Survival Analysis Using S/R

Slides für den Weiterbildungs-Lehrgang in
angewandter Statistik an der ETH Zürich

Professor Mara Tableman
Dept. of Mathematics & Statistics
Portland State University
Portland, Oregon, USA

mara.tableman@pdx.edu

16.August 2010

Chapter 1

Rationale for Survival Analysis

- Time-to-event data have as principal endpoint **the length of time until an event occurs**. The event is commonly referred to as a *failure*.

- **Censoring**: A *failure* time is not completely observed.

- **Survival Analysis**: The collection of statistical procedures that accommodate time-to-event censored data.

Example: AML study

Below are preliminary results (1977) from a clinical trial to evaluate the efficacy of maintenance chemotherapy for acute myelogenous leukemia (AML). After reaching a status of remission through treatment by chemotherapy, the patients who entered the study were assigned randomly to two groups. The first group received maintenance chemotherapy; the second, or control, group did not. **The objective of the trial was to see if maintenance chemotherapy prolonged the time until relapse**.

Group	Length of complete remission (in weeks)
Maintained	9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+
Nonmaintained	5, 5, 8, 8, 12, 16+, 23, 27, 30, 33, 43, 45

The + indicates a censored value.

- Serious bias in estimated quantities, which lowers the efficacy of the study.

  a. Throw out censored observations.
  b. Treat censored observations as exact.
  c. Account for the censoring.
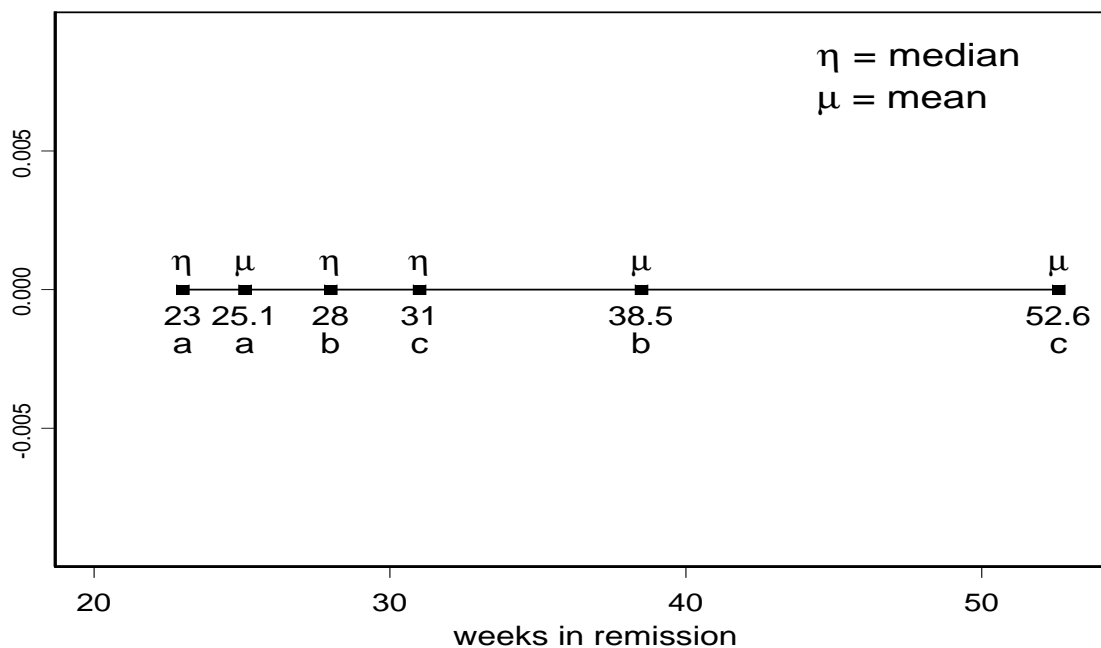
Basic Definitions & Identities

The r.v. T denotes failure time with **cdf** $F(\cdot)$ and **pdf** $f(\cdot)$.

cdf $F(\cdot)$:

$$F(t) = P(T \le t) = \int_0^t f(x)dx \quad \text{and} \quad \frac{dF(t)}{dt} = f(t)$$

That is, by definition of derivative,

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

Survivor function $S(\cdot)$:

$$S(t) = P(T > t) = 1 - F(t) = \int_t^\infty f(x)dx$$

At $t = 0$, $S(t) = 1$ and decreases to 0 as t increases to ∞.

We thus can express the pdf as

$$f(t) = -\frac{dS(t)}{dt} \, .$$

Hazard function $h(\cdot)$:

$$h(t) = \lim_{\Delta t \to 0^+} \frac{P(t < T \le t + \Delta t \,|\, T > t)}{\Delta t} = \frac{f(t)}{S(t)}$$

$$= -\frac{dS(t)/dt}{S(t)} = -\frac{d\log\left(S(t)\right)}{dt}$$

Of course, $h(t) \ge 0$ at all times t.

Cumulative hazard function $H(\cdot)$:

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

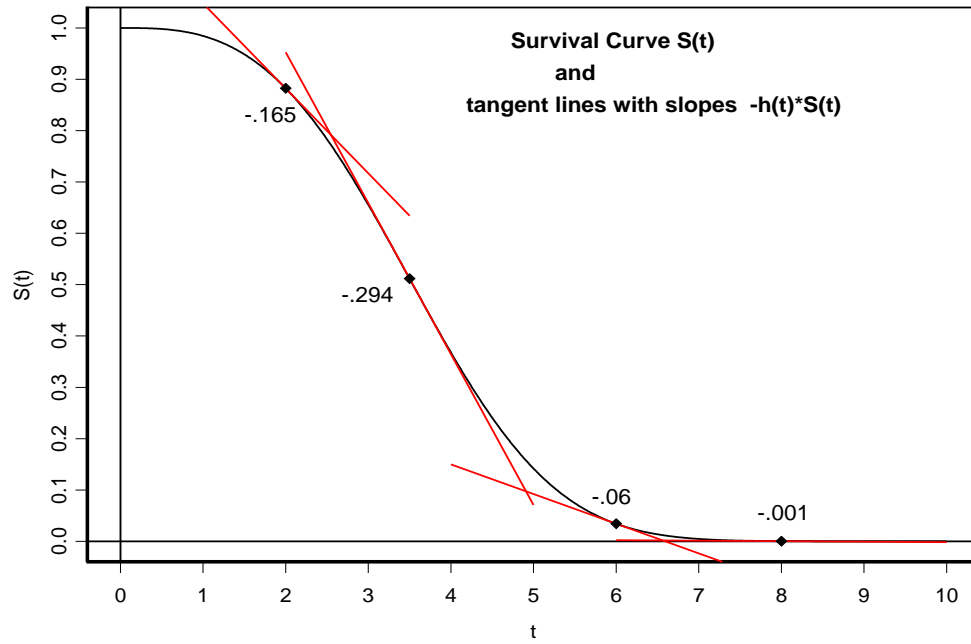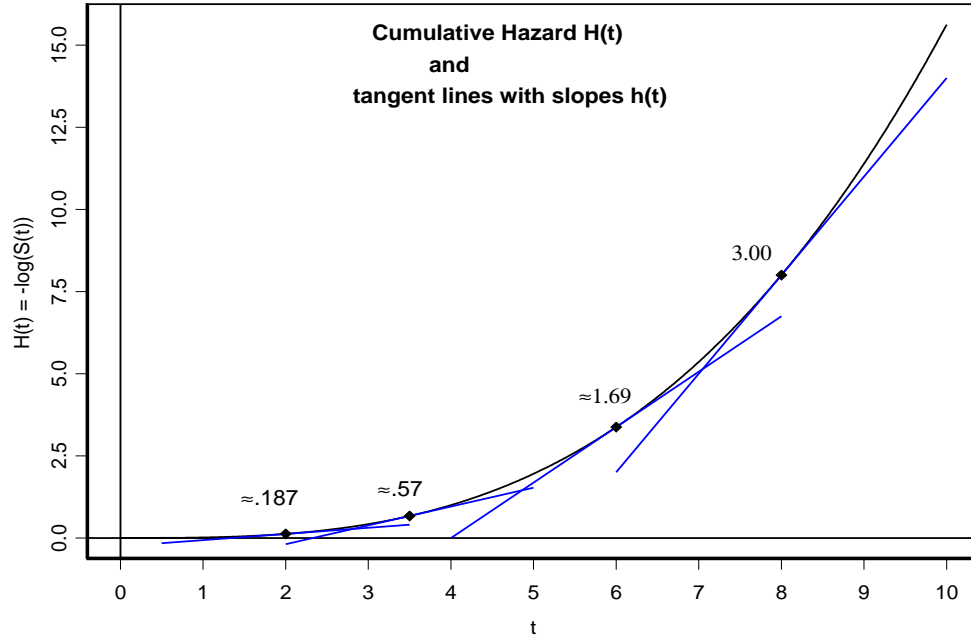At $t = 0$, $H(t) = 0$ and increases to ∞ as t increases to ∞.
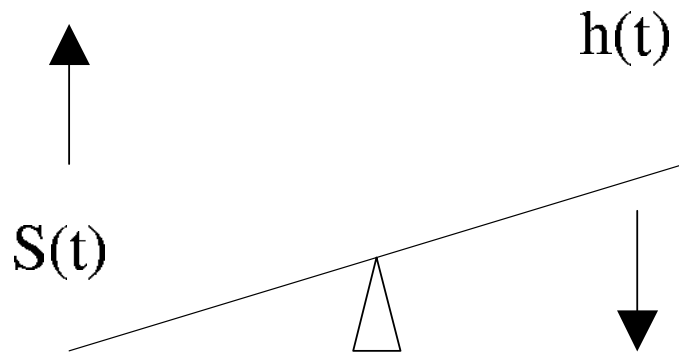
Hence, the relationship

