>
<th>Observation</th>
<th>Variable</th>
<th>Coefficient</th>
<th>SD (Coeff)</th>
<th>T</th>
<th>Co-linear</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \mu )</td>
<td>-0.55</td>
<td>0.55</td>
<td>0.55</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( \mu )</td>
<td>-0.34</td>
<td>0.34</td>
<td>0.34</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>( \mu )</td>
<td>-0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>( \mu )</td>
<td>-0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>( \mu )</td>
<td>-0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>( \mu )</td>
<td>-0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

### Report on the Most Influential Obs for Each Var

<table>
<thead>
<tr>
<th>Observation</th>
<th>Variable</th>
<th>Coefficient</th>
<th>SD (Coeff)</th>
<th>T</th>
<th>Co-linear</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \mu )</td>
<td>-0.55</td>
<td>0.55</td>
<td>0.55</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( \mu )</td>
<td>-0.34</td>
<td>0.34</td>
<td>0.34</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>( \mu )</td>
<td>-0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>( \mu )</td>
<td>-0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>( \mu )</td>
<td>-0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>( \mu )</td>
<td>-0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>( \mu )</td>
<td>-0.55</td>
<td>0.55</td>
<td>0.55</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
EMPIRICAL HAZARD FUNCTIONS

ACROSS THE TOTAL POPULATION

A SINGLE INDIVIDUAL

DISEASE FREE INTERVAL IN MONTHS

RAND CORPORATION 1982
SURVIVAL CURVES

KAPLAN-MEIER ESTIMATE

CALCULATED VALUE FROM ANALYSIS

DISEASE FREE INTERVAL IN MONTHS

RAND CORPORATION 1982