

Time to Event (Survival) Data

- Survival analysis is the analysis of observed times from a well defined *origin* to the occurrence of a particular event or *end-point*.
 - Time from entry into a clinical trial until death
 - Time to complete a task in a psychological experiment
 - Time from birth to the first appearance of a skin tumor
 - Time until an egg hatches
 - Lifetimes of electrical components
 - Time until a mortgage is refinanced

162

Special features of time to event (survival) data:

- Strictly non-negative observations
- Right skewed distributions
- Censoring

163

Censoring

- **Right censoring:** Time of event is greater than some value, but exact time is unknown
 - Individual survived at least 2 years
- **Left censoring:** Time of event is less than some value, but exact time is unknown
 - Insect eggs are collected from a cage
 - Time to hatch

164

- **Interval censoring:** Event is known to have occurred within a particular time interval, but exact time is unknown.
 - Examine subjects every 3 months
 - Discover a tumor at the 12 month exam
 - Know tumor first appeared between 9 and 12 months
- **Noninformative censoring:**
 - Reasons for censoring are unrelated to (independent of) disease process
 - Key assumption in many models and methods of analysis
 - Current research on methods for informative censoring (e.g., Scharfstein et al. 2001, Biometrics)

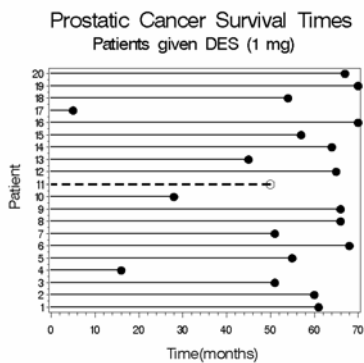
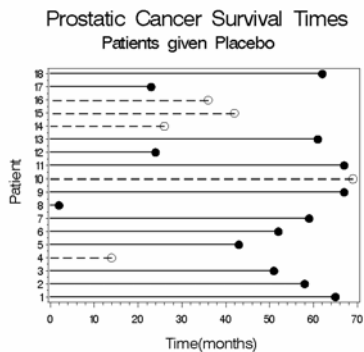
165

- **Independent responses:**

- The time to event for any subject (responding unit) is stochastically independent of the time to event of any other subject (responding unit) in the study
- Key assumption in many models and methods of analysis
- Methods for correlated response times
 - * Initially assume independence, then make adjustments (robust covariance estimation)
 - * Include random effects (frailty models)
 - * Other models

Example: Prostatic cancer

- Compare two treatments
 - placebo
 - 1.0 mg of diethylstilbestrol (DES)
- Survival times
 - start-point: when patient is randomized to a treatment
 - end-point: death from prostatic cancer
 - right censoring:
 - * death from other causes
 - * lost to follow up
 - * end of study period



- Other prognostic factors
 - age of the patient
 - serum hemoglobin level
 - size of tumor
 - Gleason index (tumor stage and grade)

Survivor Function

The observed survival time t of an individual is realization of a non-negative random variable T

- Suppose T has a probability distribution with density function $f(t)$

- The cumulative distribution function (cdf) of T , denoted by

$$F(t) = P(T < t) = \int_0^t f(u) du,$$

represents the probability that the observed survival time would be less than some value t

- The survivor function of T ,
 $S(t) = P(T \geq t) = 1 - F(t)$,
 represents the probability that the observed survival time would be greater than or equal to t .

- Note that $f(t) = \frac{\delta F(t)}{\delta t} = -\frac{\delta S(t)}{\delta t}$

169

Hazard Function

The hazard function $h(t)$ represents an instantaneous failure rate at time t conditional on survival up to time t .