$$S(t) = \exp\left(-H(t)\right).$$

The **hazard function** $h(t) \geq 0$

- specifies the instantaneous rate of failure at $T = t$ given that the individual survived up to time t. It measures the **potential** of failure in an instant at time t given the individual's survival time reaches t.

- is the slope of the tangent line to $H(t) = -\log(S(t))$ at $T = t$

- specifies the distribution of T

Cumulative Hazard H(t)
and
tangent lines with slopes h(t)

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

≈.187   ≈.57   ≈1.69   3.00

Survival Curve S(t)
and
tangent lines with slopes -h(t)*S(t)

$S(t)$

-.165   -.294   -.06   -.001

7

S(t)

h(t)

***p*th-quantile:** The value t_p such that

$$F(t_p) = P(T \leq t_p) = p.$$

That is, $t_p = F^{-1}(p)$.

Also called the $100 \times p$**th percentile**.

Mean Lifetime $E(T)$: For random variable $T \geq 0$,

$$E(T) = \int_0^\infty t \cdot f(t) dt$$

$$= \int_0^\infty S(t) dt.$$

**total area under the survivor curve**

8

Three Censoring Models

Let $T_1, T_2, \ldots, T_n$ be independent and identically distributed (iid) with distribution function (d.f.) F.

Type I censoring:

- In engineering applications, we test lifetimes of transistors, tubes, chips, etc.

- Put them all on test at time $t = 0$ and record their times to failure. Some items may take a long time to "burn out" and we do not want to wait that long to terminate the experiment.

- Terminate the experiment at a **prespecified time** t_c.

- The number of observed failure times is random. If n is the number of items put on test, then we could observe $0, 1, 2, \ldots, n$ failure times.

The following illustrates a possible trial:
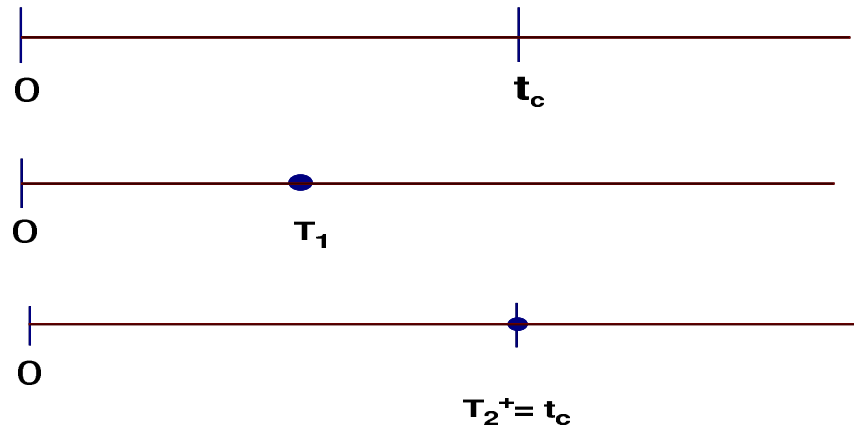
The t_c is a **fixed** censoring time.

- We **do not** observe the T_i, but **do** observe $Y_1, Y_2, \ldots, Y_n$ where

$$Y_i = \min(T_i, t_c) = \left\{ \begin{array}{ll} T_i & \text{if } T_i \leq t_c \\ t_c & \text{if } t_c < T_i. \end{array} \right.$$

- It is useful to introduce a binary random variable δ which indicates if a failure time is observed or censored,

$$\delta = \left\{ \begin{array}{ll} 1 & \text{if } T \leq t_c \\ 0 & \text{if } t_c < T. \end{array} \right.$$

We then observe the iid random pairs (Y_i, δ_i).

Type II censoring:

- In similar engineering applications as above, the experiment is run until a **prespecified fraction** r/n of the n items has failed.

- Let $T_{(1)}, T_{(2)}, \ldots, T_{(n)}$ denote the ordered values of the random sample $T_1, \ldots, T_n$.

  By plan, the experiment is terminated after the rth failure occurs. We only observe the r smallest observations in a random sample of n items.

- For example, let $n = 25$ and take $r = 15$. When we observe 15 burn out times, we terminate the experiment.

- The following illustrates a possible trial:
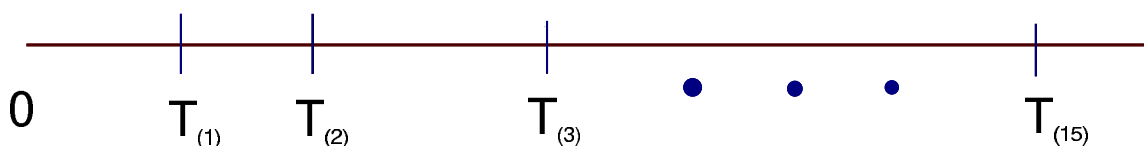
  Here the last 10 observations are assigned the value of $T_{(15)}$. Hence, we have 10 censored observations.

- Notice that we could wait an arbitrarily long time to observe the 15th failure time as $T_{(15)}$ is random; or, we could see all 15 very early on.

