

# Lecture sixteen: Parametric PH Models: The Weibull Model(III)

## 1. The Weibull PH model (cont.)

### (a) Survival, hazard, and density functions

Unlike in the Cox models, survival, hazard and density functions are determined if the parameters are estimated, say,  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p, \hat{\lambda}$  and  $\hat{\gamma}$ . The hazard and survival functions are

$$\hat{h}_i(t) = \exp(\hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}) \hat{\lambda} \hat{\gamma} t^{\hat{\gamma}-1},$$

and

$$\hat{S}_i(t) = \exp\{-\hat{\lambda} \exp(\hat{\beta}' x_i) t^{\hat{\gamma}}\}.$$

Given the particular values of the covariates, these functions can be plotted against survival time  $t$ .

### (b) Exploratory analyses

log-cumulative hazard plot can be used to assess PH assumption in Weibull model.

- i. Factors
- ii. Continuous variables
- iii. Ignore some of the factors (or continuous variables)?

If there are too many groups, then some of them will end up with insufficient number of individuals for valid estimate of survival.

- iv. Example 5.8: numerical illustration based on an artificial dataset

```
/* log-cumulative hazard plot stratified by A */
options ls=80;
libname fu '../..sdata';
data fu.t51;
infile '../..data/t51.dat';
input st A B censor;
filename x1 'ex58a.pdf';
goptions reset=all gunit=pct border ftext=swissb htitle=6
```

```

htext=2.5 gsfname=x1 ROTATE=LANDSCAPE gsfmode=replace device=pdf;
title1 'Log-cumulative hazard plot: A';
symbol1 c=black i=join v=triangle height=3;
symbol2 c=blue i=join v=star height=3;
proc lifetest notable plot = (lls) method=pl;
    time st*censor(0);
    strata A;

run;
/* log-cumulative hazard plot stratified by B:
                               Similar to A. Omit it here */
/* log-cumulative hazard plot by levels of A & B */
options ls=80;
libname fu '../..sdata';
data w;
    set fu.t51;
filename x2 'ex58ab.pdf';
goptions reset=all gunit=pct border ftext=swissb htitle=6
htext=2.5 gsfname=x2 ROTATE=LANDSCAPE gsfmode=replace device=pdf;
title1 'Log-cumulative hazard plot: A & B';
symbol1 c=black i=join v=triangle height=3;
symbol2 c=blue i=join v=star height=3;
symbol3 c=green i=join v=circle height=3;
symbol4 c=red i=join v=dot height=3;
proc lifetest notable plot = (lls) method=pl;
    time st*censor(0);
    strata A B;

run;

```

Why the log-cumulative hazard plot for B ignoring A is misleading?

```

# Check interaction between A and B: Splus program
ex58.s <- function(){
    tmpdf <- importData("../..sdata/t51.sas7bdat")
    trellis.device(motif)
    attach(tmpdf)
    interaction.plot(a, b, st, xlab="Factor A",
                    ylab="Mean of survival times", lty=1:2, col=c(3,4),
                    trace.label="Factor B")
}

```

```

detach()
}

```

- v. Checking the validity of the adequacy of the fitted models (Cox model: chapter 4, parametric (Weibull) models: chapter 7).

## 2. Comparing alternative Weibull models

Use the methods similar to the ones for the Cox PH model. The differences are that, in the Weibull model, the baseline hazard is known, and also the **likelihood** rather than the **partial likelihood** is used.

- (a) Example 5.9: Kidney cancer study (Table 3.6)

- i. sas program

```

options ls = 80 nodate;
libname fu '..../sdata';
data work;
    set fu.kidney;
    if age = 2 then A2 = 1; else A2 = 0;
    if age = 3 then A3 = 1; else A3 = 0;
    A2N=A2*neph;
    A3N=A3*neph;
proc lifereg;
    model survt*censor(0)= / covb dist=weibull;
proc lifereg;
    model survt*censor(0)= A2 A3 / covb dist=weibull;
proc lifereg;
    model survt*censor(0)= neph / covb dist=weibull;
proc lifereg;
    model survt*censor(0)= neph A2 A3 / covb dist=weibull;
proc lifereg;
    model survt*censor(0)= A2 A3 neph A2N A3N
    / covb dist=weibull;
run;

```

- ii. The log-likelihood

In the sas output from PROC LIFEREG,  $\log \hat{L}$  rather than  $-2\log \hat{L}$  is reported.

