My learning notes of Survival Analysis

Yang Ge

2021 - January - 23, 12:31

Contents

Course name: Survival Analysis

Instructor: Xiao Song

Date: Fall, 2019

Textbooks:

- Required: Modeling Survival Data in Medical Research, Third Edition. David Collett. Chapman and Hall/CRC.
- Optional:

1. Applied Survival Analysis, Second Edition. David W Hosmer, Stanley Lemeshow and Susanne May. Wiley.
2. Modeling Survival Data: Extending the Cox Model. Terry M. Therneau and Patricia M. Grambsch. Springer.
3. Survival Analysis: Techniques for Censored and Truncated Data. John P. Klein and Melvin L. Moeschberger.

Software: SAS (For survival analysis, SAS start earlier and own much more functions than R)

1 Definition

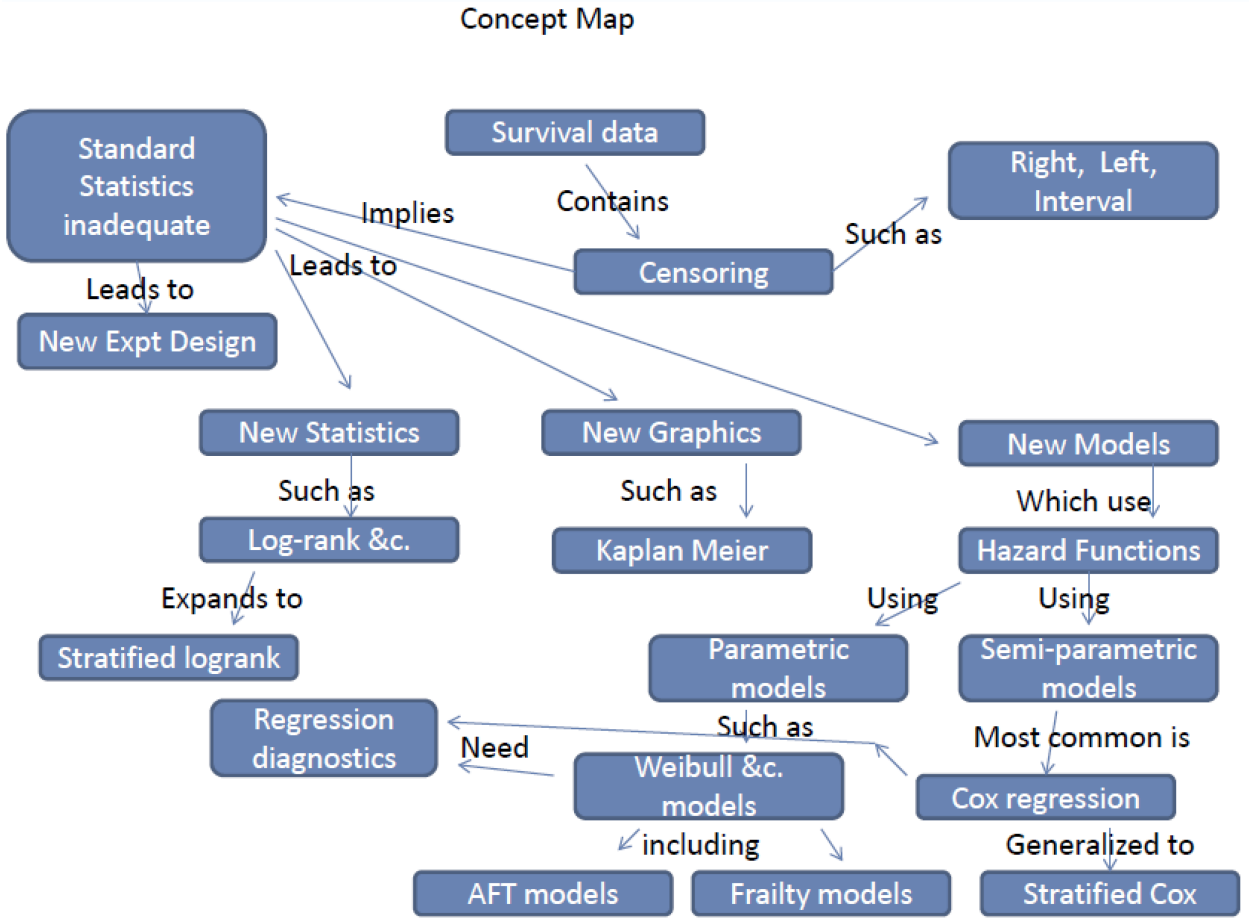


Figure 1: Basic survival analysis concept map

1.1 Censoring

An important prosperity of survival data is the occurrence of censored events. Those data censored similar as missing values may cause issues when tests were applied on mean or median, because we probably have no chance to calculate those quantities. E.g. If we have more than half participants censored, how we can get median value?

It is also wrong to just delete all censored data (missing value). Because that will produce biased results too! Therefore, we can see why survival analysis differs from other statistical analysis.

Definition:

Censoring (general): The survival time of an individual is said to be censored if the end-point of interest has not been observed for that individual (Collett, 2003).

Reasons for censoring include:

1. lost to follow-up (e.g., patient moved away).
2. withdraws from a study.
3. individual is still alive at study termination (e.g., in a study that follows patients for 5 years, any patient who lives longer than 5 years is right-censored).

4. individual's name is recorded in a death registry, but the exact time of death cannot be determined.

Types

1. Right censoring: A survival time is right censored if the individual is known to be alive at a specific time, say c, but survival status is not known for times t > c. Here, c is known as the censoring time.
2. Left censoring: An individual had event by a certain time, but the event time is not observed. Example: Event of interest is first menstrual period and study subject who enrolls at 10 years old has already experienced event.
3. Interval censoring: Observations are made at distinct times, with gaps in between. Example: Individuals contacted yearly for time-to-disease(e.g., HIV infected patients followed-up for time-to-AIDS).

Non-informative censoring

Censoring in survival analysis is non-informative if "knowledge of a censoring time for an individual provides no further information about the person's likelihood of survival at a future time had the individual continued on the study." (Klein and Moeschberg, 1997)

1.2 Truncation

Definition:

A data set of observations are truncated if it is incomplete due to a selection process inherent in the study design (Hosmer etal.). In contrast to censoring, the truncated cases are not observed at all(e.g. The selection process is typically part of the study design). Methods good for censoring are wrong for truncation which requires other approaches.

Types

1. Right truncation: occurs when the entire study population has already experienced the event of interest. Commonly length biased sampling and typically occurs when only individuals who have experienced an event are selected.
2. Left truncation: occurs when the subjects have been at risk before entering the study. Commonly delayed entry where there is an unknown delay between the start time and study entry time.

2 Probability

All proof will provided in later chapters.

2.1 Survival time T

1. T: Survival time (event happened time) for an individual is represented by the random variable T. Most parametric models assume T is continuous. Lower case letters refer to real numbers (e.g. t = 5 years).

$$T \in [0, +\infty)$$

2.2 $F(t)$

2. F: The Cumulative Distribution Function of T (event time).

$$\lim_{T \to +\infty} F(T) = 1$$

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

2.3 $S(t)$

3. S: The survival function.

$$S(t) = P(T > t), \text{ or } S(t) = P(T \geq t); \text{ because for continous variable both are same.}$$
$$S(t) = 1 - F(t)$$

2.4 $f(t)$

4. $f(t)$: Density Distribution Function of T

$$f(t) = \frac{d}{dt}F(T)_{T=t} = \lim_{h\to 0}\frac{P(t \leq T < t+h)}{h} = \frac{d}{dt}(\,1 - S(T)\,)_{T=t} = -\frac{d}{dt}S(T)_{T=t}$$

2.5 $h(t)$

5. Hazard Function

$$h(t) = \frac{f(t)}{S(t)} = -log'(S(t))$$

2.6 $H(t)$

6. Cumulative Hazard Function

$$H(t) = \int_0^t h(u)\,du = -log(S(t))$$

3 Kaplan Meier

The Kaplan Meier (a.k.a. **product limit**) method provides a dynamic **estimate of survival S(t)** that makes **no parametric** model assumptions.

3.1 No censoring

$$\hat{S}(t) = \frac{1}{n}\sum_{i=1}^{n} I(T_i > t)$$

3.2 Censoring

When censoring is present, the KM method develops $\hat{S}(t)$ from the conditional probabilities:

If no tied times

$$P(T > t_{(j)} + \Delta t \,|\, T > t_{(j)}) = 1 - P(T \leq t_{(j)} + \Delta t \,|\, T > t_{(j)}) = 1 - \frac{d_j}{n_j}$$

d_j = the number of individuals who die at time $t_{(j)}$

n_j = the number of individuals who are alive just before time $t_{(j)}$

If uncensored times are tied, then we assume the deaths occurred simultaneously at $t_{(j)}$

If censored and uncensored times are tied, then for the KM curve we assume the censored observations lived a little beyond $t_{(j)}$ contributing to the risk set.

3.3 Estimator

Kaplan-Meier estimate of $S(t)$

$$\hat{S}(t) = \begin{cases} \prod_{t_{(j)} \leq t} \frac{n_j - d_j}{n_j} & \text{if } t \geq t_{(1)} \\ 1 & \text{if } 0 \leq t \leq t_{(1)} \end{cases}$$

3.4 Assumptions

1. The survival data represents a random sample from the target population.
2. The censoring process is non-informative.

Example 2.1: Kaplan-Meier

Question 1: Severe viral hepatitis patients were entered into a 16 week study of the effects of steroid therapy. Patients were randomized to receive steroid or a control (standard of care) therapy. The survival time in weeks of the 14 patients on the steroid arm are given below.

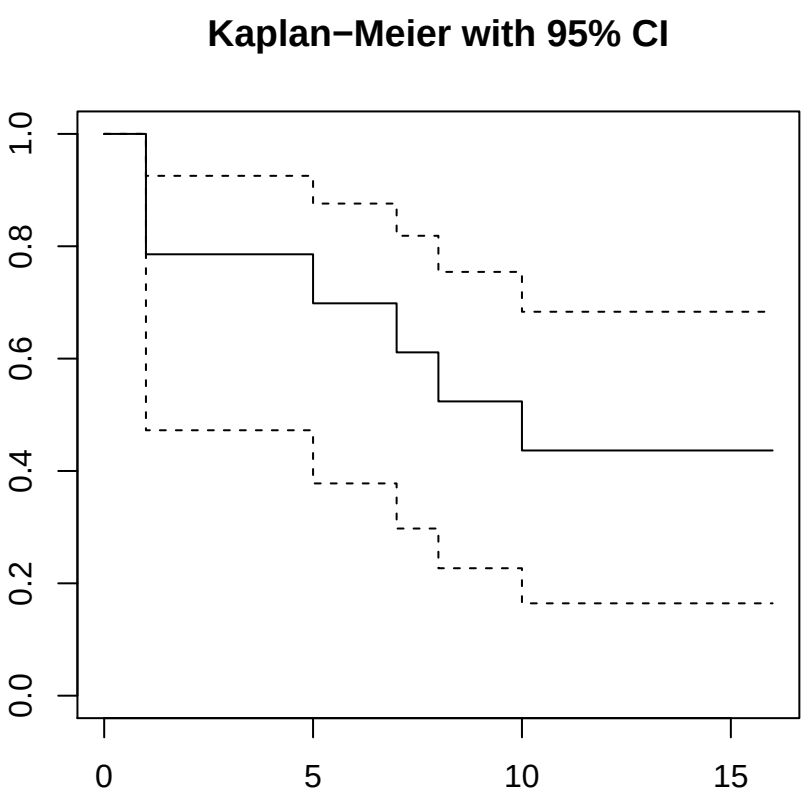
Survival times: 1, 1, 1, 1+, 4+, 5, 7, 8, 10, 10+, 12+, 16+, 16+, 16+

Note: A "+" indicates that the patient was still alive (censored).

```
library(survival)

tab1 <- round(data.frame(t = c(1, 5, 7, 8, 10), n = c(14, 9,
    8, 7, 6), d = c(3, 1, 1, 1, 1), f = c((14 - 3)/14, (9 - 1)/9,
    (8 - 1)/8, (7 - 1)/7, (6 - 1)/6), s = c((14 - 3)/14, ((14 -
    3)/14) * (9 - 1)/9, ((14 - 3)/14) * ((9 - 1)/9) * ((8 - 1)/8),
    ((14 - 3)/14) * ((9 - 1)/9) * ((8 - 1)/8) * ((7 - 1)/7),
    ((14 - 3)/14) * ((9 - 1)/9) * ((8 - 1)/8) * ((7 - 1)/7) *
        ((6 - 1)/6))), 3)

kable(tab1, row.names = FALSE, col.names = c("$t_i$", "$n_i$",
    "$d_i$", "$\\frac{n_i - d_i}{n_i}$", "$\\hat{S}_{(t_i)}$"),
    escape = FALSE)
```

t_i	n_i	d_i	$\frac{n_i - d_i}{n_i}$	$\hat{S}_{(t_i)}$
1	14	3	0.786	0.786
5	9	1	0.889	0.698
7	8	1	0.875	0.611
8	7	1	0.857	0.524
10	6	1	0.833	0.437

```
dat <- data.frame(time = c(1, 1, 1, 1, 4, 5, 7, 8, 10, 10, 12,
    16, 16, 16), censor = c(1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0,
    0, 0, 0))

myfit <- survfit(Surv(time, censor) ~ 1, conf.type = "log-log",
    error = "greenwood", data = dat)
plot(myfit, main = "Kaplan-Meier with 95% CI", sub = "Greenwood formula with S")
```



3.5 Life table

Unlike the KM estimator, the life table (actuarial) method was developed for large retrospective studies where numbers of those alive at the start of an interval, and dying and censored in the interval, are available, but not individual death times.