- More formally, we observe the following full sample.

$$
\begin{aligned}
Y_{(1)} &= T_{(1)} \\
Y_{(2)} &= T_{(2)} \\
\vdots &\quad \vdots \\
Y_{(r)} &= T_{(r)} \\
Y_{(r+1)} &= T_{(r)} \\
\vdots &\quad \vdots \\
Y_{(n)} &= T_{(r)}.
\end{aligned}
$$

The data consist of the r smallest lifetimes $T_{(1)}, \ldots, T_{(r)}$ out of the n iid lifetimes $T_1, \ldots, T_n$ with continuous p.d.f $f(t)$ and survivor function $S(t)$.

Random Right Censoring:

Random censoring occurs frequently in medical studies. In clinical trials, patients typically enter a study at different times. Then each is treated with one of several possible therapies. We want to observe their "*failure*" time but censoring can occur in one of the following ways:

1. *Loss to Follow-up*. Patient moves away. We never see him again. We only know he has survived from entry date until he left. So his survival time is $\geq$ the observed value.

2. *Drop Out*. Bad side effects forces termination of treatment. Or patient refuses to continue treatment for whatever reasons.

3. *Termination of Study*. Patient is still "alive" at end of study.

The following illustrates a possible trial:
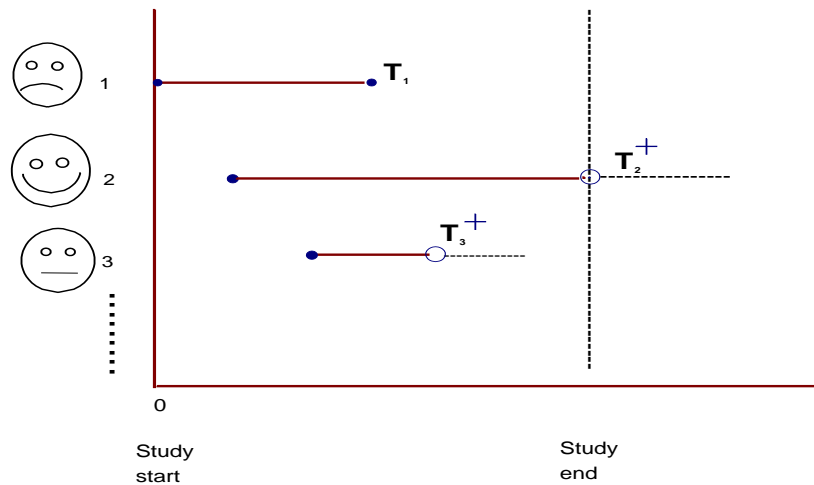
The AML study contain randomly right-censored data.

Formally: Let T denote a lifetime with d.f. F and survivor function S_f and C denote a random censor time with d.f. G, p.d.f. g, and survivor function S_g. Each individual has a lifetime T_i and a censor time C_i. On each of n individuals we observe the pair (Y_i, δ_i) where

$$Y_i = \min(T_i, C_i) \quad \text{and} \quad \delta_i = \left\{ \begin{array}{ll} 1 & \text{if } T_i \leq C_i \\ 0 & \text{if } C_i < T_i \end{array} \right. .$$

- We observe n iid random pairs (Y_i, δ_i).

- The times T_i and C_i are usually assumed to be independent.

- This is a strong assumption. If a patient drops out because of complications with the treatment (case 2 above), it is clearly offended.

14

Remarks:

• If the distribution of C does not involve any parameters of interest, then **the form of the observed likelihood function is the same for these three censoring models**.

$$L = \prod_{i=1}^{n} (f(y_i))^{\delta_i} \cdot (S_f(y_i))^{1-\delta_i}.$$

Thus, regardless of which of the three types of censoring is present, the maximization process yields the same estimated quantities.

• Here we see how censoring is incorporated to adjust the estimates. Each observed value is (y_i, δ_i). An individual's contribution is either it pdf $f(y_i)$; or $S_f(y_i) = P(T > y_i)$, the probability of survival beyond its observed censored time y_i. In the complete data setting, all $\delta_i = 1$; that is, there is no censoring. The likelihood then has the usual form

$$L = \prod_{i=1}^{n} f(y_i).$$

Major Goals

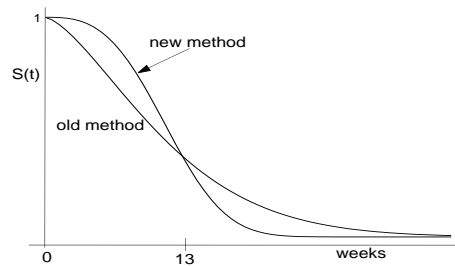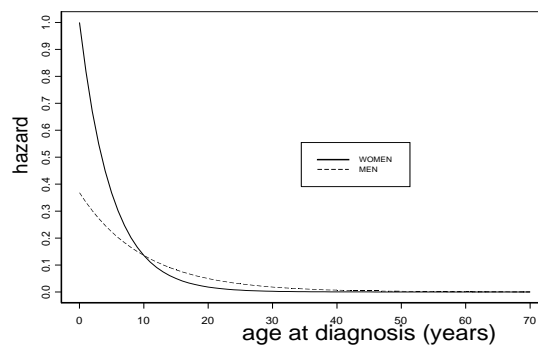**Goal 1.** To estimate and interpret survivor and/or hazard functions from survival data.
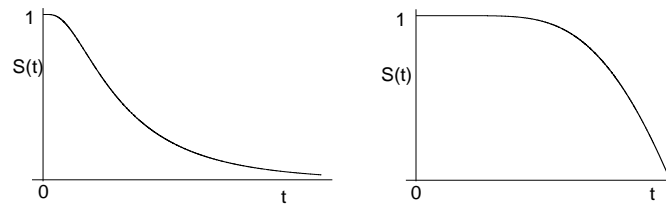
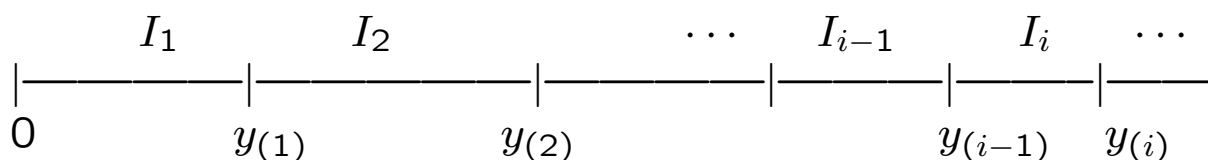**Goal 2.** To compare survivor and/or hazard functions.

Goal 3. To assess the relationship of explanatory variables to survival time, especially through the use of formal mathematical modelling.

16

Chapter 2

Kaplan-Meier Estimator of Survivor Function

$$
\begin{array}{ccccccc}
 & I_1 & I_2 & \cdots & I_{i-1} & I_i & \cdots \\
\vdash\!\!-\!\!-\!\!-\!\!\vdash\!\!-\!\!-\!\!-\!\!-\!\!\vdash\!\!-\!\!-\!\!-\!\!\vdash\!\!-\!\!-\!\!\vdash\!\!-\!\!-\!\!\vdash\!\!-\!\!- \\
0 \quad\quad y_{(1)} \quad\quad y_{(2)} \quad\quad\quad y_{(i-1)} \quad y_{(i)}
\end{array}
$$

The $y_{(i)}$: ith **distinct ordered** censored or uncensored observation and **right endpoint of the interval** I_i, $i = 1, 2, \ldots, n' \leq n$.

