TMA4275 LIFETIME ANALYSIS

Slides 14: Case study with medical data; Accelerated lifetime models

Bo Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim

http://www.math.ntnu.no/∼bo/
bo@math.ntnu.no

NTNU, Spring 2014

CASE-STUDY IN COX-REGRESSION: PBC-DATA FROM MAYO CLINIC

424 patients with PBC (primary biliary cirrhosis (rare disease))

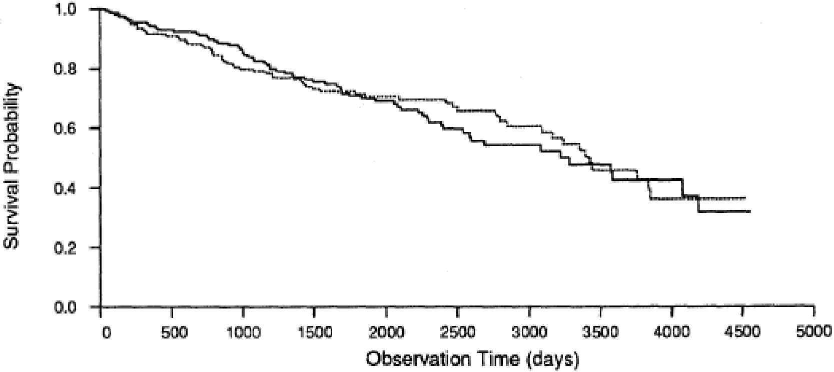
A randomized clinical trial with drug DPCA versus Placebo: 312 patients chosen

Patients included in trial: January 1974 - May 1984

Follow-up until July 1986

First: Compared DPCA group and Placebo group by Kaplan Meier.

KAPLAN-MEIER PLOTS FOR DPCS vs. PLACEBO



Time Interval

Group	0-1000	1000-2000	2000-3000	3000-4000	4000-5000
—— DPCA	23/158	22/128	13/74	5/31	2/10
······· Placebo	31/154	12/120	7/70	10/32	0/11

(# events/# at risk)

Figure 4.4.1 Estimated survival curves in DPCA and placebo groups, PBC data.

Use the same model as for the Battery Data:

x=0 for DCPA $\lambda_0(t)$
x=1 for Placebo $\lambda_0(t)e^{\beta}$
$\hat{\beta} = -0.0571$, $W = 2(l(\hat{\beta}) - l(0)) = 0.102$ (not significant)
$\widehat{SD(\hat{\beta})} = \frac{1}{-\sqrt{l''(\hat{\beta})}} = 0.1792$
95% confidence interval for β : $\hat{\beta} \pm 1.96 \cdot 0.1792$
(-0.408, 0.294)

so CI for relative risk e^{β}: (0.66, 1.34)

Conclusion: In the best case the new drug leads to 1.34 relative risk for not using it (would need at least 1.50 to do further investigations).

The data on the 312 PBC randomized patients can be used to build a statistical model for the influence of covariates on disease outcome.

The data contains 14 clinical, biochemical and histological variables.

Their model is (now $\lambda(\cdot)$ is used instead of $z(\cdot)$ for hazard rate):

$$\lambda(t;\mathbf{x}) = \lambda_0(t)e^{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k}$$

In the beginning k=14

COVARIATES

Table 4.4.1 Prognostic Factors: Summary of Univariate Statistics (312 Patients in the PBC Clinical Trial of DPCA)

Demographic	min	1st Q	med	3rd Q	max	Missing	Rao χ^2(1 d.f.)
Age (years)	26.3	42.1	49.8	56.7	78.4	0	20.86
Sex	male:	36	female:	276		0	4.27
Clinical		**Absent**		**Present**		**Missing**	**Rao χ^2(1 d.f.)**
Ascites		288		24		0	104.02
Hepatomegaly		152		160		0	40.18
Spiders		222		90		0	30.31
Edema[1]	0: 263	1/2: 29		1: 20		0	97.89
Biochemical	**min**	**1st Q**	**med**	**3rd Q**	**max**	**Missing**	**Rao χ^2(1 d.f.)**
Bilirubin	0.3	0.8	1.35	3.45	28.0	0	190.62
Albumin	1.96	3.31	3.55	3.80	4.64	0	70.83
Urine Copper	4	41	73	123	588	2	84.35
Pro Time	9.0	10.0	10.6	11.1	17.1	0	51.76
Platelet Count	62	200	257	323	563	4	12.15
Alkaline Phos	289	867	1259	1985	13862	0	2.58
SGOT	26	81	115	152	457	0	29.59
Histologic	**1**	**2**	**3**	**4**		**Missing**	**Rao χ^2(1 d.f.)**
Stage	16	67	120	109		0	46.49

$\rightarrow$ Bilirubin most significant

$\rightarrow$ Take out expensive/complicated covariates:
stage, urine, copper, SGOT

Remains 11 variables; then a step-down procedure is used to eliminate one (non-significant) variable at a time, arriving at lower table on next slide.