Notation

1. Let $\tilde{t}_{(1)} < ... < \tilde{t}_{(n)}$ be time-points (not necessarily equally spaced, e.g., 1, 2, 3, ..., 10 years).
2. Let n_i be the number alive at the beginning of interval with right endpoint $\tilde{t}_{(i)}$.
3. Let c_i be the number of censored observations in interval with right endpoint $\tilde{t}_{(i)}$.
4. Then $n_i' = n_i - \frac{c_i}{2}$ is the effective sample size, assuming censoring uniform distributed between the interval.
5. $\tilde{p}_{(i)} = \frac{n_i' - d_i}{n_i'}$ is the estimated probability of surviving the interval.

$$\hat{S}_a = \prod_{i=1}^{k} \tilde{p}_{(i)}$$

3.6 Probability

Suppose we divide the time in small intervals, and assuming no censoring (e.g. all individuals die).

Notation

1. n_i: the number at risk at the interval start from time point of i.
2. d_i: the number who die in the interval start from time point of i.
3. $s_i = n_i - d_i$: is the number who survive the interval start from time point of i.
4. $h_i = \frac{d_i}{n_i}$: is the probability of dying in the interval start from time point of i.

Distribution

The distribution of survive $(n_i - d_i)$ is **Binomial** with each **Bernoulli** event probability (1 - h_i), and trials n_i, assuming each death is independent.

For the Binomial distribution:

$$E[n_i - d_i] = n_i(1 - h_i)$$

$$E[\frac{n_i - d_i}{n_i}] = 1 - h_i$$

$$Var[n_i - d_i] = n_i(1 - h_i)h_i$$

$$Var[\frac{n_i - d_i}{n_i}] = \frac{(1 - h_i)h_i}{n_i}$$

Therefore, this gives us a way to estimate the mean and variance for that interval.

$$\hat{S}_{(t)} = \begin{cases} 1 & \text{if } t < t_{(1)} \\ \prod_{i=1}^{k} \frac{n_i - d_i}{n_i} & \text{if } t \geq t_{(1)} \end{cases}$$

3.7 Confidence interval

Now, for $\hat{S}_{(t)}$ we have a point estimation, next we need a estimation of the variance and standard error for confidence interval.

1. Greenwood's formula: By the delta method with transformation of $Ln(\hat{S}_{(t)})$, and assuming i.i.d $(t_i \neq t_j,\ cov(n_i - d_i, n_j - d_j) = 0)$.

$$\hat{V}ar[\hat{S}_{(t)}] = [\hat{S}_{(t)}]^2 \sum_{t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

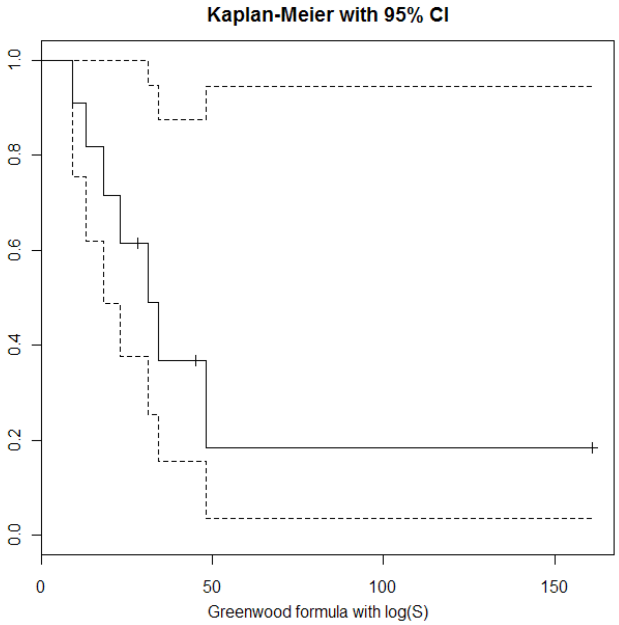


Figure 2: Kaplan Meier with Greenwood's 95%CI

Notes, The confidence intervals are:

1. Calculated based on log-transformation.
2. Truncated at 1 and 0.
3. Point-wise and not simultaneous.

Because, The log-survival ($Ln(\hat{S}_{(t)})$) has range $(-\infty, 0]$. However, ideally, since the CI is based on $\hat{S}_{(t)} \pm 1.96 * s.e.$, the range should be $(-\infty, +\infty)$, the same as a normal random variable. To achieve this, use the transformation $log(-log())$ transformation

2. log-log formula

By this transformation, $loglog(s) = log(-log(\hat{S}_{(t)})) \in (-\infty, +\infty)$, we can have a better CI estimation.

$$\widehat{Var}[loglog(s)] = \frac{1}{\left[ln(\hat{S}_{(t)})\right]^2} \sum_{t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

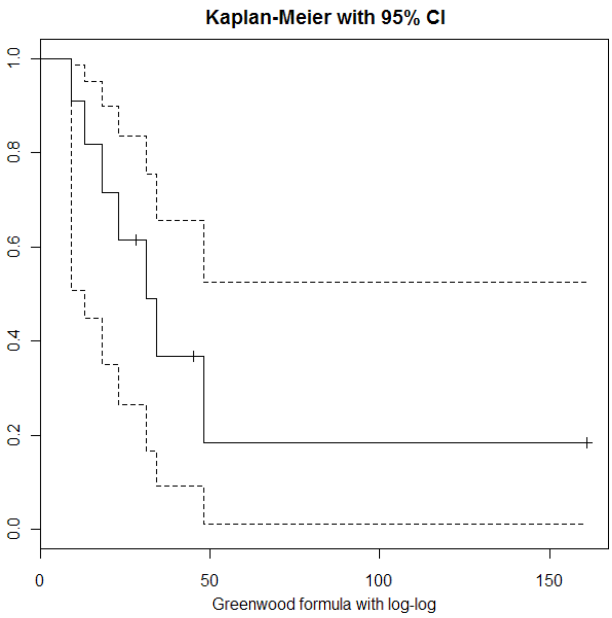


Figure 3: Kaplan Meier with log-log 95%CI

Here, the log-log method provides narrower interval that does not hit the upper bound as much.

Summary:

1. Greenwood's formula actually is an application of delta method to getting variance from a exponential transformation.
2. $s.e. = \sqrt{Var}$, because here is the estimation of a parameter from itself variance different to the s.e. estimation of average from i.i.d samples.

3.8 Codes of SAS

The "(S)" stands for estimated survival function.

```
PROC LIFETEST DATA=Q1dat PLOT=(S) METHOD=KM;
  TIME Time*Survival_status(0);
  STRATA AssayResult;
RUN;
```

Figure 4: Codes of SAS, KM estimate

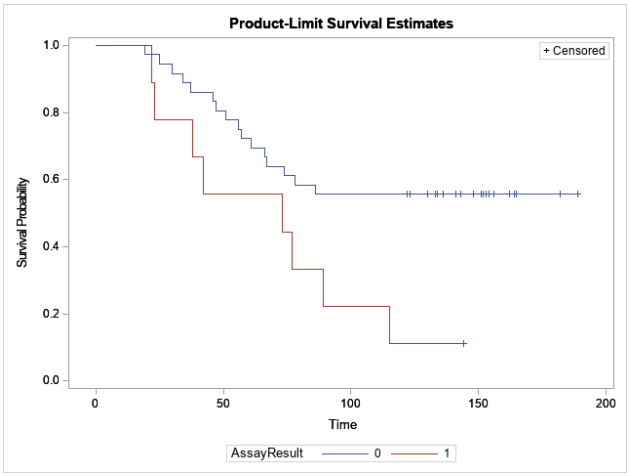


Figure 5: KM estimate results

3.9 CI band

Simultaneous vs. point-wise CI

The CI curves of $\hat{S}_{(t)} \pm 1.96 * s.e.$: just reflect a collection of CI's constructed at each event (death) time. No multiple comparison adjustment. The overall coverage across all the event times is below 95%. Simultaneous 95% CI's are constructed in such a way that the overall type I error is controlled at 5%. **Point-wise is narrower than simultaneous CI**

Simultaneous confidence bands

A 95% simultaneous confidence band for death time t_0 is a band constructed in such a way that, if the band was repeatedly constructed, 95% of the time it would contain the ENTIRE survival function up to time t_0.

Formula

1. By central limit theorem, for any t, $\sqrt{n}\Big(\hat{F}(t) - F(t)\Big) \to N(0, \sigma^2)$.
2. Also, under certain conditions $w(t)$ converges to a Wiener process.

$$w(t) = \sqrt{n}\Big(\frac{\hat{F}(t) - F(t)}{1 - \hat{F}(t)}\Big)$$

3. Wiener processes (a.k.a. Brownian motion) are common in finance and other applications; theory can be used to construct the interval.

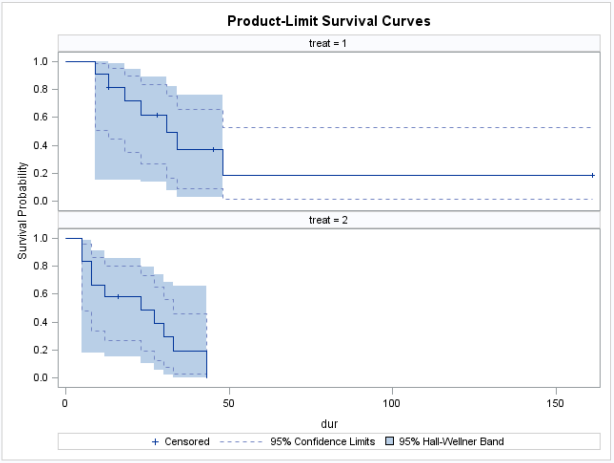


Figure 6: Simultaneous confidence bands

Notes:

1. The bands are wider than the point-wise confidence limits.
2. In R, the package "km.ci" can be used to construct simultaneous confidence bands.

4 Hazard

Density $f(t)$ is needed to measure the event possibility:

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t)}{\Delta t}$$

Hazard Function

Hazard $h(t)$ is needed to measure the instantaneous death rate conditional on current risk population:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t | T \ge t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t)}{P(T \ge t)} * \frac{1}{\Delta t} = \frac{f(t)}{S(t)}$$

$$\because \frac{f(t)}{S(t)} = \frac{F'(t)}{S(t)} = \frac{[1 - S(t)]'}{S(t)} = -\frac{S'(t)}{S(t)} = -log'(S(t))$$

$$\therefore h(t) = -log'(S(t))$$

Note: If survival time T is exponential distribution, $h(t)$ is constant to λ

Cumulative Hazard Function

$$H(t) = \int_0^t h(u)\, du = -log(S(t))$$

$$\because H(t) = \int_0^t h(u)\, du =$$

$$= \int_0^t -log'(S(u))\,du$$

$$= -\int_0^t log'(S(u))\,du$$

$$= -log(S(u))\Big|_0^t$$

$$= -log(S(t)) - (-log(S(0)))$$

$$= -log(S(t)) - (-log(1))$$

$$= -log(S(t))$$

4.1 Nelson Aalen estimation

Estimating the hazard function is analogous to estimating a density function requiring lots of data. But the cumulative hazard function can be estimated, even with small data sets, and is useful.

Nelson Aalen cumulative hazard estimate

At death time $t(i)$, the hazard for those alive is estimated by $\frac{d_i}{n_i}$, so the cumulative hazard "jumps up" at those times.

$$\tilde{H}_{(t)} = \sum_{t(i)\le t} \frac{d_i}{n_i}$$

Nelson example

Toy Example with out of battery time: 1, 2, 2, 4+, 4, 5

Table 3.1

Time t	0	1	2	4	5
n	6	6	5	3	1
d	0	1	2	1	1
d/n	0	1/6	2/5	1/3	1/1
$\tilde{H}_{(t)}$	0	1/6	1/6 + 2/5	1/6 + 2/5 + 1/3	1/6 + 2/5 + 1/3 + 1/1

Note: the cumulative hazard can exceed 1.