- Consider the conditional probability of failure between t and $t + \Delta t$ given that the unit survived to time t

$$P(t \leq T < t + \Delta t | T \geq t) = \frac{S(t) - S(t + \Delta t)}{S(t)}$$

- Divide by the length of the time interval Δt to get a rate

$$\frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

170

- Take the limit as Δt tends to zero

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{S(t) \Delta t}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \frac{1}{S(t)}$$

$$= f(t)/S(t)$$

- $h(t) \Delta t$ approximates the conditional probability of failure in the interval $[t, t + \Delta t)$ given that the unit has survived to time t (is at risk at time t)

171

Cummulative Hazard

- Note that

$$h(t) = \frac{f(t)}{S(t)} = -\frac{\delta}{\delta t} \{\log[S(t)]\}$$

- Consequently, $S(t) = \exp[-H(t)]$

where

$$H(t) = \int_0^t h(t) \delta t = -\log[S(t)]$$

is the cumulative hazard (or integrated hazard)

- Survival models: often model effects of covariates on the hazard function

172

Common Parametric Models for Survival Time Distributions

Exponential Distribution

- $f(t) = \theta \exp(-\theta t)$
- $F(t) = 1 - \exp(-\theta t)$
- $S(t) = \exp(-\theta t)$
- $h(t) = \theta$: constant hazard

173

Weibull Distribution

- $f(t) = \theta \alpha t^{\alpha-1} \exp(-\theta t^\alpha)$
- $F(t) = 1 - \exp(-\theta t^\alpha)$
- $S(t) = \exp(-\theta t^\alpha)$
- $h(t) = \theta \alpha t^{\alpha-1}$:
 - increasing hazard for $\alpha > 1$
 - decreasing hazard for $\alpha < 1$
- Expected survival time is

$$E(T) = \theta^{-1/\alpha} \Gamma(\alpha^{-1} + 1)$$

- The p -th percentile is

$$t(p) = \left[\frac{1}{\theta} \log \left(\frac{1}{1-p} \right) \right]^{1/\alpha}$$

174

Generalized Gamma Distribution

- Density function

$$f(t) = \frac{\alpha \theta^\rho t^{\rho\alpha-1} \exp(-\theta t^\alpha)}{\Gamma(\rho)}$$

- Survivor function

$$S(t) = 1 - \frac{1}{\Gamma(\rho)} \int_0^{\theta t^\alpha} u^{\rho-1} e^{-u} \delta u$$

- Special cases

- $\rho = 1 \Rightarrow$ Weibull
- $\alpha = 1 \Rightarrow$ Gamma
- $\rho \rightarrow \infty \Rightarrow$ Lognormal
- $\alpha = 1$ and $\rho = 1 \Rightarrow$ Exponential

175

Log-normal distribution

- $\log(T)$ has a normal distribution with mean μ and variance σ^2
- $f(t) = \frac{1}{\sigma\sqrt{2\pi}} t^{-1} \exp(-[\log(t) - \mu]^2 / 2\sigma^2)$
- $S(t) = 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)$
- The p -th percentile is

$$t(p) = \exp(\sigma\Phi^{-1}(p) + \mu)$$
- $h(t)$ does not have a closed form expression
- hazard is unimodal; zero at $t=0$, increases initially, then decreases to zero as $t \rightarrow \infty$

176

Log-logistic distribution

- $\log(T)$ has a logistic distribution
- $f(t) = \frac{\alpha t^{\alpha-1} \exp(\theta)}{(1+t^\alpha \exp(\theta))^2}$
- $S(t) = (1 + t^\alpha \exp(\theta))^{-1}$
- The p -th percentile is

$$t(p) = \left(\frac{p \exp(-\theta)}{1-p} \right)^{1/\alpha}$$
- $h(t) = \frac{\alpha t^{\alpha-1} \exp(\theta)}{1+t^\alpha \exp(\theta)}$
- hazard is unimodal for $\alpha > 1$;
zero at $t = 0$, increases initially, then decreases to zero as $t \rightarrow \infty$
- hazard decreases monotonically if $\alpha \leq 1$

177

Example: Breast Cancer (Leathem and Brooks, 1987, *The Lancet*)

- Survival times (in months) for female breast cancer patients:

5 8 10 13 18 24 26 26 31 35 40 41
 48 50 59 61 68 71 76* 105* 107*
 109* 113 116* 118 143 154* 162*
 188* 212* 217* 225*

- Positive reaction to HPA staining
- Indicates that cancer has metastasised to local lymph nodes
- Right censoring indicated by (*)

178

Maximum Likelihood Estimation

Observe (y_i, d_i) $i = 1, 2, \dots, n$ where
 $Y = \min(T, C)$
 T is the survival time
 C is the censoring time
 $D = I\{C > T\}$ is the failure time indicator