VARIABLE SELECTION: TABLE

Table 4.4.2 Results of variable selection procedure in 312 randomized cases with PBC.

(a) First Step, log likelihood −550.603

	Coef.	Std. Err.	Z stat.
Age	2.819 e-2	9.538 e-3	2.96
Albumin	−9.713 e-1	2.681 e-1	−3.62
Alk. Phos	1.445 e-5	3.544 e-5	0.41
Ascites	2.813 e-1	3.093 e-1	0.91
Bilirubin	1.057 e-1	1.667 e-2	6.34
Edema	6.915 e-1	3.226 e-1	2.14
Hepatomegaly	4.853 e-1	2.913 e-1	2.21
Platelets	−6.063 e-4	1.025 e-3	−0.59
Prothrombin Time	2.428 e-1	8.420 e-2	2.88
Sex	−4.769 e-1	2.643 e-1	−1.80
Spiders	2.889 e-1	2.093 e-1	1.38

(b) Last Step, log likelihood −554.237

	Coef.	Std. Err.	Z stat.
Age	0.0338	0.00925	3.65
Albumin	−1.0752	0.24103	−4.46
Bilirubin	0.1070	0.01528	7.00
Edema	0.8072	0.30775	2.62
Hepatomegaly	0.5903	0.21179	2.79
Prothrombin Time	0.2603	0.07786	3.34

Table 4.4.2: Cox with 11 variable.

Recall: *Z stat* means Coef/Std.Err.

Step-down procedure: From (a) to (b): 5 variables taken out;

Log-likelihood statistic:

$$2 \cdot \text{ difference in log likelihood } = 7.268$$

should be compared to χ^2_5 : $P(\chi^2_5 > 7.268) = 0.201$, so we do not reject the null hypothesis that all these 5 variables have coefficients equal to 0.

Then is considered log-transformations of continuous variables - four variables using logs are added to model, and this leads to increased likelihood!

Finally: Arrives at model 4.4.3(c)

FINAL MODEL (c)

Table 4.4.3 Regression models with log transformations of continuous variables, 312 randomized cases with PBC.

(a) Log likelihood −538.274

	Coef.	Std. Err.	Z stat.
Age	−0.0289	0.07141	−0.41
log(age)	3.2248	3.71828	0.87
Albumin	1.0068	1.73450	0.58
log(Albumin)	−5.8629	5.42315	−1.08
Bilirubin	−0.0461	0.03547	−1.30
log(Bilirubin)	1.0774	0.21127	5.10
Edema	0.8238	0.30386	2.71
Prothrombin Time	−0.6175	1.14523	−0.54
log(Pro Time)	10.1928	13.36131	0.76
Hepatomegaly	0.1964	0.22628	0.87

(b) Log likelihood −541.064

	Coef.	Std. Err.	Z stat.
Age	0.0337	0.00864	3.89
Albumin	−0.9473	0.23713	−3.99
log(Bilirubin)	0.8845	0.09854	8.98
Edema	0.8006	0.29914	2.68
Prothrombin Time	0.2463	0.08426	2.92

(c) Log likelihood −540.412

	Coef.	Std. Err.	Z stat.
Age	0.0333	0.00866	3.84
log(Albumin)	−3.0553	0.72408	−4.22
log(Bilirubin)	0.8792	0.09873	8.90
Edema	0.7847	0.29913	2.62
log(Prothrombin Time)	3.0157	1.02380	2.95

ESTIMATION OF SURVIVAL PROBABILITIES

Recall:

$$S(t;\mathbf{x}) = P(T > t;\mathbf{x}) = S_0(t)^{e^{\boldsymbol{\beta}'\mathbf{x}}} = e^{-\Lambda_0(t)e^R}$$

where $R = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k = \boldsymbol{\beta}'\mathbf{x}$ is called Risk Score.