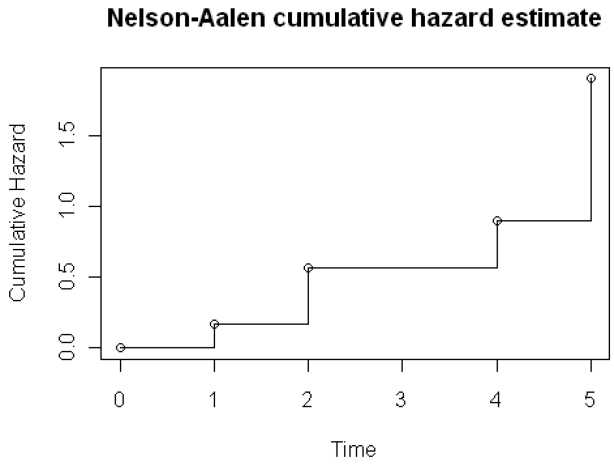


Figure 7: Simultaneous confidence bands

Variance of Nelson-Aalen Estimator

$$\widehat{Var}(\tilde{H}_{(t)}) = \sum_{t_{(i)} \leq t} \frac{d_i}{n_i^2}$$

4.2 Kaplan Meier estimation

Because, previously, we already proof that:

$$\hat{H}(t) = \int_0^t h(u)\, du = -log(S(t))$$

Then, due to the ability of $S(t)$ estimation from Kaplan Meier, we have another way to get cumulative hazard $H(t)$.

$$\hat{H}(t) = -log(S(t)) = -log\Big(\prod_{t_{(j)} \leq t} \frac{n_j - d_j}{n_j} \Big) = - \sum_{t(i) \leq t} log(\frac{n_i - d_i}{n_i})$$

4.3 Nelson Aalen VS Kaplan Meier estimation

1. Kaplan Meier is the standard graphic used for the survival function.
2. The Nelson-Aalen is the standard graphic used for the cumulative hazard.
3. Nelson-Aalen is more "optimistic" ($H_{(Nelson\ Aalen)} \leq H_{(Kaplan\ Meier)}$ or $S_{(Nelson\ Aalen)} \geq S_{(Kaplan\ Meier)}$).

Proof

$$\because S_{(Nelson\ Aalen)} = e^{-\tilde{H}(t)}$$

$$\because e^{-\tilde{H}(t)} = e^{\left(- \sum_{t(i) \leq t} \frac{d_i}{n_i} \right)} = \prod_{t_{(j)} \leq t} e^{-\frac{d_i}{n_i}}$$

$$\because S_{(Kaplan\ Meier)} = \prod_{t_{(j)} \le t} \frac{n_j - d_j}{n_j} = \prod_{t_{(j)} \le t} 1 - \frac{d_j}{n_j}$$

From the The Taylor series expansion for the function e^x is given by

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots = \sum_{n \ge 0} \frac{x^n}{n!}$$

We know that, actually, $S_{(Kaplan\ Meier)}$ is the **first two terms** of **Taylor series** expansion of $S_{(Nelson\ Aalen)}$

$$\therefore S_{(Nelson\ Aalen)} \ge S_{(Kaplan\ Meier)}$$

Example 2.1: Kaplan-Meier

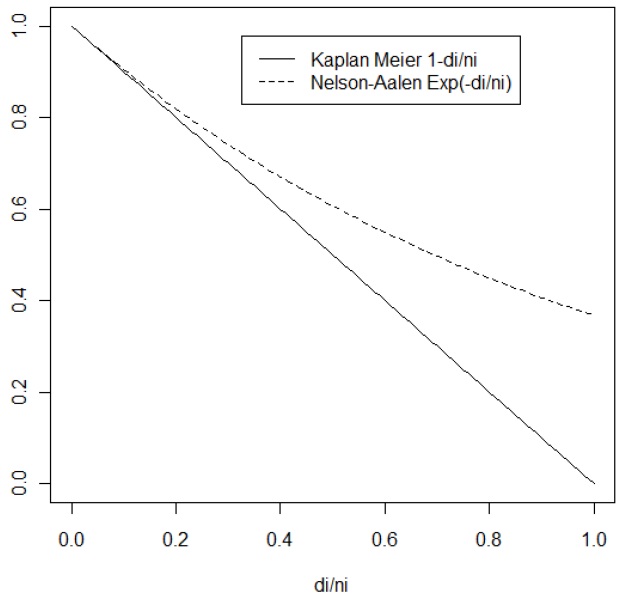


Figure 8: Nelson-Aalen is more optimistic

```
data methadone;
   input ID clinic status time prison dose;
   datalines;
1     1     1     428    0     50
2     1     1     275    1     55
...
266   1     1     47     0     45

;

proc lifetest data=methadone plots=(logsurv) nelson;
   time time*status(0);
   strata clinic;
run;
```

Figure 9: SAS codes of Nelson-Aalen

Notes

1. logsurv: plot the cumulative hazard function (I already proofed above).
2. nelson: get the Nelson-Aalen estimate.

5 Non-parametric model

People love statistical test, because we need a important quantity even always around critics, P-value!

Non-parametric models (e.g., product limit estimator for survival) make no distributional assumptions (e.g., about the survival times, or the hazard function).

5.1 The log-rank test

The p-values reported on Kaplan Meier plots are typically based on the log-rank test.

Assumptions

1. The survival data represents a random sample from the target population.
2. The censoring process is non-informative.

Hypotheses

$$H_0 : S_1(t) = S_2(t), \ \ for\ all\ t > 0$$

$$H_1 : S_1(t) \neq S_2(t), \ \ for\ all\ t > 0$$

The alternative may seem sub-optimal, since it give no hints on the direction of the difference over time. Therefore, the log-rank test has good power only when the direction stays the same.

Justification

If the people are labeled, for example as treated or untreated, and the labels are independent of survival, then the distribution of the number treated who die (d_i) at time $t(i)$ is hypergeometric.

This provides a **null distribution**.

Computation

Table 4.1 for interval i

	Group A	Group B	Total
Dead	d_{Ai}	d_{Bi}	d_i
Alive	$n_{Ai} - d_{Ai}$	$n_{Bi} - d_{Bi}$	$n_i - d_i$
Total	n_{Ai}	n_{Bi}	n_i

By linearity of expectation, the expectation of death (d) for the group A (not matter choose death or live for group A or B, because the degree of freedom here is only 1):

$$E\Big[\sum_{i=1}^{k} \widehat{D_{Ai}}\Big] = \sum_{i=1}^{k} \frac{d_i}{n_i} n_{Ai}$$

The test statistic is approximately chi-square with 1 degree of freedom under $H_0 : S_1(t) \equiv S_2(t)$. Reject H_0 in favor of $H_1 : S_1(t) \not\equiv S_2(t)$ at α if the test statistic is larger than $\chi_{1,1-\alpha}$ (the $(1-\alpha)$ 100th percentile of χ^2 distribution).

Observed: $O = \sum_{i=1}^{k} d_{Ai}$

Expected: $E = \sum_{i=1}^{k} \frac{d_i}{n_i} n_{Ai}$

Variance: $Var(O) = \sum_{i=1}^{k} \frac{n_{Ai}\, d_i}{n_i} \frac{(n_i - d_i)(n_i - n_{Ai})}{n_i(n_i - 1)}$

Statistics: $\chi = \frac{\left(|O-E|\right)^2}{Var(O)}$

Example 4.1

The data:

Table 4.2

Time (Day) in Group A	Status	Time (Day) in Group B	Status
28	1	2	1
32	1	4	1
49	1	72	1
84	1	77	1
357	1	79	1
933, 1078, 1183, 1560, 2114, 2144	0	n.a.	n.a.

The Log-rank statistics table

Table 4.3

t_i	d_{Ai}	n_{Ai}	d_i	n_i	$E[d_{Ai}]$	$Var[d_{Ai}]$
2	0	11	1	16	(1*11)/16 = 0.6875	0.215
4	0	11	1	15	0.733	0.196
28	1	11	1	14	0.786	0.168
32	1	10	1	13	0.769	0.178
49	1	9	1	12	0.750	0.188
72	0	8	1	11	0.727	0.198
77	0	8	1	10	0.800	0.160
79	0	8	1	9	0.889	0.098
84	1	8	1	8	1.000	0
357	1	7	1	7	1.000	0
Total	5				8.142	1.401

$\chi^2_{n-1=1} = \frac{\left(|5-8.142|\right)^2}{1.401} = 7.05$, Log-rank test (a.k.a score test) is 7.05, $d.f. = 1$, $p = 0.0079$

5.2 Alternative to Log-rank

Alternatives to the Log-rank test generally just differ in the weights assigned to the 2-by-2 interval tables, similar to add a column of weight in table 4.3.

With weight:

1. Wilcoxon: weights each table by n_i, the number at risk. Earlier tables get greater weight due to larger participants. It is most likely to detect early differences.
2. Tarrone and Ware (1977) weight each table by $\sqrt{n_i}$. It is also more likely to detect early differences than the log-rank test, but less likely than Wilcoxon.

3. Other weights: Peto-Prentice weight, Modified Peto-Peto weight, Harrington-Fleming(p,q) weight, etc..

Without weight: Just as Log-rank weights the table at each death time equally. Relative to Wilcoxon and Tarrone, the log-rank test is more likely to detect later differences in survival.

5.3 Log-rank test > 2 groups

Extended the log-rank test with > 2 groups

The log-rank test provides an overall p-value of whether there are any differences. Analogous to the ANOVA F-test with K groups

Hypotheses

$$H_0 : S_1(t) = S_2(t) = ... = S_k(t), \quad for\ all\ t > 0$$

$$H_1 : S_i(t) \neq S_j(t),\ forsomei \neq jandt > 0$$

Similar to Table 4.1 but with more columns for more groups, so the $d.f. = K - 1$. Correspondingly, with more complicated formulas for $E[S]$, $Var[S]$ and χ^2 (please check the textbook).

Just as with the 2-sample log-rank test, weights can be added to the extended log-rank test to produce weighted log-rank test (Wilcoxon, Tarrone-Ware, etc.).

Example 4.2:

SAS "strata" command tells SAS to treat the variable as a group indicator.

```
data lungcancer;
   input cell dur status;
   datalines;
1 72 1
1 411 1
1 228 1
1 126 1
1 118 1
...
4 49 1
;

proc lifetest data=lungcancer plots=(s);
   time dur*status(0);
   strata cell/test=(logrank wilcoxon tarone);
run;
```

Figure 10: Codes of SAS, Extended Log-rank test

Summary of the Number of Censored and Uncensored Values					
Stratum	cell	Total	Failed	Censored	Percent Censored
1	1	35	31	4	11.43
2	2	48	45	3	6.25
3	3	27	26	1	3.70
4	4	27	26	1	3.70
Total		137	128	9	6.57

Figure 11: Data table, Extended Log-rank test

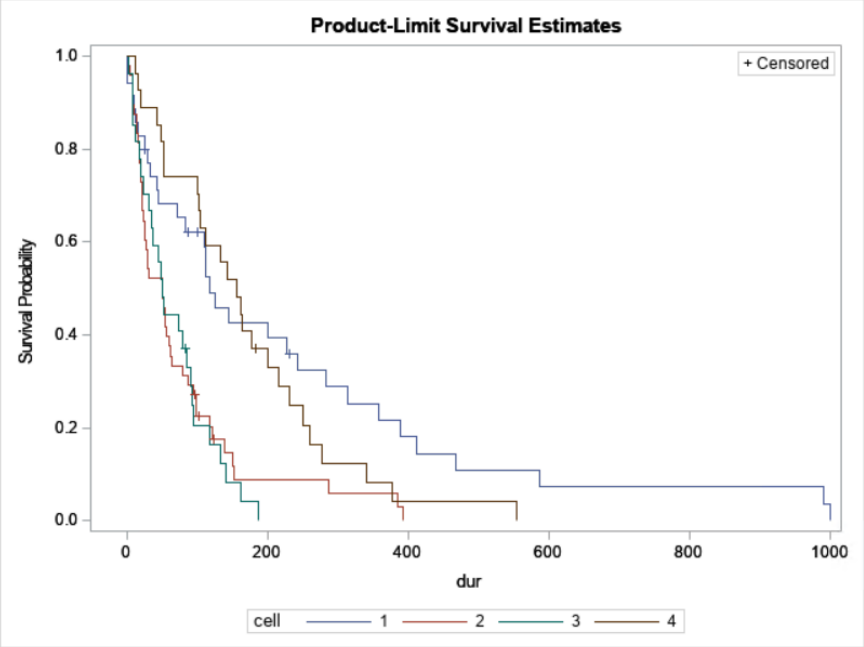


Figure 12: Kaplan Meier curves, Extended Log-rank test

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	25.4037	3	<.0001
Wilcoxon	19.4331	3	0.0002
Tarone	22.5728	3	<.0001

Figure 13: Test result, Extended Log-rank test

Conclusion: There are some differences in survival between the 4 groups. Further study would be needed

to determine where the differences lie.

5.4 Log-rank test for trend

The log-rank test for trend (ordered groups)

If the groups are ordered, then the data are ordinal instead of categorical. The simple log-rank test ignores the ordering, which may result in a loss of power.

Hypotheses

$$H_1 : S_1(t) < S_2(t) < ... < S_k(t)$$

Note:

1. As in trend tests for 2xK tables, the "distances" between the groups must be specified to do the analysis. These "distances" may not be straightforward to figure out.
2. The data will similar to analyze continuous variable, $d.f. = 1$.