Contribution to the Weibull likelihood

- $f(y_i; \alpha, \theta) = \theta \alpha y_i^{\alpha-1} \exp(-\theta y_i^\alpha)$ if $d_i = 1$
- $S(y_i; \alpha, \theta) = \exp(-\theta y_i^\alpha)$ if $d_i = 0$

The joint likelihood

$$L(\alpha, \theta) = \prod_{i=1}^n \{f(y_i; \alpha, \theta)^{d_i} S(y_i; \alpha, \theta)^{1-d_i}\}$$

179

Maximize the log-likelihood

$$\ell(\alpha, \theta) = \sum_{i=1}^n \{d_i \log(f(y_i; \alpha, \theta)) + (1 - d_i) \log(S(y_i; \alpha, \theta))\}$$

Likelihood (estimating) equations:

$$0 = \frac{\delta \ell(\alpha, \theta)}{\delta \alpha}$$

$$0 = \frac{\delta \ell(\alpha, \theta)}{\delta \theta}$$

Covariance matrix for the large sample normal distribution of the mle

$$\hat{\Sigma} = - \left[\begin{array}{cc} \frac{\delta^2 \ell(\alpha, \theta)}{\delta \alpha^2} & \frac{\delta^2 \ell(\alpha, \theta)}{\delta \alpha \delta \theta} \\ \frac{\delta^2 \ell(\alpha, \theta)}{\delta \alpha \delta \theta} & \frac{\delta^2 \ell(\alpha, \theta)}{\delta \theta^2} \end{array} \right]^{-1} \text{ at } (\hat{\alpha}, \hat{\theta})$$

180

```

/* SAS code for using lifereg to
estimate parameters in a Weibull
model for survival times of breast
cancer patients. This code is
posted as weibull1.sas */

```

```

data set1;
input d y;
datalines;
1 5
1 8
1 10
1 13
. .
. .
. .
0 188
0 212
0 217
0 225
run;

```

181

```

proc lifereg covout outest=set2;
model y*d(0)= / dist=weibull
covb
itprint
maxiter=30
converge=.0001;
output out=set3 censored=c
cdf=cdf
predicted=p
quantiles = .1 to .9 by .1
std=std;

```

```

run;

proc print data=set2; run;

proc print data=set3; run;

```

182

The LIFEREG Procedure

Iteration History for Parameter Estimates

Iter	Ridge	Loglikelihood	Intercept	Scale
0	0	-57.439543	4.033237339	1.01274866
1	0	-53.339308	5.001610136	0.68236274
2	0	-49.33453	4.896150049	0.86739760
3	0	-48.286641	4.865512502	1.02688156
4	0	-48.15151	4.865418619	1.11137314
5	0	-48.147716	4.867589108	1.12881628
6	0	-48.147712	4.867723441	1.12942698
7	0	-48.147712	4.867723441	1.12942698

Model Information

```

Data Set WORK.SET1
Dependent Variable Log(y)
Censoring Variable d
Censoring Value(s) 0
Number of Observations 32
Noncensored Values 21
Right Censored Values 11
Left Censored Values 0
Interval Censored Values 0
Name of Distribution Weibull
Log Likelihood -48.14771192

```

183

Last Evaluation of the Negative of the Gradient

```

Intercept Scale
-1.57928E-6 -0.000016155

```

Last Evaluation of the Negative of the Hessian

```

Intercept Scale
Intercept 16.462773892 -2.81573892
Scale -2.81573892 23.223744185

```

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Conf Limits	Chi-Square	Pr > ChiSq
Intercept	1	4.8677	0.2491	4.380 5.356	381.99	<.0001
Scale	1	1.1294	0.2097	0.785 1.625		

Estimated Covariance Matrix

```

Intercept Intercept Scale
Intercept 0.062029 0.007521
Scale 0.007521 0.043971

```

184

Contents of the file created by the outest= option

Obs	_MODEL_	_NAME_	_TYPE_	_DIST_	_STATUS_
1	y	PARMS	Weibull	0	Converged
2	Intercept	COV	Weibull	0	Converged
3	Scale	COV	Weibull	0	Converged

LNLIKE	y	Intercept	_SCALE_
-48.1477	-1.00000	4.86772	1.12943
-48.1477	4.86772	0.06203	0.00752
-48.1477	1.12943	0.00752	0.04397