- **death** is the generic word for the event of interest. In the AML study, a "relapse" (end of remission period) = "death"

- **Cohort** is a group of people who are followed throughout the course of the study.

- People **at risk** at the beginning of the interval I_i are those people who survived (not dead, lost, or withdrawn) the previous interval I_{i-1} .

  Let $\mathcal{R}(t)$ denote the **risk set just before time** t and let

$$
\begin{aligned}
n_i \ &= \ \# \text{ in } \mathcal{R}(y_{(i)}) \\
&= \ \# \text{ alive (and not censored) just before } y_{(i)} \\
d_i \ &= \ \# \text{ died at time } y_{(i)} \\
p_i \ &= \ P(\text{surviving through } I_i \,|\, \text{alive at beginning } I_i) \\
&= \ P(T > y_{(i)} \,|\, T > y_{(i-1)}) \\
q_i \ &= \ 1 - p_i \ = P(\text{die in } I_i \,|\, \text{alive at beginning } I_i).
\end{aligned}
$$

General multiplication rule for joint events A_1 and A_2:

$$
P(A_1 \cap A_2) = P(A_2 \mid A_1)P(A_1).
$$

From repeated application of this product rule, the survivor function is now expressed as

$$
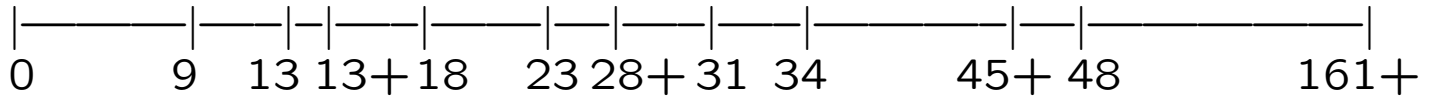S(t) = P(T > t) = \prod_{y_{(i)} \leq t} p_i.
$$

Estimates of p_i and q_i:

$$
\widehat{q_i} = \frac{d_i}{n_i} \qquad \text{and} \qquad \widehat{p_i} = 1 - \widehat{q_i} = 1 - \frac{d_i}{n_i} = \left(\frac{n_i - d_i}{n_i} \right).
$$

K-M estimator of survivor function (1958):

$$
\widehat{S}(t) = \prod_{y_{(i)} \leq t} \widehat{p_i} = \prod_{y_{(i)} \leq t} \left(\frac{n_i - d_i}{n_i} \right).
$$

For AML maintained group:

|———————|——|–|——|———|—|——|——|————————|—|————————|
0   9   13 13+ 18   23 28+ 31 34   45+ 48   161+

$$
\begin{aligned}
\widehat{S}(0) &= 1 \\
\widehat{S}(9) &= \widehat{S}(0) \times \tfrac{11-1}{11} &&= .91 \\
\widehat{S}(13) &= \widehat{S}(9) \times \tfrac{10-1}{10} &&= .82 \\
\widehat{S}(13+) &= \widehat{S}(13) \times \tfrac{9-0}{9} &&= \widehat{S}(13) &&= .82 \\
\widehat{S}(18) &= \widehat{S}(13) \times \tfrac{8-1}{8} &&= .72 \\
\widehat{S}(23) &= \widehat{S}(18) \times \tfrac{7-1}{7} &&= .61 \\
\widehat{S}(28+) &= \widehat{S}(23) \times \tfrac{6-0}{6} &&= \widehat{S}(23) &&= .61 \\
\widehat{S}(31) &= \widehat{S}(23) \times \tfrac{5-1}{5} &&= .49 \\
\widehat{S}(34) &= \widehat{S}(31) \times \tfrac{4-1}{4} &&= .37 \\
\widehat{S}(45+) &= \widehat{S}(34) \times \tfrac{3-0}{3} &&= \widehat{S}(34) &&= .37 \\
\widehat{S}(48) &= \widehat{S}(34) \times \tfrac{2-1}{2} &&= .18 \\
\widehat{S}(161+) &= \widehat{S}(48) \times \tfrac{1-0}{1} &&= \widehat{S}(48) &&= .18
\end{aligned}
$$

The K-M curve is a right continuous step function, which **steps down only at an uncensored observation**.

• Here $\widehat{S}(t)$ is a **defective survival function** as it does not jump down to zero. We cannot estimate $S(t)$ beyond $t = 48$.

19

Estimate of Variance of $\widehat{S}(t)$

Greenwood's formula (1926):

$$\widehat{\text{var}}\left(\widehat{S}(t)\right) = \widehat{S}^2(t) \sum_{y_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

Example with the AML data:

$$\widehat{\text{var}}\left(\widehat{S}(13)\right) = (.82)^2 \left(\frac{1}{11(11-1)} + \frac{1}{10(10-1)}\right) = .0136$$

$$\text{s.e.}\left(\widehat{S}(13)\right) = \sqrt{.0136} = .1166$$

- Theory tells us, for each fixed value t,

$$\widehat{S}(t) \stackrel{a}{\sim} \text{normal}\left(S(t), \widehat{\text{var}}\left(\widehat{S}(t)\right)\right).$$

- At time t, an approximate $(1 - \alpha) \times 100\%$ confidence interval for the probability of survival, $S(t) = P(T > t)$, is given by

$$\widehat{S}(t) \pm z_{\frac{\alpha}{2}} \times \text{s.e.}\left(\widehat{S}(t)\right).$$
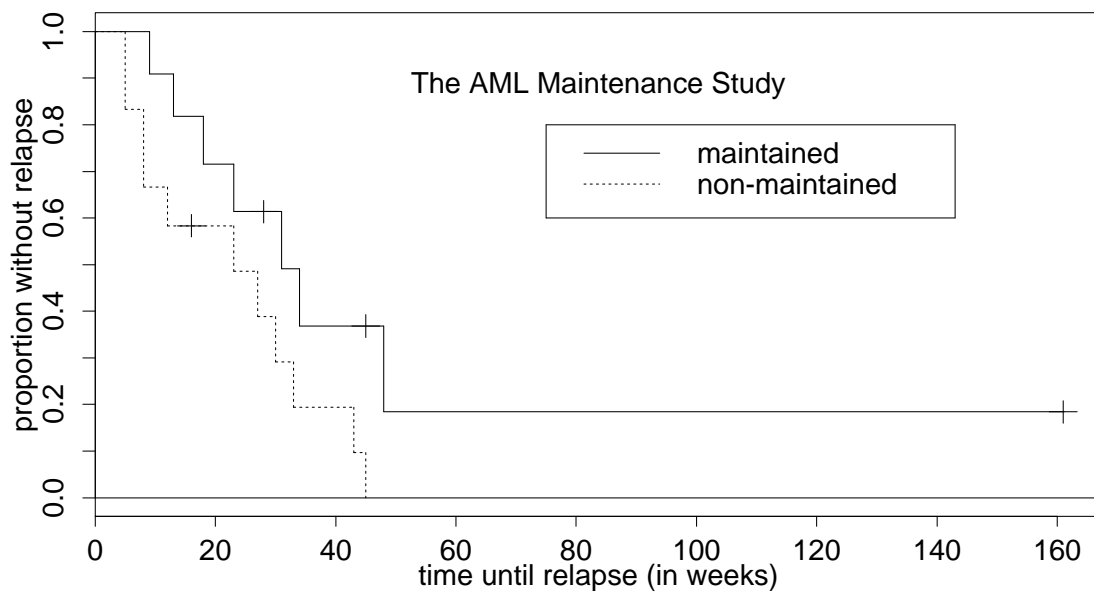
S/R Application