Table 1: Example 5.9: values of  $-2\log \hat{L}$  on five models

Terms in model	variables in model	$-2\log \hat{L}$
null model (1)	none	104.886
$\alpha_j$ (2)	A2, A3	96.4
$\nu_k$ (3)	N	96.384
$\alpha_j + \nu_k$ (4)	A2, A3, N	87.758
$\alpha_j + \nu_k + (\alpha\nu)_{jk}$ (5)	A2, A3, N, A2N, A3N	83.064

iii. Interaction?

Comparing model (4) and (5), using Weibull model:  $p = 0.096$  (some degree of interaction); using Cox model:  $p = 0.22$  (no interaction).

iv. The estimates for model (5)

#### Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits	Chi-Square	Pr>ChiSq
Intercept	1	2.5590	0.3727	1.8285 3.2894	47.15	<.0001
A2	1	0.0548	0.5259	-0.9759 1.0855	0.01	0.9170
A3	1	-0.0741	0.7437	-1.5317 1.3836	0.01	0.9207
NEPH	1	1.5676	0.4095	0.7649 2.3702	14.65	0.0001
A2N	1	-0.0781	0.6056	-1.2649 1.1088	0.02	0.8974
A3N	1	-1.6334	0.8283	-3.2568 -0.0100	3.89	0.0486
Scale	1	0.6436	0.0938	0.4837 0.8563		
Weibull Shape	1	1.5538	0.2264	1.1678 2.0674		

Don't forget to convert the estimates in terms of  $\beta$ ,  $\lambda$  and  $\gamma$  when calculate the hazard ratio.

v. hazard ratios

vi. Median survival times

(b) Example 5.10: The ovarian cancer study

SURVT: survival time in days, CENS: censoring indicator  
TREAT: treatment, Surgery + chemotherapy (1: single, 2: combined)  
AGE: age of patient in years

Table 2: Example 5.9, hazard ratios, Table 5.4 at page 211

Age-group	No nephrectomy	Nephrectomy
< 60	1	0.09
60 – 70	0.92	0.09
> 70	1.12	1.24

Table 3: Example 5.9, median survival times, Table 5.5 at page 212

Age-group	No nephrectomy	Nephrectomy
< 60	10.21	48.94
60 – 70	10.78	47.81
> 70	9.48	8.87

RDISEASE: extent of residual disease (1: incomplete, 2: complete)

PERF: performance status (1: good, 2: poor)

i. sas program

```

options ls = 80 nodate;
libname fu '..../sdata';
data fu.ovarian;
infile '..../data/ovarian.dat';
input id survt cens treat age rdisease perf;
data w;
    set fu.ovarian;
treat = treat - 1;
trtage = treat * age;
proc lifereg;
    model survt*cens(0) = age treat /covb dist=weibull;
proc lifereg;
    model survt * cens (0) = /covb dist = weibull;
proc lifereg;
    model survt * cens (0) = age /covb dist = weibull;
proc lifereg;
    model survt * cens (0) = rdisease /covb dist = weibull;
proc lifereg;
    model survt * cens (0) = perf /covb dist = weibull;
proc lifereg;

```

```

        model survt*cens(0)=age rdisease /covb dist = weibull;
proc lifereg;
        model survt * cens (0) = age perf /covb dist = weibull;
proc lifereg;
        model survt * cens (0) = age treat trtage
                                                /covb dist = weibull;

run;

