TMA4275 LIFETIME ANALYSIS

Slides 13: More on Cox regression

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

i	Y_i	x_i	δ_i
1	5	12	0
2	10	10	1
3	40	3	0
4	80	5	0
5	120	3	1
6	400	4	1
7	600	1	0

Model:

- $z(t|x) = z_0(t)\exp\{\beta x\}$

Partial likelihood:

$$L(\beta) = \frac{e^{10\beta}}{e^{10\beta}+e^{3\beta}+e^{5\beta}+e^{3\beta}+e^{4\beta}+e^{\beta}} \cdot \frac{e^{3\beta}}{e^{3\beta}+e^{4\beta}+e^{\beta}} \cdot \frac{e^{4\beta}}{e^{4\beta}+e^{\beta}}$$

$$\hat{Z}_0(t) = \sum_{T_{(j)} \leq t} \frac{1}{\sum_{i \in R_j} e^{\hat{\boldsymbol{\beta}}' \mathbf{x}_i}}$$

so

$$\begin{aligned}
\hat{Z}_0(10) &= \frac{1}{e^{10\hat{\beta}} + e^{3\hat{\beta}} + e^{5\hat{\beta}} + e^{3\hat{\beta}} + e^{4\hat{\beta}} + e^{\hat{\beta}}} = 4.57 \cdot 10^{-4} \\
\hat{Z}_0(120) &= 4.57 \cdot 10^{-4} + \frac{1}{e^{3\hat{\beta}} + e^{4\hat{\beta}} + e^{\hat{\beta}}} = 0.0304 \\
\hat{Z}_0(400) &= 0.0304 + \frac{1}{e^{4\hat{\beta}} + e^{\hat{\beta}}} = 0.0730
\end{aligned}$$

$$\hat{R}(t; \mathbf{x}) = \exp\{-\hat{Z}_0(t) e^{\hat{\beta}' \mathbf{x}}\}$$

so

$$\begin{aligned} \hat{R}(10; x) &= \exp\{-4.57 \cdot 10^{-4} e^{0.765x}\} \\ \hat{R}(120; x) &= \exp\{-0.0304 e^{0.765x}\} \\ \hat{R}(400; x) &= \exp\{-0.0730 e^{0.765x}\} \end{aligned}$$

x	$\hat{R}(10; x)$	$\hat{R}(120; x)$	$\hat{R}(400; x)$
0*	0.9995	0.9701	0.9296
1	0.9990	0.9368	0.8548
3	0.9955	0.7396	0.4846
5	0.9793	0.2483	0.0352
10	0.3829	0.0000	0.0000
KM**	0.8333	0.5556	0.2778

* Baseline survival function

**KM-estimator does not use the value of x

MODEL CHECKING IN COX REGRESSION: COX-SNELL RESIDUALS

Cox-Snell Residuals (called "Generalized residuals" by Ansell & Phillips):

$$\hat{V}_i \equiv \hat{Z}_0(Y_i)e^{\hat{\beta}' x_i},$$

which should behave like a *censored set from expon(1)* if the model is correct.

Note: Sometimes is added 1 to the censored residuals in order to include them as "uncensored". The reason for this is that if $V \sim expon(1)$, then

$$E[V|V > y] = y + E(V) = y + 1$$

by the memoryless property of the exponential distribution.

$$\hat{V}_i = \hat{Z}_0(Y_i)e^{\hat{\beta}'\mathbf{x}_i}$$

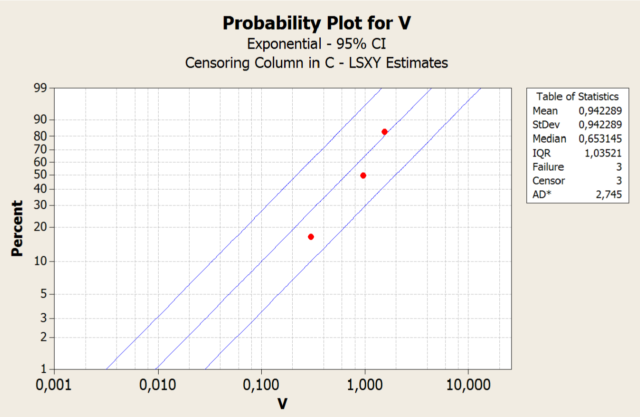
which should behave like censored data from *expon*(1)

i	Y_i	x_i	δ_i	$\hat{V}_i$	
1	5	12	0	$\hat{Z}_0(0)e^{0.765\cdot 12}$	$= 0.0000$
2	10	10	1	$\hat{Z}_0(10)e^{0.765\cdot 10}$	$= 0.9593$
3	40	3	0	$\hat{Z}_0(10)e^{0.765\cdot 3}$	$= 0.0045$
4	80	5	0	$\hat{Z}_0(10)e^{0.765\cdot 5}$	$= 0.0209$
5	120	3	1	$\hat{Z}_0(120)e^{0.765\cdot 3}$	$= 0.3017$
6	400	4	1	$\hat{Z}_0(400)e^{0.765\cdot 4}$	$= 1.5567$
7	600	1	0	$\hat{Z}_0(400)e^{0.765\cdot 1}$	$= 0.1569$