Estimated value: $\hat{S}(t;\mathbf{x}) = e^{-\hat{\Lambda}_0(t)e^{\hat{R}}}$

In the data we have the median value: $\hat{R} = 5.24$, and for this value we get the one- and five-year survival estimates:

$\hat{S}(1) = 0.982$
$\hat{S}(2) = 0.845$

A low-risk example:

Bilirubin 0.5; Albumin 4.5; Age 52; Prothrombin 10.1; edema 0; gives

$$\hat{R} = 0.879 \cdot \ln 0.5 - 3.0553 \cdot \ln 4.5 - \cdots = 3.49$$

so $\Rightarrow \hat{S}(5) = 0.97$

Suppose we want to find the distribution ($R(t)$, MTTF, etc.) for the lifetime of a product.

Problem: MTTF may be so *large* that one would need to let experiments last several years.

Solution: Increase stress, use a regression model, and then extrapolate to normal conditions.

Suppose you want to investigate the deterioration of an insulation used for electric motors. The motors normally run between 80 and 100 degrees Celsius.

To save time and money, you decide to use *accelerated life testing.*

First you gather failure times for the insulation at abnormally high temperatures: 110, 130, 150, and 170 degrees Celsius, to speed up the deterioration. With failure time information at these temperatures, you can then extrapolate to 80 and 100 degrees.

It is known that an *Arrhenius* relationship exists between temperature and failure time.

This is an example from MINITAB (next slide)

MINITAB WORKSHEET

Insulate.MTW ***

↓	C1	C2	C3	C4	C5-T	C6	C7	C8	C9	C10	C11
	Temp	ArrTemp	Plant	FailureT	Censor	Design	NewTemp	ArrNewT	NewPlant		
1	170	26,1865	1	343	F	80	80	32,8600	1		
2	170	26,1865	1	869	F	100	80	32,8600	2		
3	170	26,1865	1	244	C		100	31,0988	1		
4	170	26,1865	1	716	F		100	31,0988	2		
5	170	26,1865	1	531	F						
6	170	26,1865	1	738	F						
7	170	26,1865	1	461	F						
8	170	26,1865	1	221	F						
9	170	26,1865	1	665	F						
10	170	26,1865	1	384	C						
11	170	26,1865	2	394	C						
12	170	26,1865	2	369	F						
13	170	26,1865	2	366	F						
14	170	26,1865	2	507	F						
15	170	26,1865	2	461	F						
16	170	26,1865	2	431	F						
17	170	26,1865	2	479	F						
18	170	26,1865	2	106	F						
19	170	26,1865	2	545	F						
20	170	26,1865	2	536	F						
21	150	27,4242	1	2134	C						
22	150	27,4242	1	2746	F						
23	150	27,4242	1	2859	F						
24	150	27,4242	1	1826	C						

MINITAB SETUP

Methods and Formulas – Accelerated Life Testing

Equation	Models	Residuals
Lifetime regression	Linear	Ordinary
Response variable	Arrhenius	Standardized
Error term	Inverse temp	Cox-Snell
	Loge (Power)	

Equation

Lifetime regression

The regression model estimates the percentiles of the failure time distribution:

$$Y = \beta_0 + \beta_1 X + \sigma\varepsilon$$

where:

- Y = either failure time or log(failure time)
- β_0 = y-intercept (constant)
- β_1 = regression coefficient
- X = predictor values (may be transformed)
- σ = 1/shape (Weibull distribution) or scale (other distributions)
- ε = random error term

Response variable

Depending on the distribution, Y = failure time or log (failure time):

- For the Weibull, exponential, lognormal, and loglogistic distributions, Y = log (failure time)
- For the normal, extreme value, and logistic distributions, Y = failure time

When Y = log (failure time), Minitab takes the antilog to display the percentiles on the original scale.

Error term

The value of the error distribution also depends on the distribution chosen.

- For the normal distribution, the error distribution is the standard normal distribution – normal (0,1). For the lognormal distribution, Minitab takes the log base e of the data and uses a normal distribution.
- For the logistic distribution, the error distribution is the standard logistic distribution – logistic (0, 1). For the loglogistic distribution, Minitab takes the log of the data and uses a logistic distribution.
- For the extreme value distribution, the error distribution is the standard extreme value distribution – extreme value (0, 1). For the Weibull distribution and the exponential distribution (a type of Weibull distribution), Minitab takes the log of the data and uses the extreme value distribution.