Example 4.3:

SAS "strata" command tells SAS to treat the variable as a group indicator.

```
PROC LIFETEST DATA=dat PLOT=(S);
  TIME Time*Survival_status(0);
  STRATA  GROUP / TREND TEST=(logrank wilcoxon tarone peto);
RUN;
```

Figure 14: Codes of SAS, trend test

5.5 Log-rank test for stratification

Motivation: In two-group comparisons, heterogeneity of patients (at baseline) can impact the analysis like a confounder. E.g. if we compared one treatment in dying and healthy people, the results may different with no relationship to the intervention.

Hypotheses:

Suppose there are J strata and 2 groups, with survival functions $S_{1,j}$ and $S_{2,j}$, $j = 1, ..., J$. It assumes that the effect of group or treatment is same in each specific strata (like two parallel curves).

$$H_0 : S_{1,j}(t) = S_{2,j}(t), for j = 1, ..., J and all t$$

$$H_1 : S_{1,j}(t) \neq S_{2,j}(t), for some j and some t$$

Notes: If the effect of group is different across strata (like a modifier or interaction term), one may have little power to detect this.

Example 4.3:

SAS "strata" command tells SAS to treat the variable as an confounder (SAS will control on it), "group" command tells SAS to treat the variable as a group indicator (SAS will test its effect).

```
PROC LIFETEST DATA=Q1dat PLOT=(S) METHOD=KM;
  TIME t2death*death(0);
  STRATA race / GROUP = gender;
RUN;
```

Figure 15: Codes of SAS, trend test

Summary:

In general, the stratified log-rank test is only slightly less powerful than the log-rank test, because:

1. It may be that the noise eliminated by stratification actually makes the stratified log-rank test more powerful than the log-rank!
2. Even if there is no stratification effect, using a stratified analysis costs little in terms of power loss.
3. The only times a stratified log-rank test can be a very bad idea are when: (1) substantial heterogeneity of treatment effect across strata, or (2) there are very many strata.
4. Rule of thumb: The number of stratification variables should be kept to at most 2 or 3.

6 Semi-parametric model

A semi-parametric model has parametric and non-parametric components.

When comparing two groups, the hazards may be different, reflecting greater risk in one group.

Hazard Ratio: The hazard ratio is the instantaneous relative risk between the two groups.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{P(T \geq t)} * \frac{1}{\Delta t} = \frac{f(t)}{S(t)}$$

$$HR(t_i) = \frac{h_{At_i}}{h_{Bt_i}}$$

6.1 Cox model

The Cox proportional hazards model describes the HR in terms of a single regression parameter.

Parametric component: In Cox regression, the **ratio** between the hazards is the parametric component. The parametric component assumes the ratio does not change over time.

non-parametric component: The non-parametric component of Cox regression is the (baseline) hazard function. The baseline hazard can have any shape.

The univariate Cox (a.k.a. proportional hazards) regression model is:

$$h(t) = h_0(t)e^{\beta X}$$

where β is a regression coefficient vector, X is a covariate vector, and $h_0(t)$ is the baseline hazard function. The baseline hazard is the hazard function when the covariate is zero. We also **restrict** to covariates that are measured **ONLY ONCE** per subject (not multiple measurements over time).

Cox regression for 2 groups

Suppose x = 0 for a placebo group, and x = 1 for a drug treatment group.

For the placebo group:

$$h_p(t|x) = h_0(t)e^{\beta X}$$

For the treatment group:

$$h_t(t|x) = h_0(t)e^{\beta X}$$

The hazard ratio of treatment group vs placebo group is:

$$HR(t) = \frac{h_0(t)e^{\beta X_t}}{h_0(t)e^{\beta X_p}} = e^{\beta_t - \beta_p}$$

The proportion change in the hazard associated with moving from group treatment to group placebo.

$$HR_{proportion}(t) = \frac{h_0(t)e^{\beta X_t} - h_0(t)e^{\beta X_p}}{h_0(t)e^{\beta X_p}} = e^{(\beta_t - \beta_p)} - 1$$

Rescaling: If X is a continuous covariate, rescaling mean $e^{\beta X_p} \to e^{\beta \frac{X_p}{c}}$. It will impact the slope of the Cox regression model. But, the model is fundamentally the same.

Linear: If true relationship is $y = \sqrt{x}$, or $y = x^2$, etc., then the Cox model does not hold for y.

$$log\Big(h_t(t|x)\Big) = log\Big(h_0(t)\Big) + \beta X$$

Note: Nonlinear transformations of x lead to a different regression model may better fitting Cox regression model.

6.2 Estimation

The primary objectives of studies that utilize the Cox proportional hazards model are to:

1. Estimate β and its standard error.
2. Perform statistical inference on β, such as test the hypothesis $H_0 : \beta = 0$ vs $H_a : \beta \neq 0$, or construct a confidence interval for β.

Partial likelihood:

Suppose $\mathcal{R}_j$ is the risk set at time $t_{(j)}$. Suppose one individual dies at time $t_{(j)}$. The probability individual $i_0 \in \mathcal{R}_j$ is the one who dies is:

$$P(d_i = i_0 | i_0 \in \mathcal{R}_j) = \frac{h_0(t)e^{\beta X_{i_0}}}{\sum\limits_{i \in \mathcal{R}_j} h_0(t)e^{\beta X_i}} = \frac{e^{\beta X_{i_0}}}{\sum\limits_{i \in \mathcal{R}_j} e^{\beta X_i}}$$

These conditional probabilities can be calculated for each individual that dies at each failure time in a data set. Multiplying the probabilities together results in something similar to a likelihood, called the **partial likelihood**(Because no censoring considered in it).

Estimate β:

Assume failures occur at distinct times $t_{(1)} < ... < t_{(m)}$, and denote $\mathcal{R}_j$ as the risk set at time t_j. Let $x_{(j)}$ be the covariate value for the individual failing at time $t_{(j)}$. The **partial likelihood** is:

$$PL(\beta|p_{t_1}, p_{t_2}, ...p_{t_k}) = \prod_j \frac{e^{\beta X_j}}{\sum\limits_{j \in \mathcal{R}_j} e^{\beta X_j}}$$

The partial likelihood is maximized like a likelihood function, resulting in an estimate $\hat{\beta}$.

Note:

1. Only event time counted in the $PL(\beta|p_{t_1}, p_{t_2}, ...p_{t_k})$
2. Why can one ignore the times between deaths in the calculation? The argument is that there can be no information about the value of β contained in these intervals because $h_0(t)$ might be zero at these times.
3. Any intercept term would overparametrize the model, so no intercept is included. Because $h_0(t)e^{(\beta_0+\beta X)}$ or $h_0'(t)e^{\beta X}$ is indistinguishable through fitting.
4. If there is more than one death at a death time then the sampling is not multinomial (Multinomial sampling would allow the same person to die more than once). This sampling corresponds to the noncentral multivariate hypergeometric distribution, an extension of the hypergeometric distribution discussed in relation to the log-rank test.

Example 5.1: Toy example

Table 5.1

Time	Status	X variable
1	1	1
2	1	2
3	1	3

Because only one event in any interval:

$$PL(\beta|p_{t_1}, p_{t_2}, p_{t_3}) = \frac{e^{\beta X_1}}{e^{\beta X_1} + e^{\beta X_2} + e^{\beta X_3}} \cdot \frac{e^{\beta X_2}}{e^{\beta X_2} + e^{\beta X_3}} \cdot \frac{e^{\beta X_3}}{e^{\beta X_3}}$$

$$PL(\beta|p_{t_1}, p_{t_2}, p_{t_3}) = \frac{e^{\beta}}{e^{\beta} + e^{3\beta} + e^{2\beta}} \cdot \frac{e^{3\beta}}{e^{3\beta} + e^{2\beta}} \cdot \frac{e^{2\beta}}{e^{2\beta}}$$

This function can be maximized by taking the derivative with respect to β, setting to zero, and solving for β numerically.

6.3 Tied event time

Tied event time is that: there is more than one death at a death time. There are three methods for handling tied death times in Cox regression. In most practical settings, the three produce very similar results:

1. Exact method (This is the best one to use.)
2. Efron method (This is faster but not as good.)
3. Breslow method (This is fast, but the worst.)

Motivation:

If person A and B died at the same interval, the problem of tied events is:

$$PL_1(\beta|p_{t_1}, p_{t_2}, p_{t_3}) = \frac{e^{A\beta}}{e^{A\beta}+e^{B\beta}+e^{C\beta}} \cdot \frac{e^{B\beta}}{e^{B\beta}+e^{C\beta}} \cdot \frac{e^{C\beta}}{e^{C\beta}}$$

$$PL_2(\beta|p_{t_1}, p_{t_2}, p_{t_3}) = \frac{e^{B\beta}}{e^{A\beta}+e^{B\beta}+e^{C\beta}} \cdot \frac{e^{A\beta}}{e^{A\beta}+e^{C\beta}} \cdot \frac{e^{C\beta}}{e^{C\beta}}$$

Very commonly:

$$PL_1(\beta|p_{t_1}, p_{t_2}, p_{t_3}) \neq PL_2(\beta|p_{t_1}, p_{t_2}, p_{t_3})$$

So which one is right partial likelihood? $PL_1(\beta|p_{t_1}, p_{t_2}, p_{t_3})$ or $PL_2(\beta|p_{t_1}, p_{t_2}, p_{t_3})$

Methods

If there is n tied events, the total possible order would be n!. E.g. if 3 deaths at time $t_{(j)}$, corresponding to patients M, N, P, then equal probability is placed on each ordering: MNP, MPN, NMP, NPM, PMN, PNM. There are $3! = 6$ orderings. The result is a likelihood for a mixture distribution of there 6 different PLs.

1. **Exact method** (This is the best one to use.)

The exact method uses the average partial likelihood. If there are d_j ties at a time $t_{(j)}$, then there are $d_j!$ permutations of the tied individuals. With no other information, any of these permutations of the ties are equally likely. Each of the $d_j!$ permutations gets weight $\frac{1}{d_j}$.

$$PL_{exact} = \frac{\sum_j PL_j}{d_j!}$$

Notes: With large data sets containing many tied survival times, the computation can be slow.

2. Breslow and Efron methods

Both the Breslow and Efron methods are approximations to the exact method. With few ties, Efron and Breslow are equivalent. With severe numbers of ties, Exact or Efron is preferable. For formula please check the textbook.

6.4 Hypothesis of Cox model

Hypothesis:

$$H_0 : \beta = 0$$

$$H_0 : \beta \neq 0$$

Test

1. (Partial) **Likelihood ratio test**

Wilks (1938) showed that, under certain conditions, if, for a parameter β and data set X_n, the null hypothesis $H_0 : \beta = 0$ is true, then.

$$2Ln\Big[\frac{L(\hat{\beta}|X_n)}{L(\beta_0|X_n)}\Big] \xrightarrow{L} \chi^2_{(1)}, n \to \infty$$

This theorem can be applied to the partial likelihood.

$LPL(\hat{\beta})$: is the log partial likelihood evaluated at $\hat{\beta}$ (the LPL maximizer). $LPL(0)$ is the LPL evaluated at $\beta = 0$. Under the null hypothesis $H_0 : \beta = 0$, and under certain conditions, the statistic:

$$G = 2\Big(LPL(\hat{\beta}) - LPL(0)\Big) \xrightarrow{L} \chi^2_{(1)}$$

2. **Wald** test

A similar theorem, often called Cramer's theorem, is that, under the same conditions as the previous theorem, with I as the Fisher information:

$$\sqrt{n}\Big(\hat{\beta} - \beta_0\Big) \xrightarrow{L} N\Big(0, I^{-1}(\beta_0)\Big)$$

To test $H_0 : \beta = \beta_0$

$$Z = \frac{\hat{\beta} - \beta_0}{\widehat{se}(\hat{\beta})}, \widehat{se}(\hat{\beta}) = \frac{1}{\sqrt{I_n(\hat{\beta})}}$$

$I_n(\hat{\beta})$ is the observed information (negative second derivative of the LPL) evaluated $\hat{\beta}$

3. Score test.

Above some proofs, we know:

$$\frac{[\frac{d}{d\beta} Log(L(\hat{\beta}))]^2}{I_n(\beta_0)} \xrightarrow{L} \chi^2_{(1)}$$

Therefore, at the point $\beta = 0$ the test is:

$$z^* = \frac{\frac{d}{d\beta} LPL(\beta)}{\sqrt{I_n(\hat{\beta})}}$$

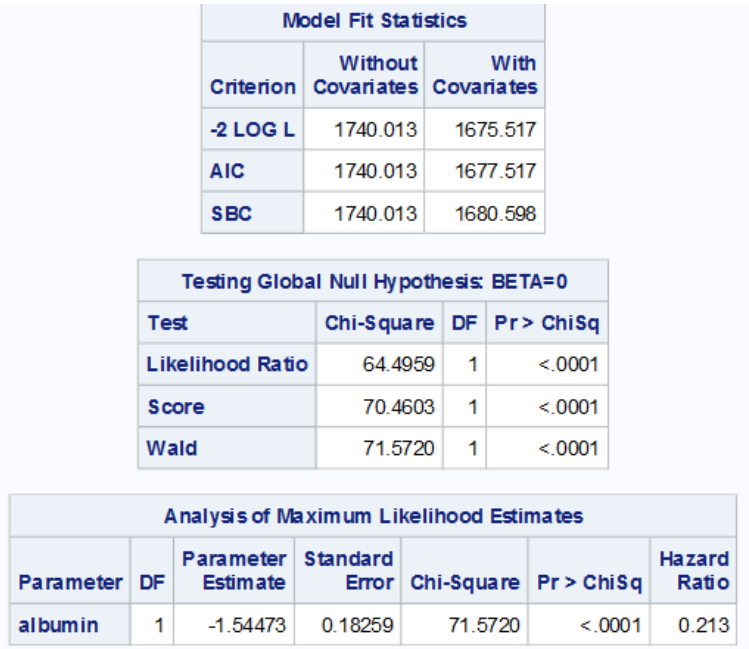
Note: The score test is **identical** to the **log-rank** test.