`survfit` **function:**

```
> library(survival) # For R users only
> km.fit <- survfit(Surv(weeks,status)~group,data=aml)
> plot(km.fit,conf.int=F,xlab="time until relapse (in weeks)",
              ylab="proportion without relapse")
> summary(km.fit) # This displays the survival probabilities
                  # for each group.
```

group=0

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
5	12	2	0.8333	0.1076	0.6470	1.000
8	10	2	0.6667	0.1361	0.4468	0.995
12	8	1	0.5833	0.1423	0.3616	0.941
23	6	1	0.4861	0.1481	0.2675	0.883
27	5	1	0.3889	0.1470	0.1854	0.816
30	4	1	0.2917	0.1387	0.1148	0.741
33	3	1	0.1944	0.1219	0.0569	0.664
43	2	1	0.0972	0.0919	0.0153	0.620
45	1	1	0.0000	NA	NA	NA

group=1

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
9	11	1	0.909	0.0867	0.7541	1.000
13	10	1	0.818	0.1163	0.6192	1.000
18	8	1	0.716	0.1397	0.4884	1.000
23	7	1	0.614	0.1526	0.3769	0.999
31	5	1	0.491	0.1642	0.2549	0.946
34	4	1	0.368	0.1627	0.1549	0.875
48	2	1	0.184	0.1535	0.0359	0.944

Plot of the Kaplan-Meier survivor curves

The AML Maintenance Study

maintained
non-maintained

proportion without relapse

time until relapse (in weeks)

Remarks:

- `survfit` fits each group to its own K-M curve.

- **Default** `std.err` is Greenwood's.

- To obtain C.I.'s for $\widehat{S}(t)$ using Greenwood's formula, specify `conf.type = "plain"` in `survfit` function.

- **Default** 95% C.I.'s in `survfit` are called `"log"` and the formula is

$$\exp\left(\log\left(\widehat{S}(t)\right) \pm 1.96\,\text{s.e.}\left(\widehat{H}(t)\right)\right),$$

  where $\widehat{H}(t)$ is the estimated cumulative hazard function (to be defined).

- Sometimes, both of these intervals give limits outside the interval $[0, 1]$. This is not so appealing as $S(t)$ is a probability! Kalbfleisch & Prentice (1980) suggest using the transformation $W = \log(-\log(\widehat{S}(t)))$ to estimate the log cumulative hazard parameter $\log(-\log(S(t)))$, and to then transform back. These intervals will always have limits within the interval $[0, 1]$. Specify `conf.type="log-log"` in the `survfit` function.

Estimate of quantiles:

As K-M is a step function, the inverse is not unique. The estimate is the first observed failure time where K-M steps below $1 - p$. That is,

$$\widehat{t}_p = \min\{t_i : \widehat{S}(t_i) \leq 1 - p\}.$$

AML maintained group: median $\widehat{t}_{.5} = 31$ weeks.

• The function `quantile.km` computes an estimated pth-quantile, its s.e., and an approximate $(1 - \alpha) \times 100\%$ C.I.

```
> quantile.km(km.fit,.25,.05,1.96) # the .25th-quantile
 [1] "summary"
   qp se.S.qp   f.qp  se.qp   LCL    UCL
 1 18  0.1397 0.0205 6.8281 4.617 31.383  # in weeks
```

Truncated mean survival time:

The estimated mean is taken to be

$$\widehat{\text{mean}} = \int_0^{y_{(n)}} \widehat{S}(t) \, dt \, ,$$

where $y_{(n)} = \max(y_i)$.

• `survfit` function provides the estimates.

```
> km.fit
          n   events   mean   se(mean) median 0.95LCL 0.95UCL
```

24

```
group=0 12     11   22.7     4.18    23        8      NA
group=1 11      7   52.6    19.83    31       18      NA
```

- Median is the preferred descriptive measure of typical survival time because it is more resistant to presence of extreme values. One large survival time can grossly inflate the mean.

Comparison of Two Survivor Curves

Test: $H_0 : S_1(\cdot) = S_0(\cdot)$ against $H_A : S_1(\cdot) \neq S_0(\cdot)$

Log-rank test:

Use S function `survdiff`.

```
> survdiff(Surv(week,status)~group,data=aml)
```

```
          N  Observed  Expected (O-E)^2/E  (O-E)^2/V
 group=1 11         7     10.69      1.27        3.4
 group=2 12        11      7.31      1.86        3.4

 Chisq= 3.4  on 1 degrees of freedom, p= 0.0653
```

At 5% level of significance, there is insufficient evidence to conclude maintenance chemo effects remission time.

Remarks:

1. log-rank stat $= MH^2$,

    where MH $=$ the Mantel-Haenszel statistic (1959).

    - Under H_0, MH $\overset{a}{\sim}$ N$(0, 1)$.
    - Use MH $= \sqrt{\log - \text{rank}}$ to test $H_A : S_1(\cdot) > S_0(\cdot)$.

    MH $= \sqrt{3.4} = 1.844$

    with p-value $= .0653/2 = .033$.

    There is mild evidence to suggest that maintenance chemotherapy prolongs the remission period, since the one-sided test is appropriate.

2. The `survdiff` function contains a `rho` parameter.
    - **Default**, `rho = 0`, conducts log-rank test.
    - When `rho = 1`, conducts Peto test .

Empirical Estimates of Hazard

Estimates of hazard (risk):

1. Estimate at a distinct ordered death time t_i:

$$\widetilde{h}(t_i) = \frac{d_i}{n_i}.$$

2. Estimate of hazard in the interval $t_i \leq t < t_{i+1}$:

$$\widehat{h}(t) = \frac{d_i}{n_i(t_{i+1} - t_i)}.$$

   This is referred to as the K-M type estimate. It estimates the rate of death per unit time in the interval $[t_i, t_{i+1})$.

3. Examples with the AML data:

$$\widetilde{h}(23) = \frac{1}{7} = .143$$

$$\widehat{h}(26) = \widehat{h}(23) = \frac{1}{7 \cdot (31 - 23)} = .018$$

Estimates of $H(\cdot)$, cumulative hazard to time t :

1. Constructed with K-M:

$$\widehat{H}(t) = -\log\left(\widehat{S}(t)\right) = -\log\prod_{y_{(i)}\leq t}\left(\frac{n_i - d_i}{n_i}\right),$$

$$\widehat{\text{var}}\left(\widehat{H}(t)\right) = \sum_{y_{(i)}\leq t}\frac{d_i}{n_i(n_i - d_i)}\,.$$

2. Nelson-Aalen estimate (1972, 1978):

$$\widetilde{H}(t) = \sum_{y_{(i)}\leq t}\frac{d_i}{n_i}\,,$$

$$\widehat{\text{var}}\left(\widetilde{H}(t)\right) = \sum_{y_{(i)}\leq t}\frac{d_i}{n_i^2}\,.$$

The Nelson-Aalen estimate is the cumulative sum of estimated conditional probabilities of death from I_1 through I_k where $t_k \leq t < t_{k+1}$. This estimate is the first order Taylor approximation to the first estimate. To see this let $x = d_i/n_i$ and expand $\log(1 - x)$ about $x = 0$.

3. Examples with the AML data:

$$\widehat{H}(26) = -\log\left(\widehat{S}(26)\right) = -\log(.614) = .488$$

$$\widetilde{H}(26) = \frac{1}{11} + \frac{1}{10} + \frac{1}{8} + \frac{1}{7} = .4588$$

- The function `hazard.km` takes a `survfit` object for its argument.