```

ii. The Log-likelihood for various models

Table 4: Example 5.10: Table 5.7

Variables in model	$-2\log \hat{L}$
null model (1)	59.534
<b>TREAT</b> (2)	58.354
<b>AGE</b> (3)	43.566
<b>RDISEASE</b> (4)	55.382
PERF (5)	58.849
<i>AGE, RDISEASE</i> (6)	41.663
AGE, PERF (7)	43.518
<b>AGE, TREAT</b> (8)	41.126
AGE, TREAT, TREAT * AGE (9)	39.708

In the univariate analysis, both AGE and RDISEASE are significant (likelihood ratio test), but RDISEASE is not significant from Wald test ( $p = 0.074$ ). By comparing model (3) and (6), we conclude that only AGE is important variable. Interaction: Is treatment difference consistent over age?

iii. sas output for model (8)

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits	Chi-Square	Pr>ChiSq
Intercept	1	10.9868	1.2763	8.4853 13.4884	74.10	<.0001
AGE	1	-0.0790	0.0198	-0.1177 -0.0402	15.97	<.0001
TREAT	1	0.5615	0.3399	-0.1048 1.2277	2.73	0.0986
Scale	1	0.5489	0.1291	0.3461 0.8705		

Weibull Shape 1 1.8218 0.4286 1.1487 2.8891

What's the hazard ratio for the two treatments?

### 3. The Gompertz PH model

$$f(t) = \lambda e^{\theta t} \exp\left[\frac{\lambda}{\theta}(1 - e^{\theta t})\right],$$

where  $\lambda, \theta > 0, t \geq 0$ .

$$S(t) = \exp\left[\frac{\lambda}{\theta}(1 - e^{\theta t})\right],$$

what if  $\theta < 0$ ? (Cure rate, Ref.: *Survival Analysis with Long-Term Survivors*, Maller; Zhou, 1996)

$$h(t) = \lambda e^{\theta t},$$

The hazard of death is to increase or decrease with time.

- (a) the median survival time

$$t(50) = \frac{1}{\theta} \log\left\{1 + \frac{\theta}{\lambda} \log 2\right\}.$$

- (b) Plot: hazard function (see Splus function `gompertz.s` on the website)
- (c) The proportional hazards property  
If  $h_0(t) = \lambda e^{\theta t}$ , then  $\psi h_0(t)$  is Gompertz hazard function with parameter  $\psi\lambda$  and  $\theta$ .
- (d) The general Gompertz PH model

$$h_i(t) = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \lambda e^{\theta t},$$

- (e) MLE and Software: Not available in SAS LIFEREG.
- (f) Example 5.12: Ovarian cancer - Gompertz PH model

SAS program:

```
options ls = 78 nodate;
libname fu '../sdata';
```

```

data w;
    set fu.ovarian;
/* Age alone */
proc nlmixed data = w;
parms lambda = 0.00002 theta = 0.0016 beta1 = 0;
    xb = beta1 * age;
    phi = exp(xb);
    St = exp((lambda * phi / theta)*(1 - exp(theta * survt)));
    ft = lambda * phi * exp(theta * survt) * St;
    ll = (cens = 1) * log(ft) + (cens = 0) * log(St);
    model ll ~ general(ll);
/* age and treatment (treat) */
proc nlmixed data = w;
parms lambda = 0.00002 theta = 0.0016 beta1 = 0 beta2 = 0.3;
    xb = beta1 * age + beta2 * treat;
    phi = exp(xb);
    St = exp((lambda * phi / theta)*(1 - exp(theta * survt)));
    ft = lambda * phi * exp(theta * survt) * St;
    ll = (cens = 1) * log(ft) + (cens = 0) * log(St);
    model ll ~ general(ll);
run;

```

Part of SAS output:

Parameter Estimates					
Parameter	Estimate	Standard Error	DF	t Value	Pr >  t
lambda	1.706E-6	.	26	.	.
theta	0.001377	0.001053	26	1.31	0.2023
beta1	0.1222	0.008692	26	14.05	<.0001
beta2	-0.8482	0.5942	26	-1.43	0.1653

Corrections:

page 103, line 6, (1.55, 8.18) should be (1.55, 28.16).

page 215, line 8,  $\hat{\lambda} = 5.645$  should be  $\hat{\lambda} = 2.028$ .