Example 5.2

The "ties=EXACT" tells it to use the exact method for breaking tied survival times.

```
PROC PHREG DATA=Q3dat2;
   MODEL time*status2(0) = albumin / TIES = exact;
RUN;
```

Figure 16: Codes of SAS, a proportional hazards regression



Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	1740.013	1675.517
AIC	1740.013	1677.517
SBC	1740.013	1680.598

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	64.4959	1	<.0001
Score	70.4603	1	<.0001
Wald	71.5720	1	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
albumin	1	-1.54473	0.18259	71.5720	<.0001	0.213

Figure 17: Results of SAS, a proportional hazards regression

Summary:

1. All 3 tests are asymptotic, and may not work well for small sample sizes.
2. They generally agree pretty well.
3. Some have argued that the LR test is preferable (e.g., Hosmer et al. 2008, Collett 2000, others). However, The score test is debatable the most commonly reported in the medical literature, perhaps because of its equivalence to the non-parametric log-rank test.
4. These 3 tests (Wald, LR, score) are presented in SAS.
5. If study have **no event**, we will have no event time and, consequently, no test due to no LPL!

6.5 Multivariate Cox model

The model still is:

$$h(t) = h_0(t)e^{\beta X}$$

β_k is the change in the log-hazard associated with a one unit increase in x_k WHEN ALL OTHER COVARIATE VALUES ARE HELD FIXED. Or, e^{β_k} is the factor change in the hazard ratio associated with the one unit increase in x_k WHEN ALL OTHER COVARIATE VALUES ARE HELD FIXED.

Hypothesis:

$$H_0 : \beta_i = 0, \beta_{\notin i} = any$$

$$H_1 : \beta_i = any, \beta_{\notin i} = any$$

Stratified Cox Regression

After stratification, proportional hazards assumption can relax some. In particular, between-strata hazards need not be proportional but within strata it need.

Recall the partial likelihood:

$$PL(\beta|p_{t_1}, p_{t_2}, ...p_{t_k}) = \prod_j \frac{e^{\beta X_j}}{\sum\limits_{j \in \mathcal{R}_j} e^{\beta X_{jk}}}$$

For stratified analyses, the PL is calculated separately within each stratum, then combined.

$$PL_{stratified}(\beta|p_{t_1}, p_{t_2}, ...p_{t_k}) = \prod_j PL_j(\beta)$$

Note that is the same for all strata.

Example 5.3

The "ties=EXACT" tells it to use the exact method for breaking tied survival times.

```
PROC PHREG data=retire2;
 model diffage*status(0)=gender /ties=exact;
 strata ageentryq;
run;
```

Figure 18: Codes of SAS, Stratified Cox Regression

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	3.7799	1	0.0519
Score	4.0384	1	0.0445
Wald	4.0064	1	0.0453

Figure 19: Results of SAS, Stratified Cox Regression

6.6 Paired survival data

Stratification can be used to analyze data for matched pairs experiment, or a case-control study.

```
PROC PHREG DATA=Q3dat2;
   MODEL time*status2(0) = albumin / TIES = exact;
   STRATA pair;
RUN;
```

Figure 20: Codes of SAS, paired analysis

6.7 Cumulative baseline hazard function

Breslow estimator:

$$\hat{H}_0(t) = \sum_{t_j \leq t} \frac{d_j}{\sum\limits_{l \in \mathcal{R}_j} e^{\beta X_k}}$$

7 Proportional hazards diagnostics

Because regression models can be misleading by some influential points or easily overfiting, it is necessary to check some assumption and to make model diagnosis.

For a proportional hazards model, we already know that:

$$h(t) = h_0(t)e^{\beta X}$$

$$\therefore \int h(t)dt = e^{\beta X} \int h_0(t)dt$$

$$\therefore H(t) = e^{\beta X} H_0(t)$$

$$\therefore \log\Big(H(t)\Big) = \log\Big(e^{\beta X} H_0(t)\Big) = log\Big(H_0(t)\Big) + \beta X$$

$$\because H(t) = -log\Big(S(t)\Big)$$

$$\therefore \log\Big(S(t)\Big) = e^{\beta X} \log\Big(S_0(t)\Big)$$

$$\therefore S(t) = \Big(S_0(t)\Big)^{e^{\beta X}}$$

Because of the proof above, we can know that if a proportional hazards model hold, $\log\Big(H(t)\Big) = log\Big(H_0(t)\Big) + \beta X$. Log cumulative hazard function versus a function of time should produce **parallel curves** with different x_1.

However, because $H(t)$ or $h(t)$ is unknown, functions above cannot help a lot when diagnosis.

7.1 Schoenfeld residuals

Schoenfeld residuals (Schoenfeld, 1982) are defined as:

$$r_S = \delta_i(x_i - \hat{a}_i), \hat{a}_i = \frac{\sum\limits_{l \in \mathcal{R}_{(t_j)}} x_l e^{\hat{\beta} X_l}}{\sum\limits_{l \in \mathcal{R}_{(t_j)}} e^{\beta X_l}}, if x_i is censored \delta_i = 0$$

Note: Schoenfeld residuals are zero (missing) for censored observations.

Example 6.1:

Table 6.1

Time	Status	X variable
12	1	0
13	1	0
14	1	1
15	1	1
16	1	0

If from cox model we fitted $\widehat{\beta}$ is -0.07, then Schoenfeld residuals are:

1. $r_1 = 1 * (0 - \frac{0*e^{0*(-0.07)}+0*e^{0*(-0.07)}+1*e^{1*(-0.07)}+1*e^{1*(-0.07)}+0*e^{0*(-0.07)}}{e^{0*(-0.07)}+e^{0*(-0.07)}+e^{1*(-0.07)}+e^{1*(-0.07)}+e^{0*(-0.07)}}) = -\frac{2e^{1*(-0.07)}}{3+2e^{(-0.07)}} = -0.38$
2. $r_2 = 1 * (0 - \frac{0*e^{0*(-0.07)}+1*e^{1*(-0.07)}+1*e^{1*(-0.07)}+0*e^{0*(-0.07)}}{e^{0*(-0.07)}+e^{1*(-0.07)}+e^{1*(-0.07)}+e^{0*(-0.07)}}) = -\frac{2e^{1*(-0.07)}}{2+2e^{(-0.07)}} = -0.48$
3. $r_3 = 1 * (1 - \frac{1*e^{1*(-0.07)}+1*e^{1*(-0.07)}+0*e^{0*(-0.07)}}{e^{1*(-0.07)}+e^{1*(-0.07)}+e^{0*(-0.07)}}) = -\frac{2e^{1*(-0.07)}}{1+2e^{(-0.07)}} = 0.35$
4. $r_4 = 1 * (1 - \frac{1*e^{1*(-0.07)}+0*e^{0*(-0.07)}}{e^{1*(-0.07)}+e^{0*(-0.07)}}) = -\frac{e^{1*(-0.07)}}{1+e^{(-0.07)}} = 0.51$
5. $r_5 = 1 * (0 - \frac{0*e^{0*(-0.07)}}{e^{0*(-0.07)}}) = -\frac{0}{e^0} = 0$

The plot

```
tab <- data.frame(Time = c(12, 13, 14, 15, 16), Residuals = c(-0.38,
    -0.48, 0.35, 0.51, 0), ID = c("r1", "r2", "r3", "r4", "r5"))
p <- ggplot(tab, aes(x = Time, y = Residuals, col = ID)) + geom_point(size = 3) +
    labs(x = "Event time", y = "Schoenfeld residuals", col = "Event ID") +
    plot_theme

p
```

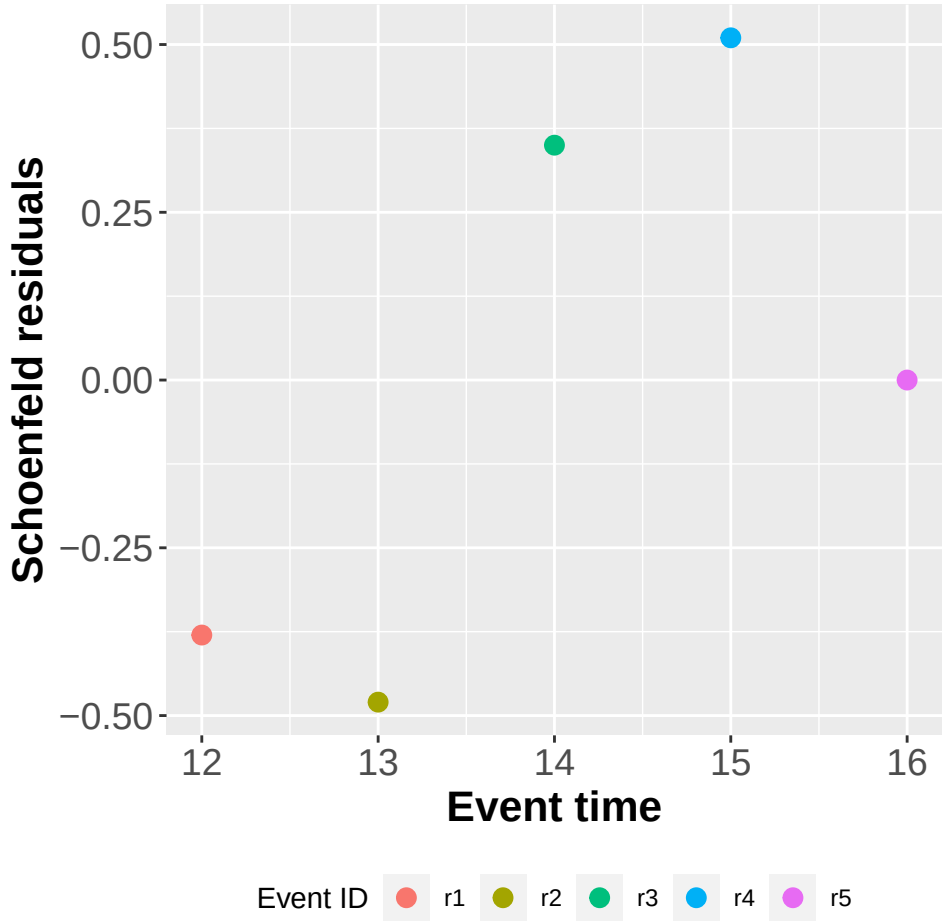


Figure 21: Schoenfeld residuals plot

Notes:

1. If model have more covariates there will be a SEPARATE SET OF SCHOENFELD RESIDUALS FOR EACH COVARIATE $\hat{\beta}_1 ... \hat{\beta}_k$.
2. The Schoenfeld residuals should, if the model is true, be independent of time.
3. A plot that shows a non-random pattern over time is evidence of violation of the proportional hazards assumption.
4. In single covariate model, scaled Schoenfeld residuals are used to assess the overall fit of the model.
5. In multivariate models, scaled Schoenfeld residuals are used to assess the model fit to individual covariates.
6. In either case, the scaled Schoenfeld residuals can be used to assess departures from proportional hazards.
7. A line can be fit to the plot followed by a test for zero slope; a nonzero slope is an evidence against proportional hazards.
8. Schoenfeld residuals around **zero**

Example 6.1:

```
PROC PHREG DATA=Q2dat;
    MODEL lenfol*fstat(0) = age    bmi     gender/ TIES = exact;
    OUTPUT OUT=a WTRESSCH=schgroup;
RUN;

PROC SGPLOT DATA=a;
LOESS X = lenfol Y = schgroup;
WHERE schgroup^=.;
RUN;
```