Worksheet 1 ***

↓	C1	C2	C3
	V	C	V*
1	0,9593	1	0,9593
2	0,0045	0	1,0045
3	0,0209	0	1,0209
4	0,3017	1	0,3017
5	1,5567	1	1,5567
6	0,1569	0	1,1569
7			

COX-SNELL RESIDUALS IN SIMPLE EXAMPLE: PROBABILITY PLOT OF CENSORED RESIDUALS



COX-SNELL RESIDUALS IN SIMPLE EXAMPLE: PROBABILITY PLOT WITH 1 ADDED TO CENSORED RESIDUALS

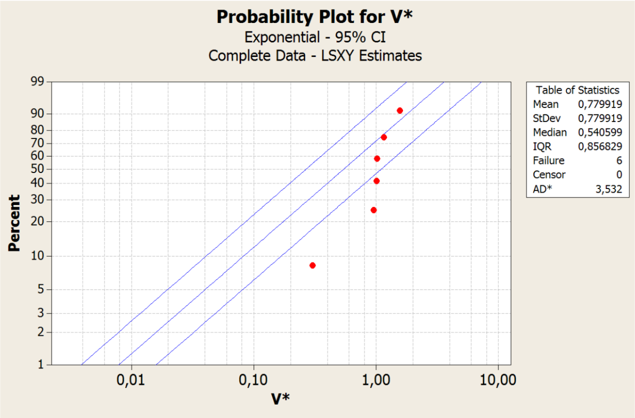


Table 3.2. Lifetimes (in cycles) of sodium sulphur batteries

Batch 1	164	164	218	230	263	467	538	639	669
	917	1148	1678+	1678+	1678+	1678+			
Batch 2	76	82	210	315	385	412	491	504	522
	646+	678	775	884	1131	1446	1824	1827	2248
	2385	3077							

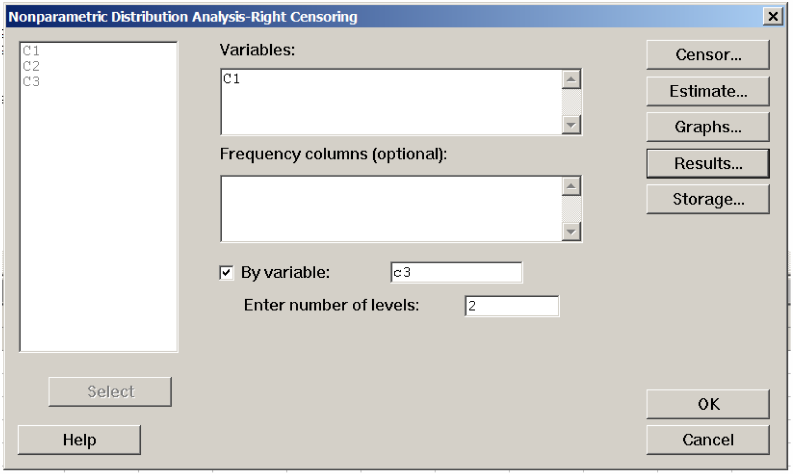
Note: Lifetimes with + are right censored observations, not failures.

C1=Obs. times, C2=censoring, C3="batch" no.

Ansell32.MTW ***

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
1	164	1	1									
2	164	1	1									
3	218	1	1									
4	230	1	1									
5	263	1	1									
6	467	1	1									
7	538	1	1									
8	639	1	1									
9	669	1	1									
10	917	1	1									
11	1148	1	1									
12	1678	0	1									
13	1678	0	1									
14	1678	0	1									
15	1678	0	1									
16	76	1	2									
17	82	1	2									
18	210	1	2									
19	315	1	2									
20	385	1	2									
21	412	1	2									
22	491	1	2									
23	504	1	2									
24	522	1	2									
25	646	0	2									

MINITAB-analysis of data from Ansell & Phillips:



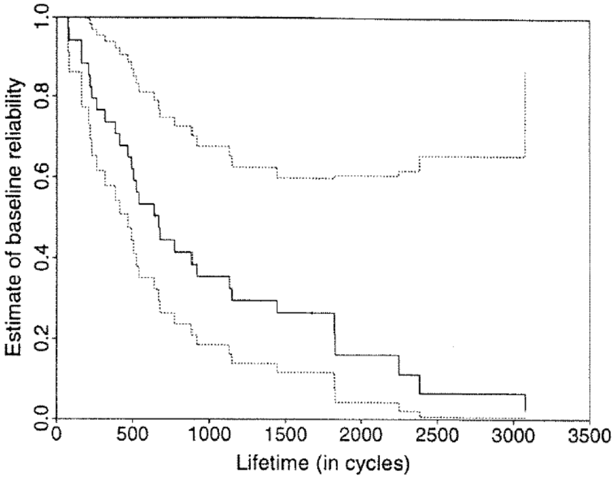


Fig. 3.3. Plot of the baseline reliability function for the proportional hazards model for the sodium sulphur battery data with 95% confidence limits

1 is added to the censored residuals

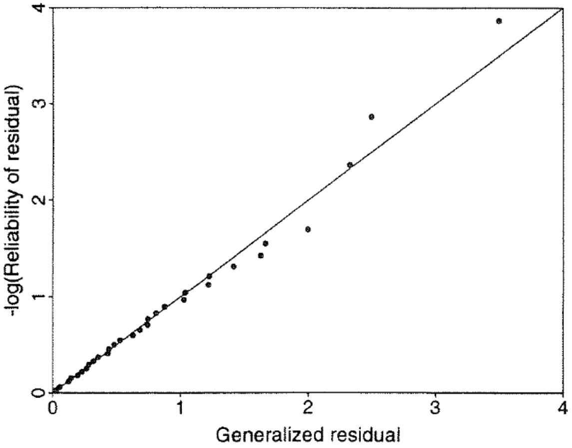


Fig. 3.5. Plot of the generalized residuals of the proportional hazards model for the sodium sulphur battery data

Recall the Cox-model: $z(t; \mathbf{x}) = z_0(t)e^{\boldsymbol{\beta}'\mathbf{x}}$

As we have seen, the effect of increasing, e.g., covariate number 1 by 1 unit, is to multiply the likelihood by e^{β_1}, independently of time t.

In practice one might imagine, however, that β_1 could depend on t, like $\beta_1(t)$; for example the risk of smoking could depend on the age, t, of a person, with $\beta_1(t)$ approaching 0 for high ages t.

The *Schoenfeld residual* (see 3.5.2 p. 77 in the book chapter on regression) compares, for each failure time $T_{(j)}$, the values of the covariates of the unit that fails, with what would be expected if the Cox-model with constant $\boldsymbol{\beta}$ is correct.

SCHOENFELD RESIDUALS FOR THE CASE OF A SINGLE COVARIATE

"...compares, for each failure time $T_{(j)}$, the values of the covariates of the unit that fails, with what would be expected if the Cox-model with constant β is correct."

For each failure time $T_{(j)}$, with unit ℓ_j failing and risk set R_j, we compute

$$
\begin{aligned}
s_j &= x_{\ell_j} - \sum_{i \in R_j} x_i P(\text{unit } i \text{ fails at } T_{(j)}) \\
&= x_{\ell_j} - \sum_{i \in R_j} x_i \frac{e^{\hat{\beta} x_i}}{\sum_{v \in R_j} e^{\hat{\beta} x_v}} \\
&= x_{\ell_j} - \frac{\sum_{i \in R_j} x_i e^{\hat{\beta} x_i}}{\sum_{i \in R_j} e^{\hat{\beta} x_i}}
\end{aligned}
$$

If the model is correct, the s_j are supposed to vary around 0.

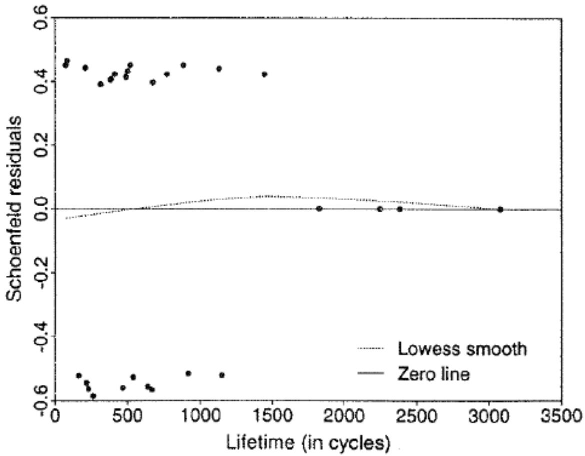


Fig. 3.7. Plot of the Schoenfeld residuals for batch of the proportional hazards model for the sodium sulphur battery data

$$s_j = x_{\ell_j} - \text{expected } x\text{-covariate of failing unit in } R_j$$