Contents of the file created by the OUTPUT statement

Obs	d	y	_PROB_	p	std	cdf	c
1	1	5	0.1	10.238	5.128	0.05433	0
2	1	5	0.2	23.895	8.889	0.05433	0
3	1	5	0.3	40.584	12.393	0.05433	0
4	1	5	0.4	60.889	16.312	0.05433	0
5	1	5	0.5	85.951	21.474	0.05433	0
6	1	5	0.6	117.800	29.106	0.05433	0
7	1	5	0.7	160.353	41.300	0.05433	0
8	1	5	0.8	222.561	62.613	0.05433	0
9	1	5	0.9	333.520	108.156	0.05433	0
.
.
.

185

For our parametrization of the Weibull distribution

$$\hat{\alpha} = \frac{1}{scale} = \frac{1}{1.1294} = .8854$$

$$\hat{\theta} = \exp\left(\frac{-intercept}{scale}\right) = \exp\left(\frac{-4.8677}{1.1294}\right) = .013434$$

186

Use the delta method to estimate the large sample covariance matrix for $\hat{\alpha}$ and $\hat{\theta}$

$$\hat{\theta} = \exp\left(\frac{-intercept}{scale}\right) = \exp\left(\frac{-4.8677}{1.1294}\right) = .013434$$

$$\hat{\alpha} = \frac{1}{scale} = \frac{1}{1.1294} = .8854$$

Let

$$\begin{bmatrix} \theta \\ \alpha \end{bmatrix} = \begin{bmatrix} \exp\left(\frac{-intercept}{scale}\right) \\ \frac{1}{scale} \end{bmatrix}$$

187

Compute

$$G = \begin{bmatrix} \frac{\delta\theta}{\delta int} & \frac{\delta\theta}{\delta scale} \\ \frac{\delta\alpha}{\delta int} & \frac{\delta\alpha}{\delta scale} \end{bmatrix}$$

$$= \frac{-1}{scale^2} \begin{bmatrix} (scale)\exp\left(\frac{-int}{scale}\right) & -(int)\exp\left(\frac{-int}{scale}\right) \\ 0 & 1 \end{bmatrix}$$

and estimate the covariance matrix for $(\hat{\theta}, \hat{\alpha})$ as

$$V = \hat{G} \begin{bmatrix} 0.062099 & 0.007521 \\ 0.007521 & 0.043971 \end{bmatrix} \hat{G}'$$

$$= \begin{bmatrix} 0.0001152 & -0.001697 \\ -0.001697 & 0.0270231 \end{bmatrix}$$

188

```

/* Keep only the columns containing the
   estimates and the covariance matrix */

data set2; set set2;
drop _MODEL_ _NAME_ _TYPE_ _DIST_
     _STATUS_ _LNLIKE_ y;
run;

/* Compute parameter estimates and
   apply the delta method */

proc iml;
start btrans;

use set2;          /* Enter the data */
read all into zz;

n=ncol(zz);       /* Number of parameters */
nr=nrow(zz);

b=zz[1, ];        /* Parameter estimates */

v = zz[2:nr, ];  /* Covariance matrix */

```

189

```

/* Transform to parameters for the
   Weibull model in the notes */

a = - b / b[1,n];
a[1,n] = 1/b[1,n];
a[1,1] = exp(-b[1,1]/b[1,n]);

/* Apply delta method to get a large
   sample covraiance matrix for a */
G = diag(j(1,n,-a[1,n])) ;
G[1,1] = -a[1,1]*a[1,n];
G[ ,n] = -t(a)*a[1,n];
G[1,n] = -a[1,1]*a[1,n]*log(a[1,1]);
G[n,n] = -(a[1,n])##(2);

parms = t(a);
covparms = G*V*t(G);
stdparms = sqrt(vecdiag(covparms));
print parms stdparms, covparms;

finish;

run btrans;