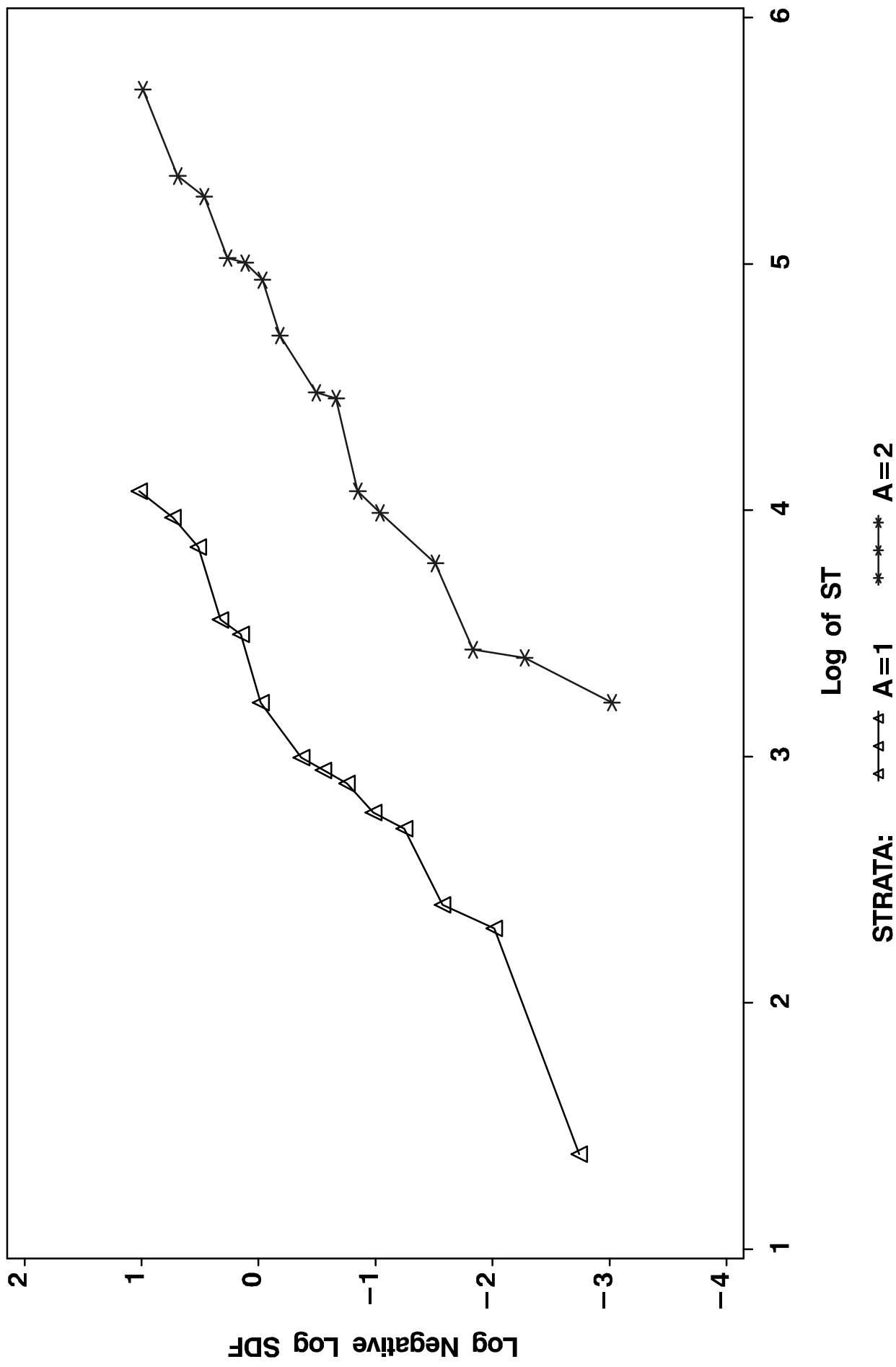


page 218, line 14 from bottom, 1.686 (P = 0.184) should be 1.7 (P = 0.192)  
page 225, line 10,  $f(t)$  should be