$$\begin{aligned}
s_1 &= 10 - \frac{10e^{10\hat{\beta}} + 3e^{3\hat{\beta}} + 5e^{5\hat{\beta}} + 3e^{3\hat{\beta}} + 4e^{4\hat{\beta}} + e^{\hat{\beta}}}{e^{10\hat{\beta}} + e^{3\hat{\beta}} + e^{5\hat{\beta}} + e^{3\hat{\beta}} + e^{4\hat{\beta}} + e^{\hat{\beta}}} \\
&= 10 - 9.7646 = 0.2354 \\
s_2 &= 3 - \frac{3e^{3\hat{\beta}} + 4e^{4\hat{\beta}} + e^{\hat{\beta}}}{e^{3\hat{\beta}} + e^{4\hat{\beta}} + e^{\hat{\beta}}} = 3 - 3.5099 = -0.5099 \\
s_3 &= 4 - \frac{4e^{4\hat{\beta}} + e^{\hat{\beta}}}{e^{4\hat{\beta}} + e^{\hat{\beta}}} = 4 - 3.7254 = 0.2756
\end{aligned}$$

The Schoenfeld residuals appear to be rather small, so there is apparently no reason to reject the model based on them.

See book chapter on regression, p. 63.

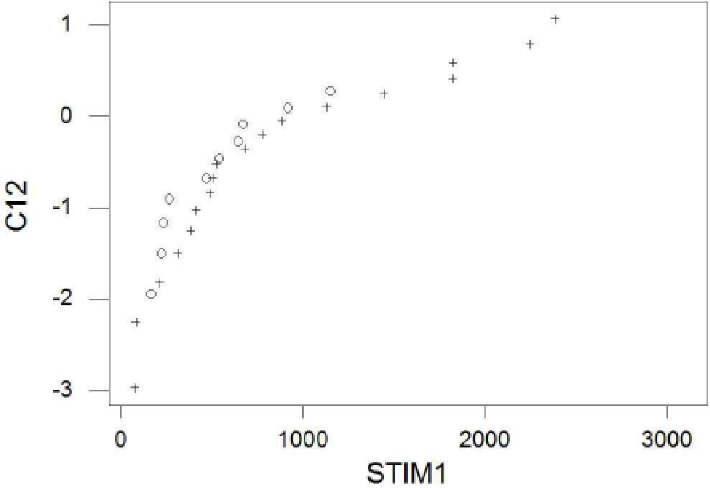
Model used in Battery example:

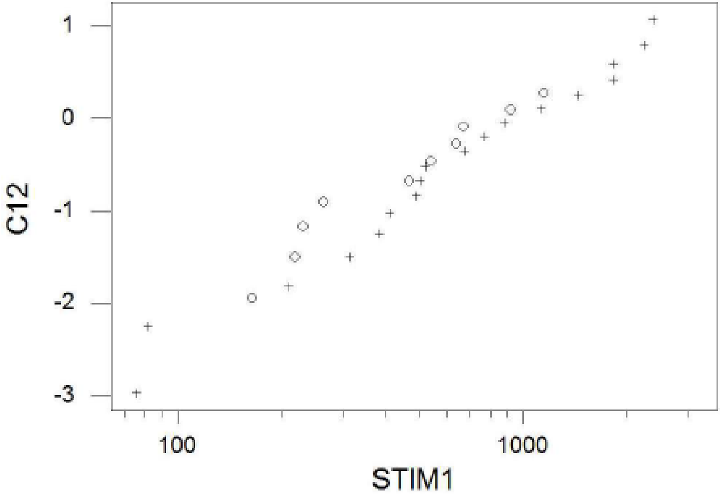
$$\text{Batch1}: \; z(t|0) = z_0(t)$$
$$\text{Batch2}: \; z(t|1) = z_0(t)e^{\beta}$$

Thus:

$$\text{Batch1:} \;\; R_1(t) = e^{-Z_0(t)} \Rightarrow \ln(-\ln R_1(t)) = \ln Z_0(t)$$
$$\text{Batch2:} \;\; R_2(t) = e^{-Z_0(t)e^{\beta}} \Rightarrow \ln(-\ln R_2(t)) = \ln Z_0(t) + \beta$$

- Thus if we compute KM-estimates $\hat{R}_{KM,1}$ and $\hat{R}_{KM,2}$ for each of the two batches, and plot $(t, \ln(-\ln \hat{R}_{KM,1}(t)))$ and $(t, \ln(-\ln \hat{R}_{KM,2}(t)))$, then the two "curves" will be in constant distance (theoretically equal to β) from each other.
- Often one plots instead $(\ln t, \ln(-\ln \hat{R}_{KM,1}(t)))$ and $(\ln t, \ln(-\ln \hat{R}_{KM,2}(t)))$, in which case straight lines will indicate Weibull distributions.


C12
1
0
-1
-2
-3
0
1000
2000
3000
STIM1


1
0
-1
-2
-3
C12
100
1000
STIM1