Figure 22: Codes of SAS, Schoenfeld residuals

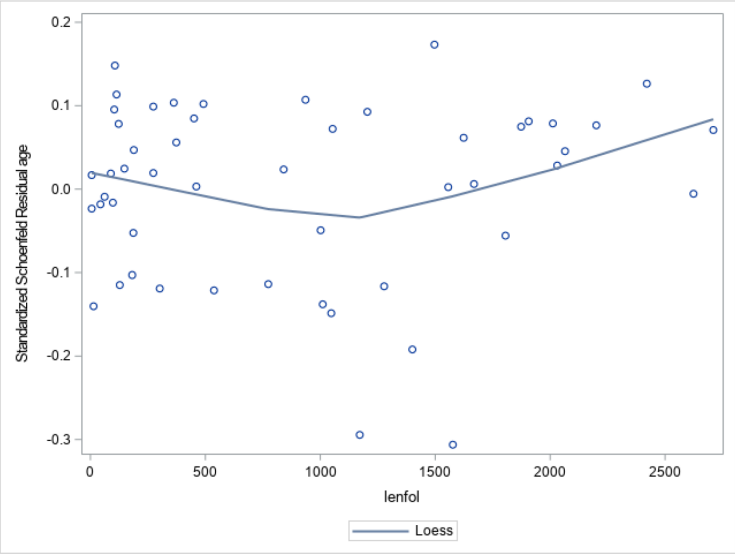


Figure 23: Results of SAS, Schoenfeld residuals of age

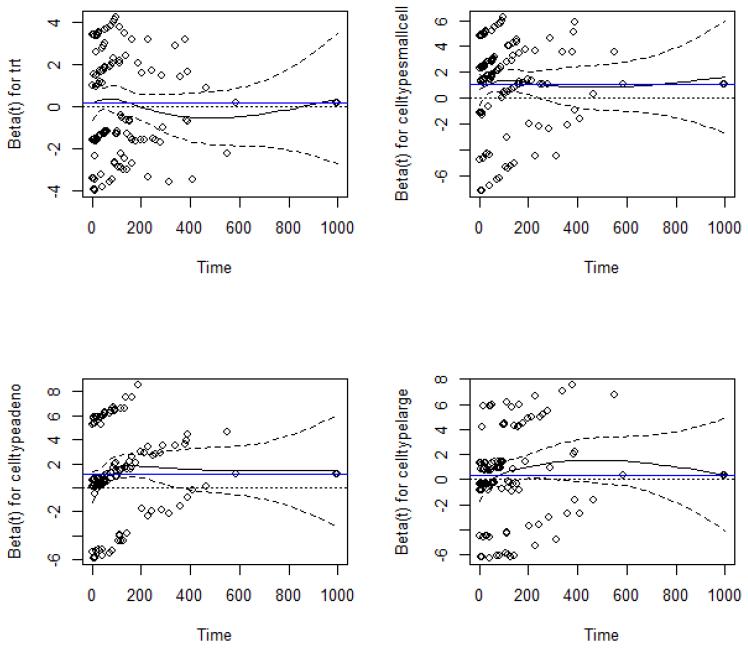


Figure 24: Results of SAS, More Schoenfeld residuals of age

If a zero slope line located in the point-wise confidence interval. it still suggest a proportional hazards hold model.

7.2 Martingale residuals

A Martingale is like a continuous version of a random walk.

$$r_M = \delta_i - \hat{H}(t_i, x_i, \beta), \hat{H}(t_i, x_i, \hat{\beta}) = e^{\hat{\beta}' x_i} \hat{H}_0(t_i), if x_i is censored \delta_i = 0$$

$$\hat{H}_0(t) = \sum_{t_j \leq t} \frac{1}{\sum_{j \in \mathcal{R}_{(t_i)}} e^{\beta' x_j}}$$

Notes:

1. Unlike Schoenfeld residuals, there is just one set of Martingale residuals for a fitted model.
2. Censored observations always have negative Martingale residuals.
3. Martingale residuals take values between $(-\infty, 1]$.
4. Martingale residuals sum to zero.
5. Values near 1 or with large negative numbers are potential outliers.
6. The asymmetry makes the plot a bit hard to interpret.
7. Martingale residuals can also be plotted against individual covariates in the model. The covariate itself is **excluded** from the model from which the residuals are calculated.
8. Unlike other plots, linear trends when a Martingale residual is plotted against a variable are **not** a model violation. But **curvature** is a model violation.

Example 6.2:

```
PROC PHREG DATA=Q2dat;
    MODEL lenfol*fstat(0) = bmi    gender/ TIES = exact;
    OUTPUT OUT=a RESMART= MART;
RUN;

PROC SGPLOT DATA=a;
LOESS X = age Y = MART;
WHERE MART^=.;
RUN;
```

Figure 25: Codes of SAS, Martingale residuals

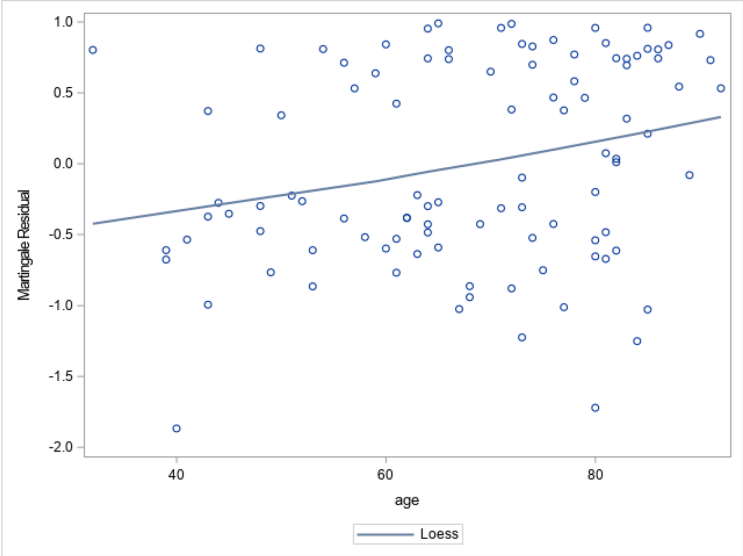


Figure 26: Results of SAS, Martingale residuals of age

7.3 Influential points

The classical way of assessing a point's influence in regression on an estimate $\hat{\beta}$ is to simply:

1. Delete the point, say i
2. Re-estimate the slope on the n $-$ 1 remaining individuals, $\hat{\beta}_i$
3. Calculate "dfbetas" $\hat{\beta} - \hat{\beta}_i$

7.4 Time-dependent covariates

The Cox PH model is:

$$h(t) = h_0(t)e^{\beta X}$$

If a single variable x_j changes over time, then the result is a time-dependent covariate model. Only introduce time-dependent covariates into a model if they are really required.

$$h(t) = h_0(t)e^{\beta f_{x_j}(t)+\beta_i x_i}$$

Hazard ratio

1. HR at time t:

Let r and s be different individuals with covariates $x_r(t)$ and $x_s(t)$ at time t.

$$HR = \frac{h(t, x_r(t), \beta)}{h(t, x_s(t), \beta)} = e^{[x_r(t)-x_s(t)]}$$

2. HR at time t_1 and t_2:

$$HR = \frac{h(t, x_r(t_1), \beta)}{h(t, x_s(t_2), \beta)} = \frac{h_0(t_1)}{h_0(t_2)}e^{[x_r(t)-x_s(t)]}$$

Partial likelihood:

$$PL(\beta) = \prod_{i=1}^{n} \frac{e^{\beta x_i(t_i)}}{\sum\limits_{k\in\mathcal{R}_i} e^{\beta x_k(t_i)}}$$

Note: Only calculate on event time not on censored time.

Example 6.3:

The following small data set contains survival information from four patients and smoking status x(1), x(2), x(3), and x(4) at each death time.

Time (month)	Status	X(1)	X(2)	X(3)	X(4)
3	1	1	0	0	.
2	0	1	0	.	.
1	1	1	.	.	.
4	1	0	0	1	0

Figure 27: Four patients survival information

$$PL(\beta) = (\frac{e^{\beta*1}}{e^{\beta*1} + e^{\beta*1} + e^{\beta*1} + e^{\beta*0}}) * (\frac{e^{\beta*0}}{e^{\beta*0} + e^{\beta*1}}) * (\frac{e^{\beta*0}}{e^{\beta*0}})$$

Types of time-dependent covariates:

Internal covariate: An internal covariate is a measurement that is taken on a "living" patient (e.g. Lung function).

External covariate: An external covariate is a measurement that does not require a "living" patient (e.g. environmental pollen density).

Example 6.4:

Time is measured from the time of first being put on the transplant list to death. The covariates are transplant status (transstat, 0=none, 1=received), prior surgery (1=surgery,0=none), and age in years.

```
Data transplant;
INPUT time status wait psurg age;
DATALINES;
1   1   0   0   53
2   1   0   0   43
2   1   0   0   53
....
RUN;

PROC PHREG DATA=transplant;
MODEL time*status(0) = transstat   psurg  age / TIES = exact;
RUN;
```

Figure 28: Codes of SAS, Withou cox time-dependent

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
transstat	1	-1.66139	0.27591	36.2587	<.0001	0.190
psurg	1	-0.74211	0.44216	2.8169	0.0933	0.476
age	1	0.05920	0.01494	15.6964	<.0001	1.061

Figure 29: Result of SAS, Withou cox time-dependent

The effect of transplant when treated as a baseline covariate significant. Transplants seem to be saving lives.

Next

Transplant status is treated as a time-varying covariate. It is zero before the transplant, and 1 after the transplant.

```
PROC PHREG DATA=transplant;
MODEL time*status(0) = transstat   psurg  age / TIES = exact;
IF wait>=time OR wait=. THEN transstat=0; ELSE transstat=1;
/* This is a loop for each failure time, sas will repeat each t for
transtat result, not the time in dataset.
Transplant status is treated as a time-varying covariate.
It is zero before the transplant, and 1 after the transplant */
RUN;
```

Figure 30: Codes of SAS, COX time-dependent

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
transstat	1	-0.08745	0.30361	0.0830	0.7733	0.916
psurg	1	-1.10550	0.42976	6.6170	0.0101	0.331
age	1	0.03359	0.01403	5.7326	0.0167	1.034

Figure 31: Result of SAS, COX time-dependent

Now, there is no evidence of a beneficial effect of the heart transplant surgery. Because we only count who get surgery for analysis in non-time-dependent, the effect point to the surgery group which had longer survival may only due to an invert causal relationship that healthy patients can wait longer for donation.

8 Parametric model

A parametric model (e.g., for survival times) summarizes the whole distribution in a small number of parameters. Parametric models can be used to model large data sets. The drawback of parametric models is that they add an assumption about the hazard to the semi-parametric approach.

Three common parametric models:

8.1 Exponential model

Exponential model have Constant hazard over time.

For $\lambda > 0, t > 0$

$S(t) = e^{-\lambda t}$

$F(t) = 1 - e^{-\lambda t}$

$f(t) = \lambda e^{-\lambda t}$

$h(t) = \frac{f(t)}{S(t)} = \lambda$, Constant hazard over time

$H(t) = \int h(u)du = \lambda t$

Mean:

$E[T] = \frac{1}{\lambda}$

Median:

$S(t) = e^{-\lambda t} = 0.5 \rightarrow t_m = \frac{ln(2)}{\lambda}$

8.2 Weibull model

Weibull model: Increasing or decreasing hazard over time.

For $\lambda > 0, \alpha > 0$

$S(t) = e^{-\lambda t^\alpha}$

$F(t) = 1 - e^{-\lambda t^\alpha}$

$f(t) = \lambda \alpha t^{\alpha-1} e^{-\lambda t}$

$h(t) = \frac{f(t)}{S(t)} = \lambda \alpha t^{\alpha-1}$, Hazard could increase or decrease monotonic over time.

$H(t) = \int h(u)du = \lambda t^\alpha$

Note: If $\alpha = 1$, then the distribution is exponential.

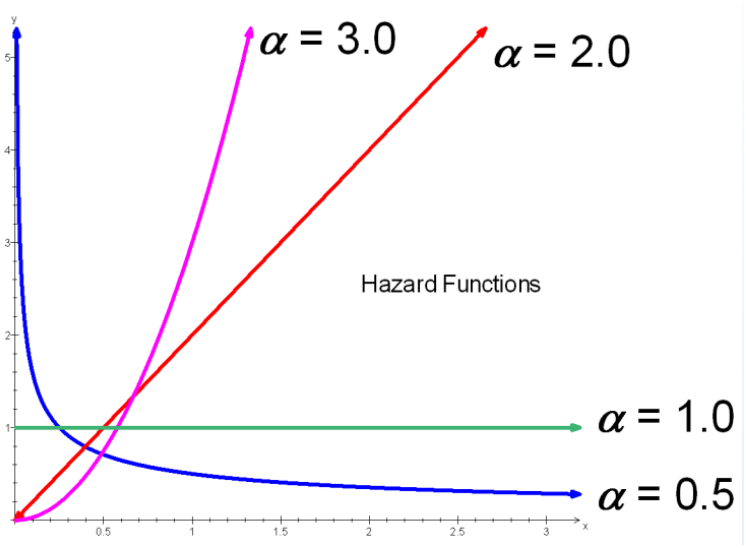


Figure 32: Different alpha, Weibull distribution h(t)

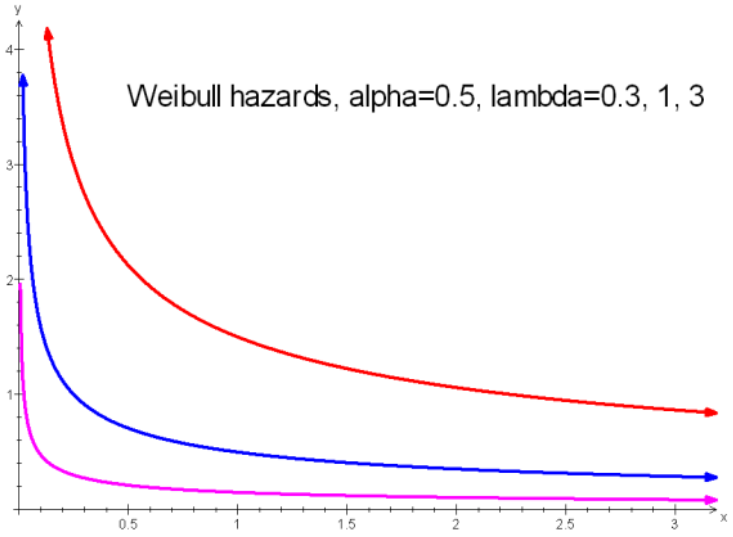


Figure 33: Different lambda, Weibull distribution h(t)

Mean:

$E[T] = \frac{\Gamma(1+\frac{1}{\alpha})}{\lambda^{\frac{1}{\alpha}}}$

$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt$. For positive integer $\Gamma(n) = (n-1)!$

Median:

$S(t) = e^{-\lambda t^\alpha} = 0.5 \rightarrow t_m = (\frac{ln(2)}{\lambda})^{\frac{1}{\alpha}}$

8.3 Log-logistic model

The Log-logistic model: Permits non-monotone hazard over time.