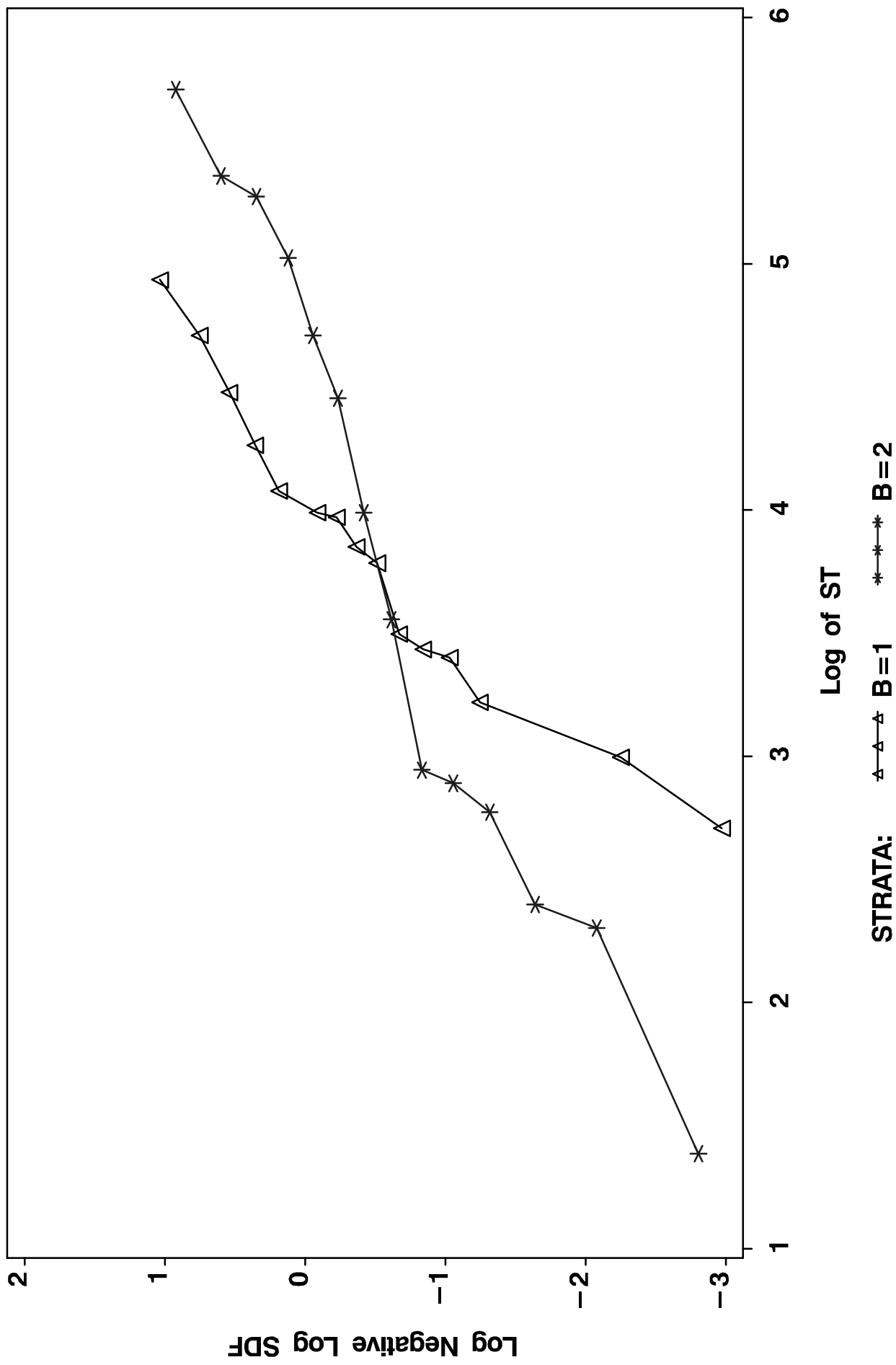
$$f(t) = \left(\frac{\lambda}{2\pi t^3}\right)^{1/2} \exp\left\{-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right\},$$

**Assignment nine:** Using Weibull proportional hazards model to fit the data of recurrence of bladder cancer (Table B.2, page 493). Provide your program and output of your program. Calculate the hazard ratio between two treatment groups (i.e. placebo versus thiotepa) and its 95% confidence interval and interpret your results.