```

190

PARMS	STDPARMS
0.0134348	0.010732
0.8854047	0.1643872

COVPARMS	
0.0001152	-0.001697
-0.001697	0.0270231

191

```

# This file is stored as weibull1.R

# This is code for analyzing time to death
# or censoring for women with breast cancer.
# Data are posted as bcancer.dat

# The variables are as follows:

#Number Name      Description
#-----
# 1 d      Failure indicator:
#          1=death 0=censored
# 2 y      Observed time
#-----

# Enter the data into a data frame.

set1 <- read.table("c:/stat565/data/bcancer1.dat",
                  header=F, col.names=c("d", "y"))

set1$z <- rep(1, length(set1$y))

```

192

```
set1
```

```
  d  y z
1  1  5 1
2  1  8 1
3  1 10 1
4  1 13 1
.  .  . .
.  .  . .
.  .  . .
31 0 217 1
32 0 225 1
```

```
# Attach the survival analysis library
```

```
library(survival)
```

```
weib.out <- survreg( Surv(y,d) ~ z , data=set1,
  dist=c("weibull"))
```

193

```
summary(weib.out)
```

```
Call:
```

```
survreg(formula = Surv(y, d) ~ z, data = set1,
  dist = c("weibull"))
```

	Value	Std. Error	z	p
(Intercept)	4.868	0.249	19.545	4.59e-85
z	0.000	0.000	NaN	NaN
Log(scale)	0.122	0.186	0.656	5.12e-01

```
Scale= 1.13
```

```
Weibull distribution
```

```
Loglik(model)= -122.6
```

```
Loglik(intercept only)= -122.6
```

```
Chisq= 0 on 1 degrees of freedom, p= 1
```

```
Number of Newton-Raphson Iterations: 4
```

```
n= 32
```

194

```
names(weib.out)
```

```
[1] "coefficients" "icoef" "var"
[4] "loglik"       "iter"  "linear.predictors"
[7] "df"           "scale" "idf"
[10] "df.residual"  "terms" "means"
[13] "call"         "dist"  "y"
```

```
# Obtain estimates of the parameters
# in our parameterization of the Weibull
# distribution
```

```
parms <- weib.out$coef
```

```
parms
(Intercept)          z
  4.867724    0.000000
```

```
parms[2] <- 1/(weib.out$scale)
parms[1] <- exp(-parms[1]/weib.out$scale)
```

```
parms
(Intercept)          z
 0.01343485    0.88540417
```

195

```
# Obtain a large sample estimate of the
# covariance matrix for the parameters
# using the delta method. First delete the
# row and column of the covariance matrix
# corresponding to the zero coefficient for z
```

```
V <- weib.out$var[-2, -2]
```

```
# Convert from log-scale to match
# SAS output
```

```
G <- matrix(0, nrow=2, ncol=2)
G[1,1] <- 1
G[2,2] <- weib.out$scale
```

```
V <- G%*%V%*%t(G)
```

```
V
          [,1]      [,2]
[1,] 0.062029512 0.007520732
[2,] 0.007520732 0.043971340
```

196

```

# Convert to our Weibull parameterization

G <- matrix(0, nrow=2, ncol=2)
G[1,1] <- -parms[1]*parms[2]
G[1,2] <- -parms[1]*log(parms[1])*parms[2]
G[2,2] <- -(parms[2])^(2)

covparms <- G%*%V%*%t(G)
stdparms <- sqrt(diag(covparms))

parms
(Intercept)          z
0.01343485  0.88540417
stdparms
[1] 0.01073203 0.16438719

covparms
          [,1]      [,2]
[1,] 0.0001151764 -0.001697105
[2,] -0.0016971050  0.027023149

```

197

```

# Obtain percentiles of the estimated
# Weibull distribution and standard
# errors

p <- c(.1, .25, .5, .75, .9)

tp <- ((log(1/(1-p)))/parms[1])^(1/parms[2])

G2 <- matrix(0, nrow=length(tp), ncol=2)
G2[,2] <- -tp*log(tp)/parms[2]
G2[,1] <- -tp/(parms[1]*parms[2])
stdtp <- sqrt(diag(G2%*%covparms%*%t(G2)))

tp
[1] 10.2379  31.8351  85.95069 188.03600 333.52006

stdtp
[1] 5.12773  10.63239  21.47450  50.32690 108.15592

```

198

Example of a proportional hazards model

- Compare different survival distributions by including information on values for covariates X_1, X_2, \dots, X_k