$S(t) = \frac{1}{1+\lambda t^{\alpha}}$

$F(t) = 1 - \frac{1}{1+\lambda t^{\alpha}}$

$f(t) = \frac{\alpha t^{\alpha-1}\lambda}{(1+\lambda t^{\alpha})^2}$

$h(t) = \frac{\alpha t^{\alpha-1}\lambda}{1+\lambda t^{\alpha}}$, Permits non-monotone hazard over time.

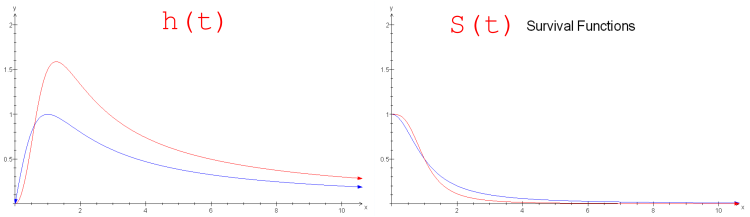


Figure 34: Log-logistic distribution; red: lambda =1, alpha = 2; blue: lambda =1, alpha = 3

Mean:

$E[T] = \frac{\pi}{sin(\frac{\pi}{\alpha})\alpha\lambda^{\frac{1}{\alpha}}}, \alpha > 1$

Median:

$t_m = \lambda^{-\frac{1}{\alpha}}$

Notes:

1. The log-logistic distribution models a non-monotone, non-constant hazard.
2. It is possible that there is a "high risk" time period after which hazard drops off like this.

8.4 Diagnosis

Is a hazard function Weibull ?

Because from data we can estimation cumulative hazard with NA or KM method, then if Weibull hold, it should follow as

$$H(t) = \int h(u)du = \lambda t^{\alpha}$$

. Therefore,

$$log\Big(H(t)\Big) = log(\lambda) + \alpha log(t)$$

. When we plot it, it should be a straight line with slope α and intercept $ln(\lambda)$. If a straight line, then it's OK to weibull distribution.

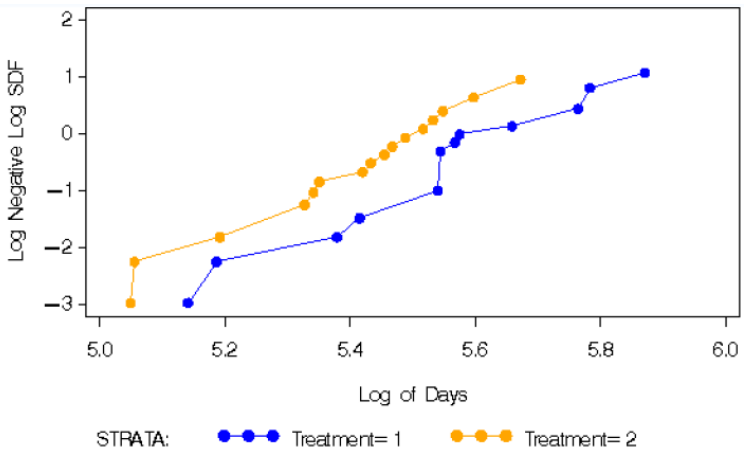


Figure 35: Check H(t), could be similar alpha and different lambda

Weibulls and PH

When two groups have same α, the proportional hazards assumption holds:

$h_1(t) = \lambda_1 \alpha t^{\alpha-1}$

$h_2(t) = \lambda_2 \alpha t^{\alpha-1}$

Is a hazard function log-logistic ?

$$\because Log\Big(\frac{S(t)}{1-S(t)}\Big) = Log\Big(\frac{\frac{1}{1+\lambda t^{\alpha}}}{1-\frac{1}{1+\lambda t^{\alpha}}}\Big) = -Log(\lambda) - \alpha log(t)$$

Therefore, with $log(t)$ as x, $Log\Big(\frac{S(t)}{1-S(t)}\Big)$ as y, we can diagnosis with plot. Or we can also check the shape of $h(t)$.

8.5 Examples

Parametric survival models require stronger assumptions than non-parametric and semi-parametric models. An advantage of parametric models is they permit full description of the hazard function.

Example 7.1: Exponential model

```
PROC LIFEREG DATA=myaml;
MODEL dur*status(0) = / DIST=exponential;
RUN;
```

Figure 36: Codes of SAS, Fiting exponential model

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.1355	0.3015	2.5445	3.7264	108.14	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		
Weibull Scale	1	23.0000	6.9348	12.7374	41.5312		
Weibull Shape	0	1.0000	0.0000	1.0000	1.0000		

Figure 37: Results of SAS, Fiting exponential model

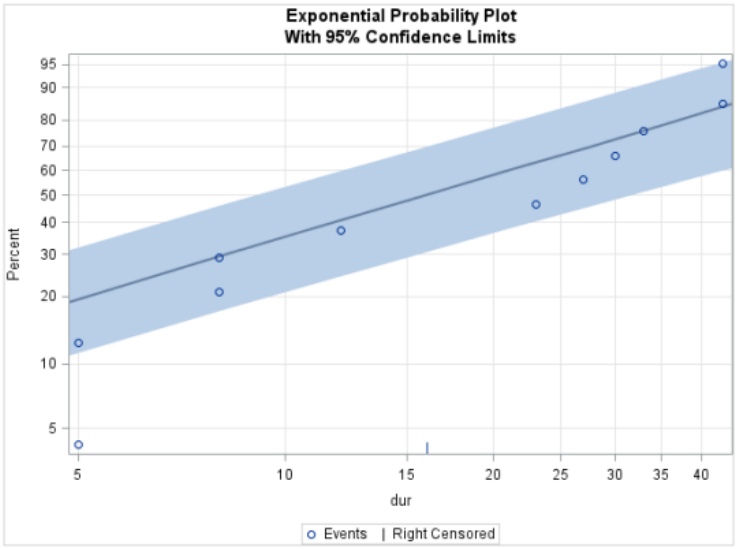


Figure 38: Results of SAS, Dignosis of exponential model

Notes:

1. From the result we had $\hat{\lambda} = \frac{1}{23}$.
2. Exponential model is a special Weibull model ($\alpha = 1$).
3. Mean = 23, Median = 15.9
4. Because $log\Big(H(t)\Big) = log(\lambda) + \alpha log(t)$ and $H(t) = -log(S(t))$. The plot of $log(t)$ vs $log\Big(log(S(t))\Big)$ should close to a straight line with slope = $\alpha = 1$ if exponential model hold.

Example 7.2: Weibull model

```
PROC LIFEREG DATA=myaml;
MODEL dur*status(0) = / DIST=weibull;
RUN;
```

Figure 39: Codes of SAS, Weibull model

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.2146	0.1958	2.8308	3.5985	269.42	<.0001
Scale	1	0.6278	0.1538	0.3884	1.0146		
Weibull Scale	1	24.8937	4.8753	16.9584	36.5421		
Weibull Shape	1	1.5930	0.3902	0.9856	2.5746		

so, 1 may good enough

Figure 40: Results of SAS, Fiting Weibull model

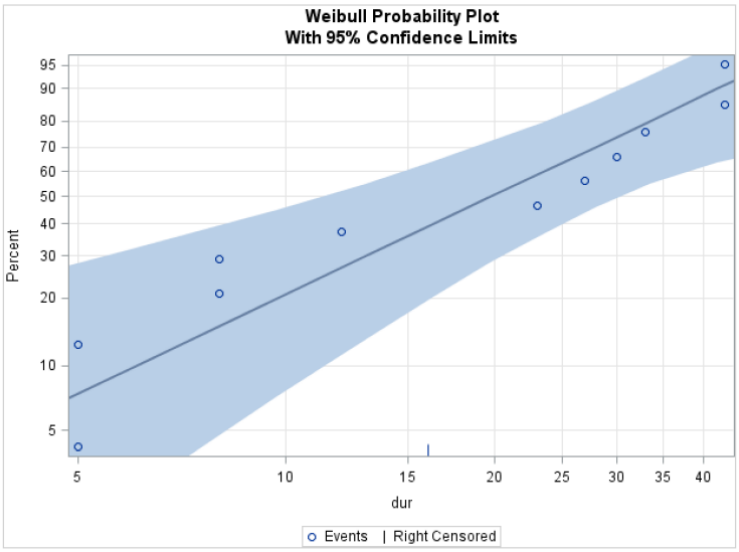


Figure 41: Results of SAS, Dignosis of Weibull model

Notes:

1. Rather than $S(t) = e^{-\lambda t^{\alpha}}$, SAS uses the parameterization $S(t) = e^{-(\frac{1}{\tau})^{\alpha} t^{\alpha}}$
2. τ is the Weibull scale parameter in the SAS output,
3. α is the Weibull shape parameter.
4. $\lambda = (\frac{1}{\tau})^{\alpha}$.
5. Because $log\Big(H(t)\Big) = log(\lambda) + \alpha log(t)$ and $H(t) = -log(S(t))$. The plot of $log(t)$ vs $log\Big(log(S(t))\Big)$ should close to a straight line with slope $= \alpha$ if Weibull model hold.
6. To test if the exponential model is adequate, we can test $H_0 : \alpha = 1$. Note that the confidence interval for α is (0.9856,2.5746), we do not reject H_0.

Example 7.2: Log-logistic model

```
PROC LIFEREG DATA=myaml;
MODEL dur*status(0) = / DIST=llogistic;
RUN;
```

Figure 42: Codes of SAS, Log-logistic model

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	5.0200	0.2652	4.5001	5.5398	358.21	<.0001
Scale	1	0.7533	0.1629	0.4930	1.1510		

Figure 43: Results of SAS, Fiting Log-logistic model

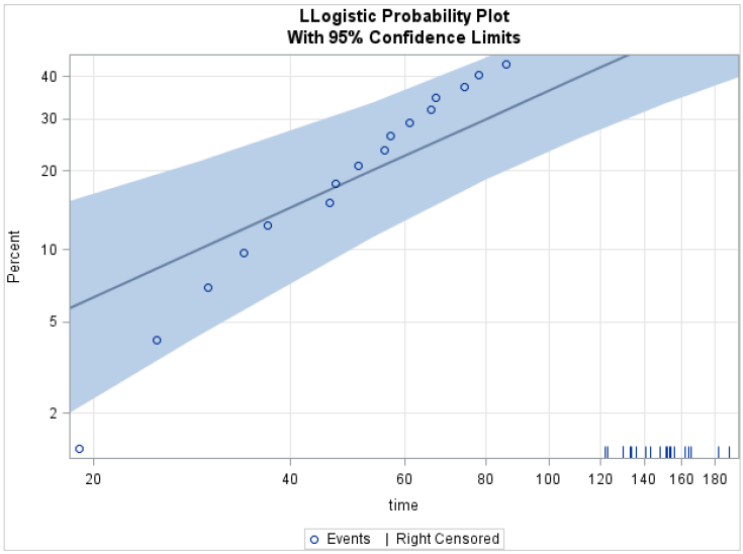


Figure 44: Results of SAS, Dignosis of Log-logistic model

Notes:

1. For SAS, $\lambda = \frac{1}{Scale} = \frac{1}{0.7533} = 1.327$
2. For SAS, $\alpha = e^{\frac{-intercept}{Scale}} = e^{\frac{-5.0200}{0.7533}} = 0.00128$
3. From the plot, $Log\left(\frac{S(t)}{1-S(t)}\right) = -Log(\lambda) - \alpha log(t)$ didn't suggest a violated model.