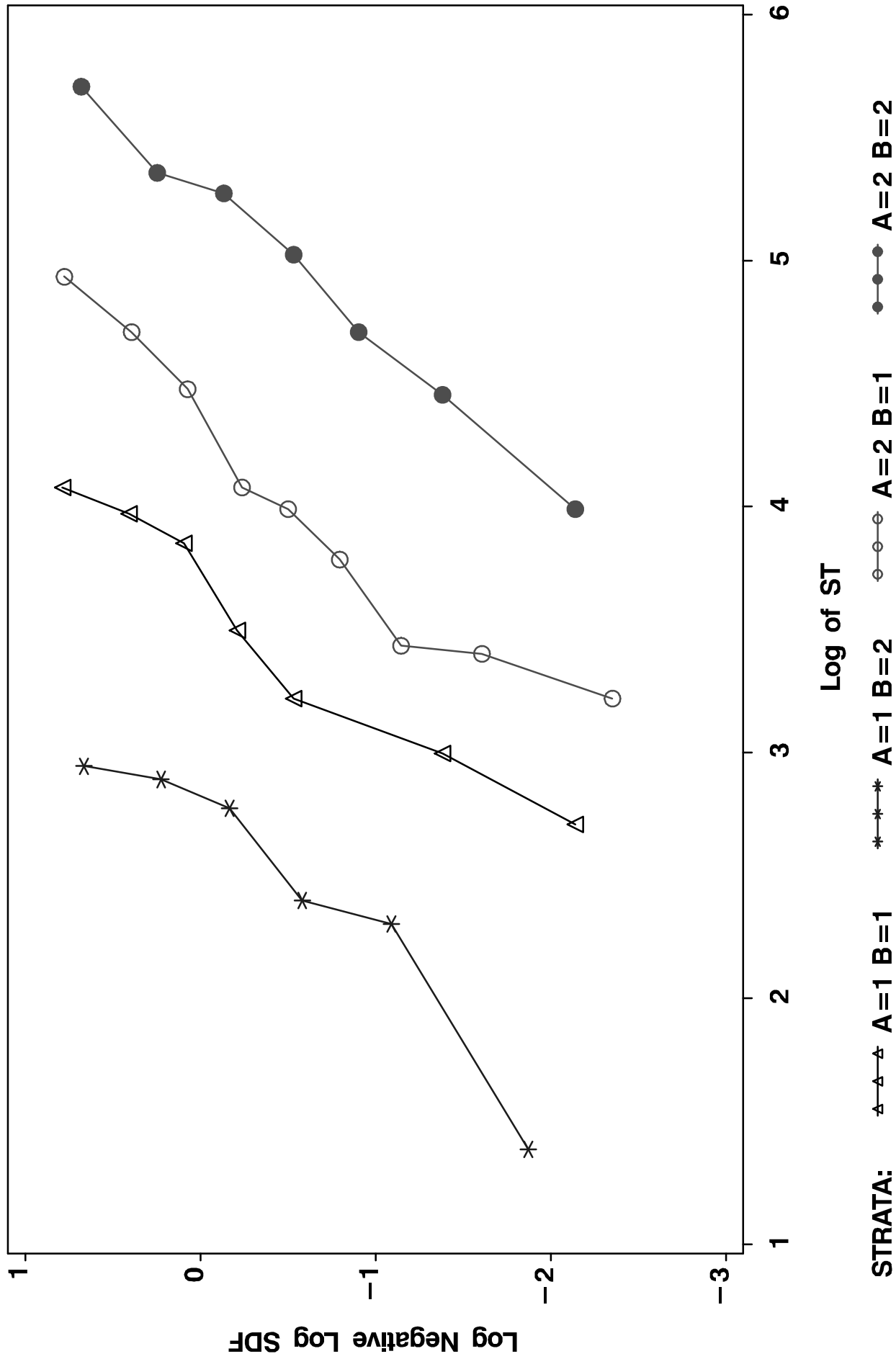
# Log – cumulative hazard plot: A



# Log – cumulative hazard plot: B



# Log - cumulative