- Baseline hazard (Weibull distribution where the values of all covariates are zero)

$$h_0(t) = \theta \alpha t^{\alpha-1}$$

- Proportional hazards model

$$h(t) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

- Survivor function

$$S(t) = \exp(-\theta t^\alpha \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k))$$

199

- PROC LIFEREG in SAS and other software transform to a model for $\log(T)$ with survivor function $S(t) =$

$$\exp\left\{-\exp\left(\frac{\log(t) - \mu - (\gamma_1 X_1 + \dots + \gamma_k X_k)}{\sigma}\right)\right\}$$

where

$$\alpha = \sigma^{-1}$$

$$\theta = \exp(-\mu/\sigma)$$

$$\beta_j = -\gamma_j/\sigma$$

- An accelerated failure time model

$$S(t) = S_0(t/\exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k))$$

where $S_0(t)$ is the baseline survivor function. This is a focus of Stat 533.

200

Example: Breast Cancer (Leathem and Brooks, 1987, *The Lancet*)

- Survival times (in months) for female breast cancer patients with positive HPA staining ($X_1 = 1$):

5 8 10 13 18 24 26 26 31 35 40 41
 48 50 59 61 68 71 76* 105* 107*
 109* 113 116* 118 143 154* 162*
 188* 212* 217* 225*

- Survival times (in months) for female breast cancer patients with negative HPA staining ($X_1 = 0$):

23 47 69 70* 71* 100* 101* 148
 181 198* 208* 212* 224*

- Positive reaction to HPA staining indicates that cancer has metastasised to local lymph nodes
- Right censoring indicated by (*)

201

```
/* SAS code for using lifereg to fit a
Weibull model to assess the association
between HPA staining and survival times of
breast cancer patients. This code is
posted as weibull2.sas */
```

```
data set1;
input d y x;
datalines;
1 5 1
1 8 1
1 10 1
. . .
. . .
0 217 1
0 225 1
1 23 0
1 47 0
1 69 0
. . .
. . .
0 212 0
0 224 0
run;
```

202

```
proc lifereg covout outest=set2;
model y*d(0)= x / dist=weibull
covb itprint
maxiter=30
converge=.0001;
output out=set3 censored=c
cdf=cdf
predicted=p
quantiles = .1 to .9 by .1
std=std;
run;
proc print data=set2; run;
proc print data=set3; run;
```

203

Last Evaluation of the Negative of the Gradient

Intercept	x	Scale
-0.00002361	-0.000016995	-0.000045781

Last Evaluation of the Negative of the Hessian

	Intercept	x	Scale
Intercept	22.84694192	18.453297299	-6.329668214
x	18.453297299	18.453297299	-2.740357185
Scale	-6.329668214	-2.740357185	34.70647604

Type III Analysis of Effects (Wald test)

Effect	DF	Chi-Square	Pr > ChiSq
x	1	3.3555	0.0670

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Conf Limits	Chi-Square	Pr > ChiSq
Intercept	1	5.8544	0.4989	4.877 6.832	137.71	<.0001
x	1	-0.9967	0.5441	-2.063 0.070	3.36	0.0670
Scale	1	1.0668	0.1786	0.768 1.481		

204

Estimated Covariance Matrix

	Intercept	x	Scale
Intercept	0.248878	-0.245010	0.026044
x	-0.245010	0.296036	-0.021310
Scale	0.026044	-0.021310	0.031880

Output file created by the OUTPUT statement

Obs	d	y	x	_PROB_	p	std	cdf	c
1	1	5	1	0.1	11.670	5.1540	0.04648	0
2	1	5	1	0.2	25.987	8.7131	0.04648	0
3	1	5	1	0.3	42.859	12.0477	0.04648	0
4	1	5	1	0.4	62.873	15.7708	0.04648	0
5	1	5	1	0.5	87.070	20.5504	0.04648	0
6	1	5	1	0.6	117.265	27.3280	0.04648	0
7	1	5	1	0.7	156.917	37.7092	0.04648	0
8	1	5	1	0.8	213.868	55.2011	0.04648	0
9	1	5	1	0.9	313.382	91.3482	0.04648	0
289	1	23	0	0.1	31.618	17.118	0.07521	0
290	1	23	0	0.2	70.405	34.668	0.07521	0
291	1	23	0	0.3	116.116	55.574	0.07521	0
292	1	23	0	0.4	170.337	81.384	0.07521	0
293	1	23	0	0.5	235.892	114.126	0.07521	0
294	1	23	0	0.6	317.698	157.114	0.07521	0
295	1	23	0	0.7	425.125	216.624	0.07521	0
296	1	23	0	0.8	579.417	307.083	0.07521	0
297	1	23	0	0.9	849.023	476.132	0.07521	0