For the maintained group:

```
> hazard.km(km.fit)
   time ni di  hihat hitilde   Hhat se.Hhat Htilde se.Htilde
 1     9 11  1 0.0227  0.0909 0.0953  0.0953 0.0909    0.0909
 2    13 10  1 0.0200  0.1000 0.2007  0.1421 0.1909    0.1351
 3    18  8  1 0.0250  0.1250 0.3342  0.1951 0.3159    0.1841
 4    23  7  1 0.0179  0.1429 0.4884  0.2487 0.4588    0.2330
 5    31  5  1 0.0667  0.2000 0.7115  0.3345 0.6588    0.3071
 6    34  4  1 0.0179  0.2500 0.9992  0.4418 0.9088    0.3960
 7    48  2  1     NA  0.5000 1.6923  0.8338 1.4088    0.6378
```

Hazard Ratio

For two treatment groups, say 0 and 1, their **hazard ratio** (HR) is

$$\text{HR}(t|1, 0) = \frac{h(t|1)}{h(t|0)}.$$

- **The HR is a numeric measure that describes the treatment effect on survival over time.**

- We say their **hazards are proportional** (PH) when

  $\text{HR}(t|1, 0) = k > 0,$ constant over follow-up time.

- Crossing hazard curves $\Rightarrow$ hazards are **not** PH.

- The function `g.emphaz` produces two plots, one for each of the two types of hazard estimates.

```
> Surv0 <- Surv(aml$weeks[aml$group==0],
                aml$status[aml$group==0])
> Surv1 <- Surv(aml$weeks[aml$group==1],
                aml$status[aml$group==1])
> data <- list(Surv1,Surv0)
> g.emphaz(data,text="ht",main="hitilde",
       legend=c("maintained","nonmaintained"))
> g.emphaz(data,text="ht",main="hihat",
       legend=c("maintained","nonmaintained"))
```

[[1]]: maintained

time	ht	hhat
1	9	0.091	0.023
2	13	0.100	0.020
3	18	0.125	0.025
4	23	0.143	0.018
5	31	0.200	0.067
6	34	0.250	0.018
7	48	0.500	0.018

[[2]]: nonmaintained

time	ht	hhat
1	5	0.167	0.056
2	8	0.200	0.050
3	12	0.125	0.011
4	23	0.167	0.042
5	27	0.200	0.067
6	30	0.250	0.083
7	33	0.333	0.033
8	43	0.500	0.250
9	45	1.000	0.250

$$\frac{\widehat{h}_{\mathrm{nm}}(15)}{\widehat{h}_{\mathrm{m}}(15)} = \frac{.011}{.020} = .55 \qquad \text{and} \qquad \frac{\widehat{h}_{\mathrm{nm}}(25)}{\widehat{h}_{\mathrm{m}}(25)} = \frac{.042}{.018} = 2.33 \; .$$

The nonmaintained group has 55% of the risk of those maintained of relapsing at 15 weeks. However, on the average, those nonmaintained have 2.33 times the risk of those maintained of relapsing at 25 weeks.

hitilde   hihat
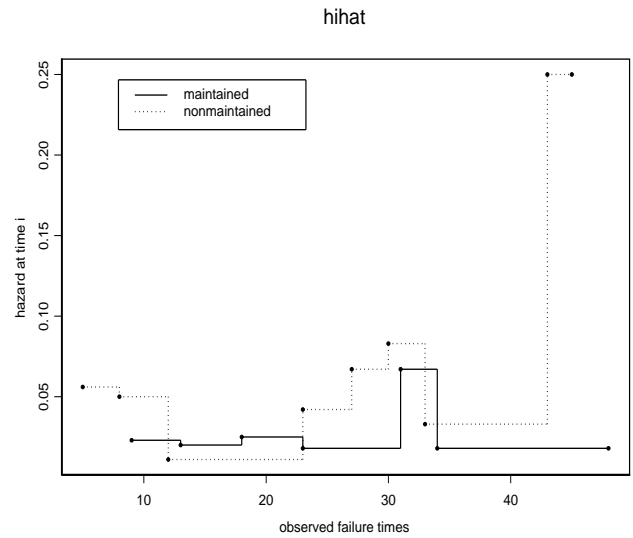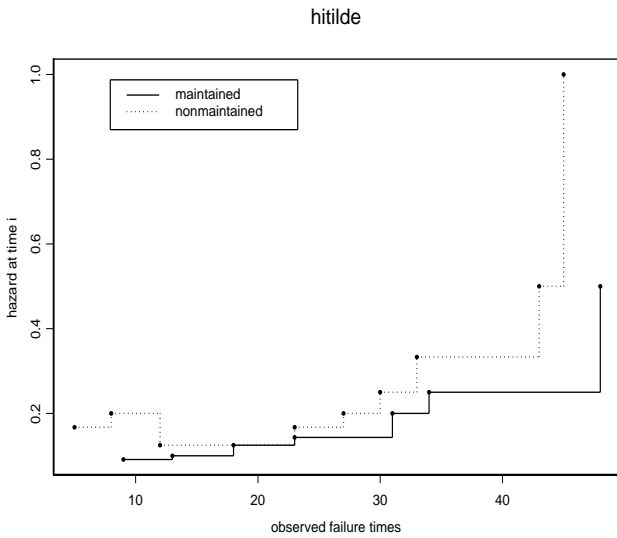
With larger datasets these plots will be chaotic. A far more useful picture is provided by:

The **kernel estimator** of $h(t)$ is given by

$$\widetilde{h}^{kernel}(t) = \frac{1}{b} \sum_{i=1}^{n'} \mathcal{K}\left(\frac{t - y_{(i)}}{b}\right) \frac{d_i}{n_i}.$$

The *kernel function* $\mathcal{K}$ is a bounded function which vanishes outside $[-1, 1]$ and has integral 1. The *bandwidth* or *window size* b is a positive parameter. It **smoothes** the increments d_i/n_i of the Nelson-Aalen estimator $\widetilde{H}(t)$. One frequently used kernel is the Epanechnikov kernel $\mathcal{K}(t) = 0.75(1 - t^2)$, $|t| \leq 1$. Another is the biweight kernel $\mathcal{K}(t) = (15/16)(1 - t^2)^2$, $|t| \leq 1$.

The essential pieces of R code follow: Let $g = 0, 1$.

```
> fit.g <- summary(survfit(Surv(weeks,status),
        subset=group==g,conf.type="n",data=aml),censor=T)
> u.g <- fit.g$time
> weight.g <- fit.g$n.event/fit.g$n.risk
> smooth.g <- density(u.g,kernel="epanechnikov",
                weights=weight.g,n=50,from=0,to=50)
> plot(smooth.g$x,smooth.g$y,type="l",...)
```

Kernel Estimates of Hazard: AML data
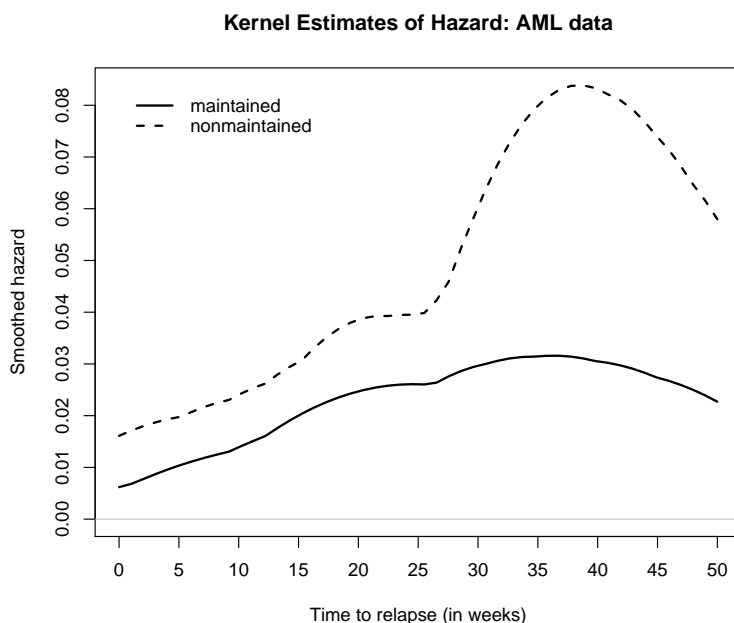
Smoothed estimates, $\widetilde{h}_g^{kernel}(t)$, $g = 0, 1$, of hazards. The Epanechnikov kernel $\mathcal{K}(t) = 0.75(1 - t^2), |t| \leq 1$ was used.