hazard plot: A & B



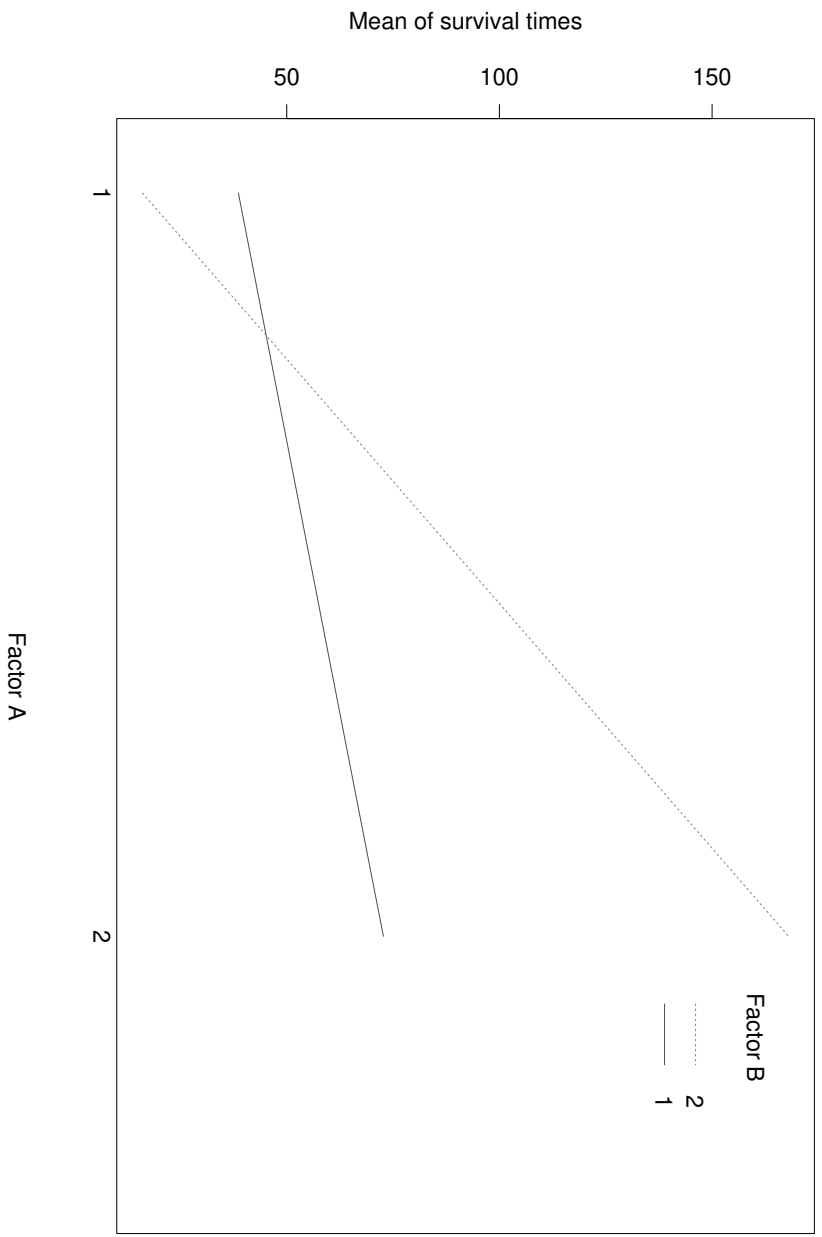


Figure 4:

Figure 5:

A Gompertz distribution with a median of 20