205

Maximum likelihood estimates

$$\hat{\mu} = 5.8544$$

$$\hat{\sigma} = 1.0668$$

$$\hat{\gamma}_1 = -.9967 \Rightarrow$$

$$\hat{\alpha} = \frac{1}{\hat{\sigma}} = \frac{1}{1.0668} = .9373$$

$$\hat{\theta} = \exp\left(\frac{-\hat{\mu}}{\hat{\sigma}}\right) = \exp\left(\frac{-5.8544}{1.0668}\right) = .004137$$

$$\hat{\beta}_1 = \frac{-\hat{\gamma}_1}{\hat{\sigma}} = \frac{.9967}{1.0668} = .9343$$

with std. err. \approx .50

Use the delta method to obtain approximate standard errors and construct approximate confidence intervals.

206

```
/* Keep only the columns containing the
   estimates and the covariance matrix */
```

```
data set2; set set2;
drop _MODEL_ _NAME_ _TYPE_ _DIST_
     _STATUS_ _LNLIKE_ y;
run;
```

```
proc iml;
start btrans;
```

```
use set2; /* Enter the data */
read all into zz;
```

```
n=ncol(zz); /* Number of parameters */
nr=nrow(zz);
```

```
b=zz[1, ]; /* Parameter estimates */
```

```
v = zz[2:nr, ]; /* Covariance matrix */
```

207

```
/* transform to parameters for weibull
   model in the notes */
```

```
a = - b / b[1,n];
a[1,n] = 1/b[1,n];
a[1,1] = exp(-b[1,1]/b[1,n]);
```

```
/* Apply delta method to get a large
   sample covariance matrix for a */
```

```
G = diag(j(1,n,-a[1,n])) ;
G[1,1] = -a[1,1]*a[1,n];
G[ ,n] = -t(a)*a[1,n];
G[1,n] = -a[1,1]*a[1,n]*log(a[1,1]);
G[n,n] = -(a[1,n])##(2);
```

```
parms = t(a);
covparms = G*V*t(G);
stdparms = sqrt(vecdiag(covparms));
print parms stdparms, covparms;
```

```
finish;
```

208

```
run btrans;
```

PARMS	STDPARMS
0.0041365	0.0037256
0.9342766	0.4995984
0.9374048	0.1568974

COVPARMS

0.0000139	-0.000971	-0.000507
-0.000971	0.2495986	0.0069813
-0.000507	0.0069813	0.0246168

209

Note that $\hat{\beta}_1 = .9343$ is an estimate of the natural logarithm of a ratio of hazards.

- An approximate 95% confidence interval is

$$.9343 \pm (1.96)(.50) \\ \Rightarrow (-.0457, 1.9143)$$

- $\exp(\hat{\beta}_1) = \exp(.9343) = 2.55$ is an estimate of a hazards ratio

- An approximate 95% confidence interval for the hazard ratio is (0.95, 6.78)

210

```
# This file is stored as weibull2.R
```

```
# This is code for analyzing time to death  
# or censoring for women with breast cancer.  
# Data are posted as bcancer2.dat
```

```
# The variables are as follows:
```

#Number	Name	Description
# 1	d	Failure indicator: 1=death 0=censored
# 2	y	Observed time
# 3	x	Treatment: 1=treated 0=control

```
# Enter the data into a data frame.
```

```
set1 <- read.table("c:/stat556/bcancer2.dat",  
header=F, col.names=c("d", "y", "x"))
```

211

```
set1  
  d y x  
1 1 5 1  
2 1 8 1  
3 1 10 1  
4 1 13 1  
.  
.  
.  
31 0 217 1  
32 0 225 1  
33 1 23 0  
34 1 47 0  
.  
.  
.  
44 0 212 0  
45 0 224 0
```