9 Accelerated failure time regression models

9.1 AFT

For a random time-to-event T, an accelerated failure time (AFT) model proposes the following relationship between covariates and $Y = log(T)$:

$$log(T) = \beta_0 + \beta_1 x_1 + ... + +\beta_p x_p + \sigma\epsilon$$

Now if all the $x_i = 0$ and $\beta_0 = 0$, then we define the baseline T_0 as:

$$log(T_0) = \sigma\epsilon$$

So time-to-event T of AFT model is same to:

$$log(T) = \beta_0 + \beta_1 x_1 + ... + + \beta_p x_p + log(T_0)$$

or

$$T = T_0 \, e^{\beta_0 + \beta_1 x_1 + ... + + \beta_p x_p}$$

or

$$E[T|x_i + 1] = e^{\beta} E[T|x_i]$$

or

$$S(t) = S_0(\frac{t}{e^{\beta_0 + \beta_1 x_1 + ... + + \beta_p x_p}})$$

or

$$h(t) = e^{-(\beta_0 + \beta_1 x_1 + ... + + \beta_p x_p)} h_0(\frac{t}{e^{\beta_0 + \beta_1 x_1 + ... + + \beta_p x_p}})$$

Under the AFT model, the time scale "speeds up" (accelerates) or "slows down" (decelerates). $e^{-\beta}$ is called the "acceleration factor".

9.2 Diagnosis

Because $T = T_0 \, e^{\beta_0 + \beta_1 x_1 + ... + + \beta_p x_p}$, then $t(p_{th}) = e^{\beta_0 + \beta_1 x_1 + ... + + \beta_p x_p} \, t_0(p_{th})$ A quantile-quantile plot of the survival times should approximate a straight line through the origin with slope of $e^{\beta_0 + \beta_1 x_1 + ... + + \beta_p x_p}$.

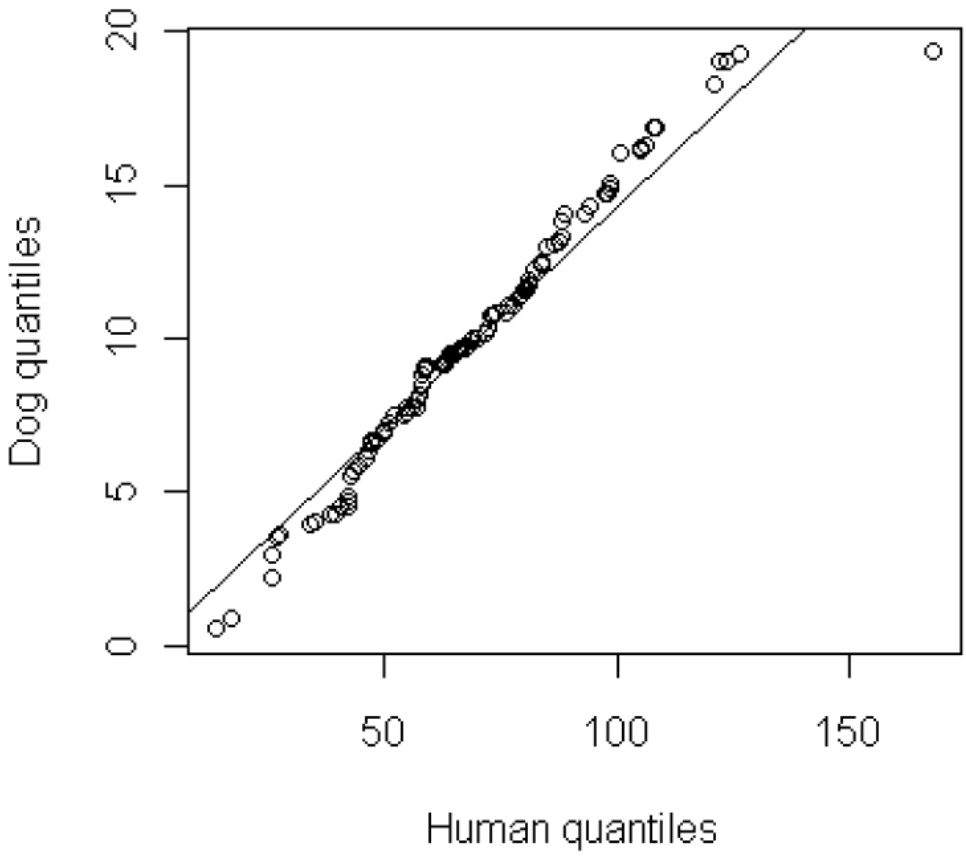


Figure 45: Results of SAS, QQplot of Weibull AFT model

Note: Diagnostics are not as well studied as for the Cox proportional hazards model.

9.3 Weibull AFT model

The Weibull AFT model is

$$log(T) = \beta_0 + \eta_0 + \sigma\epsilon$$

$$S_T(t) = e^{-e^{\frac{log(t)-\beta_0-\eta_x}{\sigma}}} = e^{-(t^{\frac{1}{\sigma}})(e^{\frac{-\beta_0-\eta_x}{\sigma}})}$$

This is a Weibull distribution with $\alpha = \frac{1}{\sigma}$ and $\lambda = e^{\frac{-\beta_0-\eta_x}{\sigma}}$

The hazard function is $h(t) = \frac{1}{\sigma}t^{(\frac{1}{\sigma}-1)}e^{\frac{-\beta_0-\eta_x}{\sigma}}$. This is a **PH model**.

Weibull AFT and PH

The PH baseline hazard is:

$$h_0(t) = e^{\beta_0^*}t^{\alpha^*}$$

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

Table 8.1

σ	α^*	Hazard $h_0(t)$
$0 < \sigma < 0.5$	$\alpha^* > 1$	Increasing at increasing rate
$\sigma = 0.5$	$\alpha^* = 1$	Linear increasing in time
$0.5 < \sigma < 1$	$0 < \alpha^* < 1$	Increasing at decreasing rate
$\sigma = 1$	$\alpha^* = 0$	Constant hazard
$1 < \sigma$	$\alpha^* < 0$	Decreasing in time

9.4 Log-logistic AFT model

The log-logistic AFT model is (with ϵ has the logistic distribution with CDF)

$$Log(T) = \beta_0 + \eta_0 + \sigma\epsilon$$

$$S_T(t) = \frac{1}{1 + t^{\frac{1}{\sigma}} e^{\frac{-\beta_0 - \eta_x}{\sigma}}}$$

The odds of survival at time t is:

$$\frac{S(t, x, \beta)}{1 - S(t, x, \beta)} = t^{-\frac{1}{\sigma}} e^{\frac{\beta_0 + \eta_x}{\sigma}}$$

This is a **proportional odds model** but not a **proportional hazard model**: Consider a trial with single covariate x, where x = 0 for the placebo group and x = 1 for the treatment group. The ratio of the odds of survival in the treatment group relative to the control group at any time t is $e^{\frac{\beta_1}{\sigma}}$.

Log-logistic AFT and PH

With $\alpha = \frac{1}{\sigma}$ and $\lambda = e^{-\alpha(\beta_0 + \sum_{i=1}^{p} \beta_i x_i)}$:

$$h(t) = \frac{\alpha\lambda t^{\alpha-1}}{1 + \lambda t^{\alpha}}$$

Table 8.2

σ	Form
$\sigma < 1$	Starts at 0, rises to peak, descends toward 0
$\sigma = 1$	Starts at λ, descends to 0
$\sigma > 1$	Starts at ∞ , descends to 0

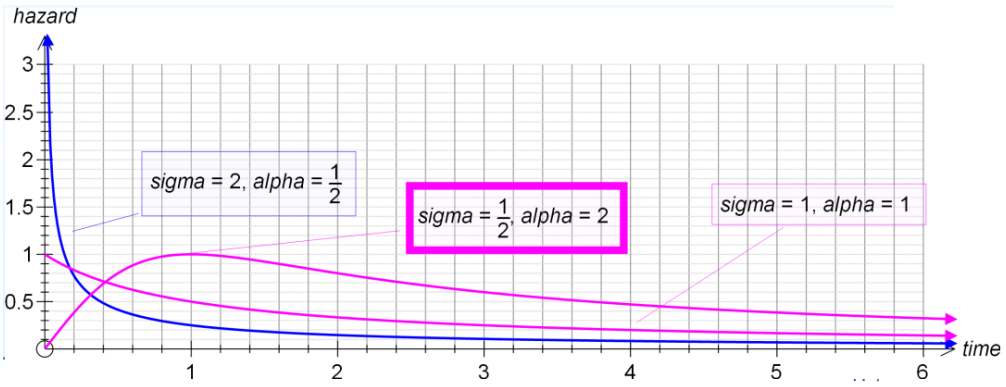


Figure 46: Results of SAS, Dignosis of Log-logistic model

9.5 Examples

Fitting the AFM Weibull in SAS

```
PROC LIFEREG DATA=larynx;
class stage;
MODEL time*status(0) = stage age / distribution=weibull;
PROBPLOT;
output out=b sresidual=resids;
run;
```

Figure 47: Codes of SAS, AFM Weibull

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.5288	0.9041	1.7567	5.3008	15.23	<.0001
stage2	1	-0.1477	0.4076	-0.9465	0.6511	0.13	0.7171
stage3	1	-0.5866	0.3199	-1.2136	0.0405	3.36	0.0668
stage4	1	-1.5441	0.3633	-2.2561	-0.8321	18.07	<.0001
age	1	-0.0175	0.0128	-0.0425	0.0076	1.87	0.1717
Scale	1	0.8848	0.1084	0.6960	1.1250		

Figure 48: Results of SAS, AFM Weibull

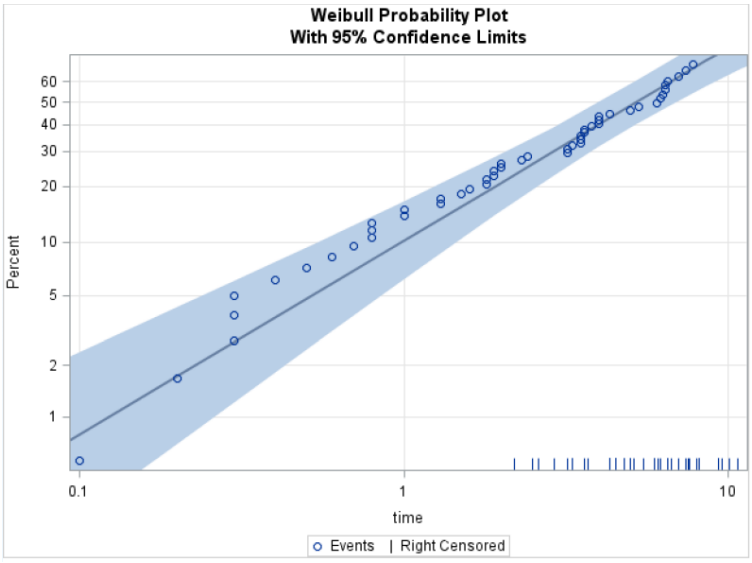


Figure 49: Results of SAS, AFM Weibull

Notes:

1. The acceleration factor for a stage IV patient, compared to a stage I patient, is estimated to be $e^{-(-1.54)} = 4.68$.
2. The median/mean lifetime for a stage I patient is estimated to be 4.68 times greater than for a stage IV patient.
3. Unlike PH, negative coefficients mean larger T at baseline $S(t)$, less $S(t)$, worse survival.

AFM Log-logistic

```
PROC LIFEREG DATA=larynx;
class stage;
MODEL time*status(0) = stage age / distribution=llogistic;
PROBPLOT;
run;
```

Figure 50: Codes of SAS, AFM Log-logistic

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.2656	0.7418	1.8117	4.7195	19.38	<.0001
type	1	-0.0808	0.4481	-0.9592	0.7975	0.03	0.8568
Scale	1	1.1846	0.1437	0.9339	1.5025		

Figure 51: Results of SAS, AFM Log-logistic

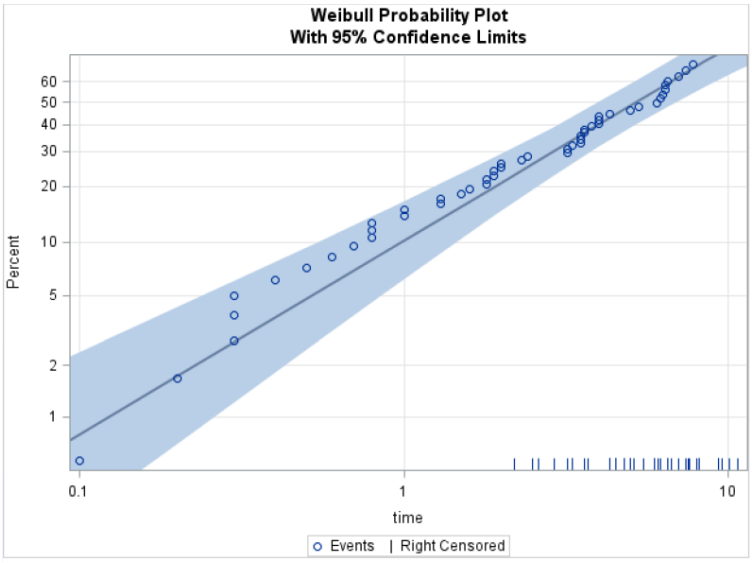


Figure 52: Results of SAS, AFM Log-logistic

Notes:

1. The scale parameter is $\sigma = 1.1846$. So, the estimated hazard is therefore monotone decreasing.
2. At ten months, the estimated odds ratio of survival between the two groups (auto vs allo) is $e^{\frac{\beta_1}{\sigma}} = e^{\frac{-0.0808}{1.1846}} = e^{-0.0682} = 0.93$, with autologous doing worse.

10 End