33

Ratio of Smoothed Hazards: AML data
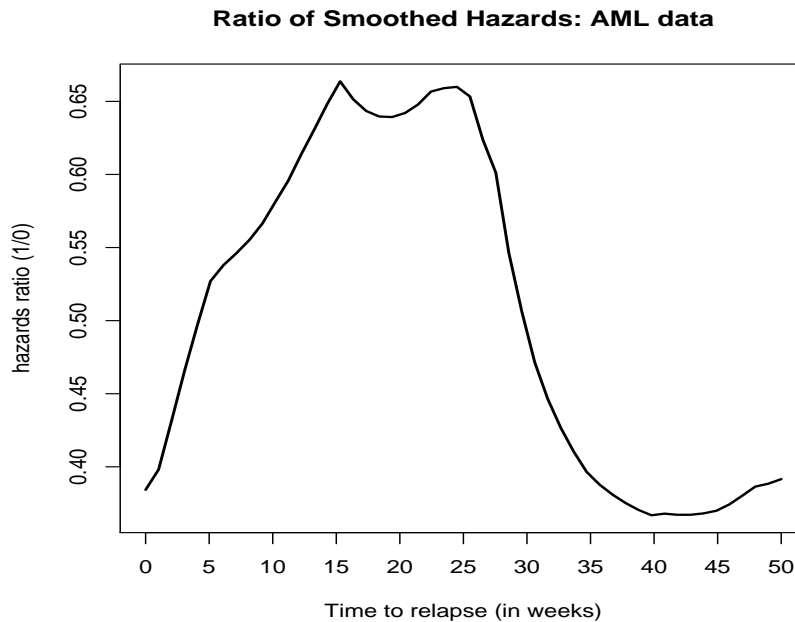
At 15 weeks we estimate the maintained group has about 66% of the risk of those nonmaintained of relapsing; or, those nonmaintained have 1.52 times the risk of those maintained of relapsing at 15 weeks. At 25 weeks the risk is slightly higher.

The plot is only an illustration of how to visualize and interpret HR's. Statistical accuracy (confidence bands) should be incorporated as these comments may not be statistically significant. Pointwise 95% bootstrap confidence limits for the log-HR are commonly reported.

Stratifying on a covariate:

- Stratifying on a particular covariate is one method that can account for its possible confounding and/or interaction effects with the treatment of interest on the response.

- Such effects **can mask the "true" effects of the treatment of interest.**

  Thus, stratification can provide us with stronger (or weaker) evidence, or more importantly, reverse the sign of the effect. That is, it is possible for the aggregated data to suggest treatment is favorable when in fact, in every subgroup, it is highly unfavorable; and vice versa. This is known as **Simpson's paradox** (Simpson, 1951).

Suppose we knew the gender of the patients in the AML study. We could stratify to see if there is a masked effect.

```
> attach(aml)
> survfit(Surv(weeks,status)~group+strata(gender))
  # Gives four separate survival probability tables.
  # plot(....) produces four K-M curves on the same plot.

> survdiff(Surv(weeks,status)~group+strata(gender))
  # Conducts the log-rank test pooled over the two strata.
> detach()
```

Chapter 3

Parametric Models and Methods

Models:

- Weibull (includes the exponential model)

- log-normal

- log-logistic

- gamma

Remark:

Except for the gamma distribution, these lifetime distributions have the property that the distribution of the log-transform $\log(T)$ is a member of the **location and scale family of distributions**.

Summary:

The common features are:

- The time T distributions have two parameters —
$$\text{scale} = \lambda > 0 \quad \text{and} \quad \text{shape} = \alpha > 0 \,.$$

- In log-time, $Y = \log(T)$, the distributions have two parameters —
$$\text{location} = \mu = -\log(\lambda) \quad \text{and} \quad \text{scale} = \sigma = \frac{1}{\alpha} \,.$$

- Each can be expressed in the form
$$Y = \log(T) = \mu + \sigma Z \,,$$
where Z is the standard member; that is,
$$\mu = 0 \ (\lambda = 1) \quad \text{and} \quad \sigma = 1 \ (\alpha = 1) \,.$$

- They are log-linear models.

The three distributions are summarized as follows:

T		$Y = \log(T)$
Weibull	$\Longleftrightarrow$	extreme minimum value
log-normal	$\Longleftrightarrow$	normal
log-logistic	$\Longleftrightarrow$	logistic

$$y_p = \log(t_p) = \mu + \sigma z_p \,,$$

t_p quantile	$y_p = \log(t_p)$ quantile	form of standard quantile z_p
Weibull	extreme value	$\log(-\log(S(t_p)))$
		$= \log(H(t_p))$
		$= \log(-\log(1-p))$
log-normal	normal	$\Phi^{-1}(p)$, where Φ denotes the standard normal d.f.
log-logistic	logistic	$-\log\left(\dfrac{S(t_p)}{1-S(t_p)}\right)$
		$= -\log(\text{odds})$
		$= -\log\left(\dfrac{1-p}{p}\right)$

Weibull Distribution

p.d.f. $f(t)$	survivor $S(t)$	hazard $h(t)$
$\lambda\alpha(\lambda t)^{\alpha-1}\times$ $\exp\left(-(\lambda t)^{\alpha}\right)$	$\exp\left(-(\lambda t)^{\alpha}\right)$	$\lambda\alpha(\lambda t)^{\alpha-1}$
mean $E(T)$	variance $Var(T)$	pth-quantile t_p
$\lambda^{-1}\Gamma(1+\frac{1}{\alpha})$	$\lambda^{-2}\Gamma(1+\frac{2}{\alpha})$ $-\lambda^{-2}(\Gamma(1+\frac{1}{\alpha}))^2$	$\lambda^{-1}\left(-\log(1-p)\right)^{\frac{1}{\alpha}}$ $\lambda > 0$ and $\alpha > 0$