Back to top

Models

Linear

$Y = \beta_0 + \beta_1$ * accelerating variable + $\sigma\varepsilon$

where:

- Y is the failure time or log failure time
- σ is the reciprocal of the shape parameter (Weibull distribution) or the scale parameter (other distributions)
- ε is the random error term

Arrhenius

$Y = \beta_0 + \beta_1$ * [11604.83/Degrees Celsius + 273.16)] + $\sigma\varepsilon$

where:

- Y is the failure time or log failure time
- σ is the reciprocal of the shape parameter (Weibull distribution) or the scale parameter (other distributions)
- ε is the random error term

Inverse temp

$Y = \beta_0 + \beta_1$ * [1/(Degrees Celsius + 273.16)] + $\sigma\varepsilon$

where:

- Y is the failure time or log failure time
- σ is the reciprocal of the shape parameter (Weibull distribution) or the scale parameter (other distributions)
- ε is the random error term

Loge (Power)

$Y = \beta_0 + \beta_1$ * log(accelerating variable) + $\sigma\varepsilon$

where:

- Y is the failure time or log failure time
- σ is the reciprocal of the shape parameter (Weibull distribution) or the scale parameter (other distributions)
- ε is the random error term

Back to top

MINITAB Help

File Edit Bookmark Options Help

Help Topics | Back | Print | << | >> | Glossary | Exit

Example of Accelerated Life Testing

main topic interpreting results session command see also

Suppose you want to investigate the deterioration of an insulation used for electric motors. The motors normally run between 80 and 100° C. To save time and money, you decide to use accelerated life testing.

First you gather failure times for the insulation at abnormally high temperatures – 110, 130, 150, and 170° C – to speed up the deterioration. With failure time information at these temperatures, you can then extrapolate to 80 and 100° C. It is known that an Arrhenius relationship exists between temperature and failure time.

To see how well the model fits, you will draw a probability plot based on the standardized residuals.

1 Open the worksheet INSULATE.MTW.

2 Choose **Stat > Reliability/Survival > Accelerated Life Testing**.

3 In **Variables/Start variables**, enter **FailureT**. In **Accelerating variable**, enter **Temp**.

4 From **Relationship**, choose **Arrhenius**.

5 Click **Censor**. In **Use censoring columns**, enter **Censor**, then click **OK**.

6 Click **Graphs**. In **Enter design value to include on plot**, enter **80**. Click **OK**.

7 Click **Estimate**. In **Enter new predictor values**, enter **Design**, then click **OK** in each dialog box.

Session window output

Regression with Life Data: FailureT versus Temp

```
Response Variable:  FailureT

Censoring Information                  Count
Uncensored value                          66
Right censored value                      14
Censoring value:  Censor = C

Estimation Method:  Maximum Likelihood
Distribution:  Weibull
Transformation on accelerating variable:  Arrhenius
```

Regression Table

Predictor	Coef	Standard Error	Z	P	95.0% Normal CI Lower	95.0% Normal CI Upper
Intercept	-15.1874	0.9862	-15.40	0.000	-17.1203	-13.2546
Temp	0.83072	0.03504	23.71	0.000	0.76204	0.89940
Shape	2.8246	0.2570			2.3633	3.3760

Log-Likelihood = -564.693

Anderson-Darling (adjusted) Goodness-of-Fit

At each accelerating level

Level	Fitted Model
110	*
130	*
150	*
170	*

Table of Percentiles

Percent	Temp	Percentile	Standard Error	95.0% Normal CI Lower	95.0% Normal CI Upper
50	80.0000	159584.5	27446.85	113918.2	223557.0
50	100.0000	36948.57	4216.511	29543.36	46209.94

In this example we need connection between lifetime (T) and temperature.

Some standard relations are known to be useful in accelerated testing:

$$\ln T = \beta_0 + \beta_1(\text{function of accelerated variable}) + \sigma W$$

i.e. $\ln T = \beta_0 + \beta_1 g(s)$ for some function $g(\cdot)$ of the stress.

The model used here is the Arrhenius model:

$$\ln T = \beta_0 + \beta_1 \cdot \frac{11604.83}{s + 273.16} + \frac{1}{\alpha} W$$