212

```

library(survival)

weib2.out <- survreg( Surv(y,d) ~ x ,
                    data=set1, dist=c("weibull"))

summary(weib2.out)
Call:
survReg(formula = Surv(y, d) ~ x, data = set1,
        dist = c("weibull"))

                Value Std. Error      z      p
(Intercept)  5.8544    0.499  11.735 8.43e-032
x            -0.9967    0.544  -1.832 6.70e-002
Log(scale)   0.0646    0.167   0.386 6.99e-001

Scale= 1.07

Weibull distribution
Loglik(model)= -156.7
  Loglik(intercept only)= -158.8
Chisq= 4.14 on 1 degrees of freedom, p= 0.042
Number of Newton-Raphson Iterations: 4
n= 45

```

213

```

# Obtain estimates of the parameters
# in our parameterization of the Weibull
# distribution

parms <- c(weib2.out$coef,weib2.out$scale)

parms
(Intercept)      x
      5.854364 -0.9966647 1.066777

parms[3] <- 1/parms[3]
parms[2] <- -parms[2]*parms[3]
parms[1] <- exp(-parms[1]*parms[3])

parms
(Intercept)      x
0.004136523 0.9342767 0.9374033

```

214

```

# Obtain a large sample estimate of the
# covariance matrix for the parameters
# using the delta method. First delete the
# row and column of the covariance matrix
# corresponding to the zero coefficient for z

V <- weib2.out$var

# Convert from log-scale to match
# SAS output

np <- length(parms)
G <- diag(rep(1,np))
G[np,np] <- weib2.out$scale

V <- G%*%V%*%t(G)
V
      [,1]      [,2]      [,3]
[1,] 0.24887909 -0.24501138 0.02604437
[2,] -0.24501138 0.29603783 -0.02130995
[3,] 0.02604437 -0.02130995 0.03188056

```

215

```

# Convert to our Weibull parameterization

G <- diag(rep(-parms[np],np))
G[1,1] <- -parms[1]*parms[np]
G[ , np] <- -parms*parms[np]
G[1,np] <- -parms[1]*parms[3]*log(parms[1])
G[np,np] <- -(parms[np])^(2)

covparms <- G%*%V%*%t(G)
stdparms <- sqrt(diag(covparms))
parms
(Intercept)      x
0.004136523 0.9342767 0.9374033

stdparms
[1] 0.003725655 0.499598763 0.156897455

covparms
      [,1]      [,2]      [,3]
[1,] 0.0000138805 -0.0009712012 -0.0005073963
[2,] -0.0009712012 0.2495989243 0.0069812811
[3,] -0.0005073963 0.0069812811 0.0246168113

```

216

```

# Obtain percentiles of the estimated
# Weibull distribution and standard
# errors for the treated population

p <- c(.1, .25, .5, .75, .9)
x <- 1

tp <- ((log(1/(1-p)))/(parms[1]*
      exp(parms[2]*x)))^(1/parms[np])

G2 <- matrix(0, nrow=length(tp), ncol=np)
G2[ ,np] <- -tp*log(tp)/parms[np]
G2[ ,2] <- -x*tp/parms[np]
G2[ ,1] <- -tp/(parms[1]*parms[np])
stdtp <- sqrt(diag(G2%*%covparms%*%t(G2)))
tp
[1] 11.67045 34.07632 87.06994 182.38958 313.38283

stdtp
[1] 5.153977 10.368615 20.550402 45.190065 91.348762

```

```

# Obtain percentiles of the estimated
# Weibull distribution and standard
# errors for the treated population

p <- c(.1, .25, .5, .75, .9)
x <- 0

tp <- ((log(1/(1-p)))/(parms[1]*
      exp(parms[2]*x)))^(1/parms[np])

G2 <- matrix(0, nrow=length(tp), ncol=np)
G2[ ,np] <- -tp*log(tp)/parms[np]
G2[ ,2] <- -x*tp/parms[np]
G2[ ,1] <- -tp/(parms[1]*parms[np])
stdtp <- sqrt(diag(G2%*%covparms%*%t(G2)))
tp
[1] 31.618 92.321 235.893 494.135 849.026

stdtp
[1] 17.118 44.608 114.127 256.425 476.136

```