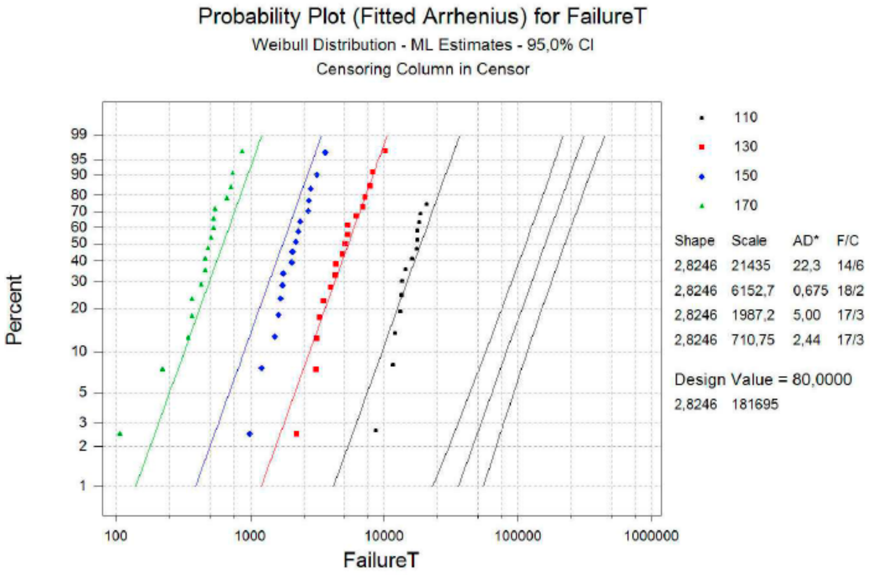
where W is Gumbel and $s =$ temperature in $^\circ\mathrm{C}$,
so $s + 273.16 =$ temp in $^\circ\mathrm{K}$ (absolute temperature).

This is the same as computing a transformed covariate.
$x = \frac{11604.83}{s+273.16}$.

Estimated model:

$$\ln T_s = -15.874 + 0.83072 \cdot \frac{11604.83}{s + 273.16} + \frac{1}{2.8246} W$$

PROBABILITY PLOTS



Normal temperature is 80-100C.

Experiment temperatures: 110, 130, 150, 170. Needs to extrapolate to 80-100, using Arrhenuis model.

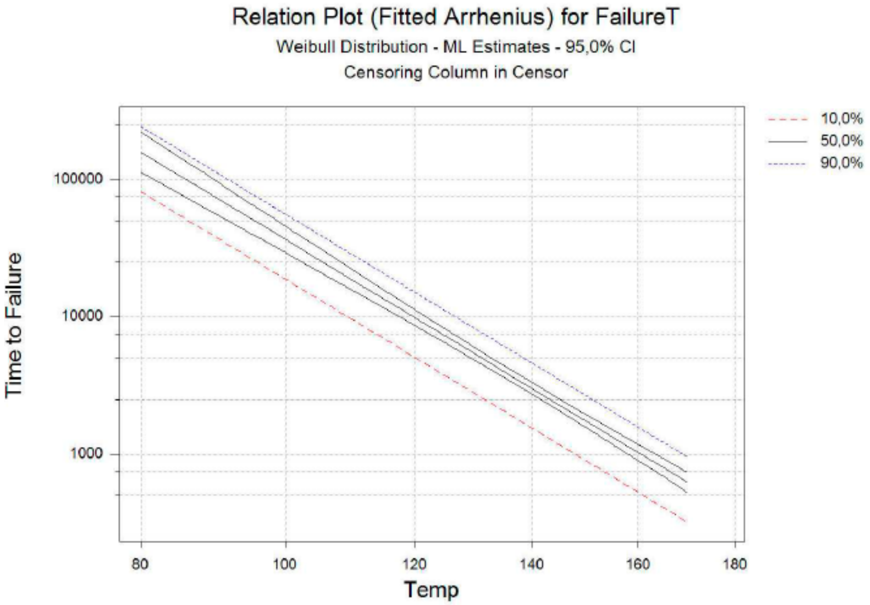
Recall probability plot for Weibull:

$\ln(-\ln R(t)) = \alpha \ln T - \alpha \ln \theta$

So:

- Slope α is the same for all lines
- Scale $\theta = \beta_0 + \beta_1 \cdot \frac{11604.83}{s+273.16}$ depends on experiment temperature s.

RELATION PLOT



EXPLANATION OF RELATION PLOT

Plot $\hat{t}_p(s)$ as function of s.

Recall general formula:

$$\ln \hat{t}_p(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}'\mathbf{x} + \sigma\Phi^{-1}(p)$$

where for Weibull/Gumbel we have $\Phi^{-1}(p) = \ln(-\ln(1-p))$. Here:

$$\ln \hat{t}_p(s) = -15.1874 + 0.83072 \cdot \frac{11.60483}{s + 273.16} + \frac{1}{2.8246} \cdot \ln(-\ln(1-p))$$

Figure shows median, $p = 0.50$, together with 95% confidence curves; and in addition the curves for $p = 0.10$ and $p = 0.90$.

ADDING THE FACTOR "PLANT"