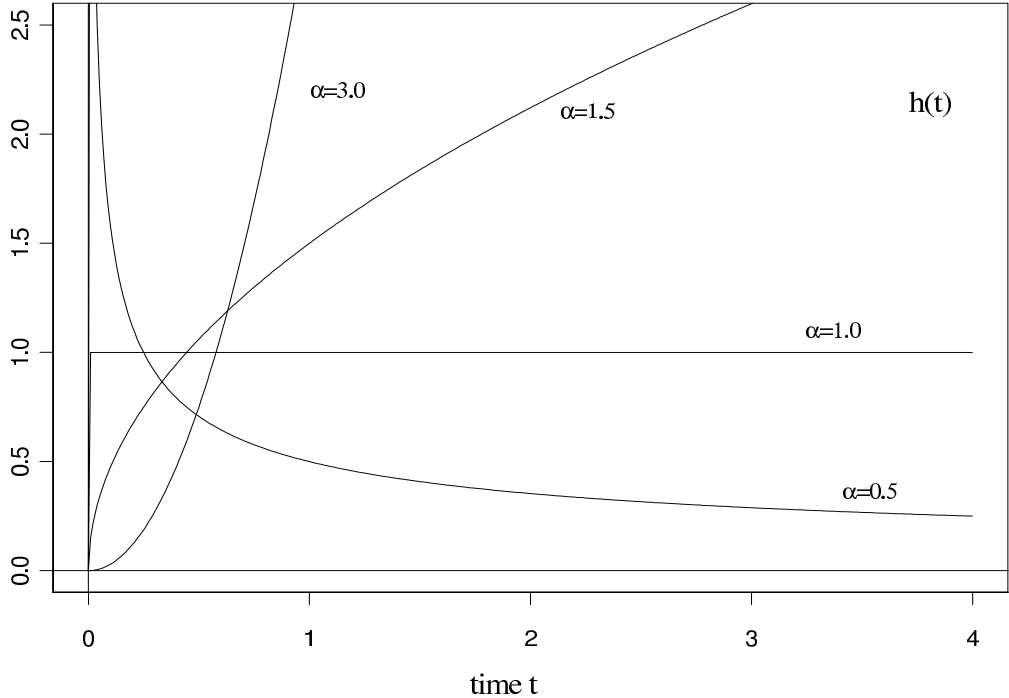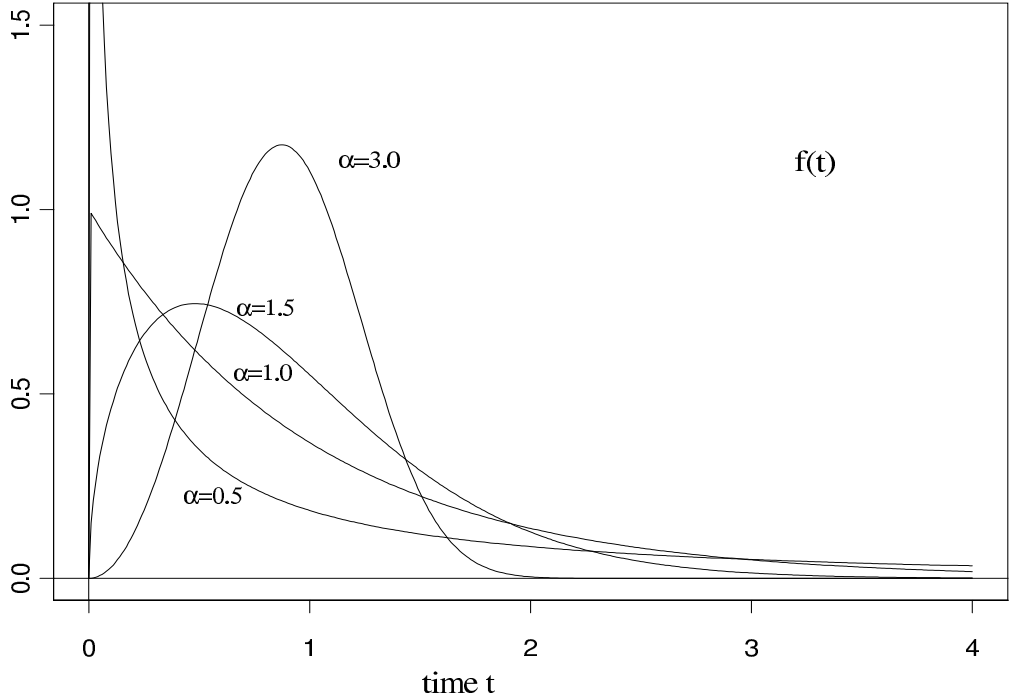$\Gamma(k) = \int_0^\infty u^{k-1}e^{-u}du,\ k > 0.$

The form of the survivor function $S(t)$ gives us

$$\log(t) = -\log(\lambda) + \sigma\log\left(-\log(S(t))\right)$$
$$= -\log(\lambda) + \sigma\log\left(H(t)\right),$$

where $\sigma = 1/\alpha$.

• When $\alpha = 1$, the Weibull is just the **exponential distribution**.

Density and hazard curves for Weibull model with $\lambda = 1$.
Exponential curves correspond to $\alpha = 1$.

Remarks:

- $h(t)$ is monotone increasing when $\alpha > 1$, decreasing when $\alpha < 1$, and constant for $\alpha = 1$. The parameter α is called the shape parameter as the shape of the p.d.f., and hence the other functions, depends on the value of α.

- The λ is a scale parameter in that the effect of different values of λ is just to change the scale on the horizontal (t) axis, not the basic shape of the graph.

- An increasing Weibull hazard may be useful for modelling survival times of leukemia patients not responding to treatment, where the event of interest is death. As survival time increases for such a patient, and as the prognosis accordingly worsens, the patient's potential for dying of the disease also increases.

- A decreasing Weibull hazard may well model the death times of patients recovering from surgery. The potential for dying after surgery usually decreases as the time post-surgery increases, at least for a while.

- Plot of log-time against the log-cumulative hazard is a straight line with slope $\sigma = 1/\alpha$ and y-intercept $\mu = -\log(\lambda)$. Use for a graphical check.

Q-Q plot – diagnostic tool for model adequacy

Recall: $y_p = \log(t_p) = \mu + \sigma z_p$

- t_i, $i = 1, \ldots, r \leq n$, denote the ordered uncensored observed failure times.

- Use K-M to get estimated failure probabilities. That is, get

$$\widehat{p}_i = 1 - \widehat{S}(t_i)$$

  for each $y_i = \log(t_i)$.

- Get the parametric standard quantile z_i by

$$F_{0,1}(z_i) = P(Z \leq z_i) = \widehat{p}_i,$$

  where $F_{0,1}$ is the d.f. of the standard parametric model ($\mu = 0, \sigma = 1$) under consideration.

- Plot the points (z_i, y_i). If the proposed model is adequate, points should lie close to a straight line with slope σ and y-intercept μ.

- An appropriate line to compare the plot pattern to is one with the maximum likelihood estimates $\widehat{\sigma}$ and $\widehat{\mu}$ – **the MLE line**; that is, $\widehat{y}_p = \widehat{\mu} + \widehat{\sigma} z_p$.

- Equivalently, we can plot the points (z_i, e_i) where the e_i is the ith ordered residual

$$e_i = \frac{y_i - \widehat{\mu}}{\widehat{\sigma}}$$

and z_i is the corresponding log-parametric standard quantile of either the Weibull, log-normal, or log-logistic distribution. If the model under study is appropriate, the points (z_i, e_i) should lie very close to the 45°-line through the origin.

The function `qq.reg.resid.r` draws this plot. This function is also useful for checking adequacy of a regression model.

Example: AML maintained group

- The R function `qweibull(p,`α`,`λ^{-1}`)` computes quantiles for the Weibull. Take log to get the extreme value quantiles.

t_i	y_i	K-M(t_i)	$\widehat{p}_i$	z_i
	$\log(t_i)$			$\log(\text{qweibull}(\widehat{p}_i,1,1))$
9	2.197	.909	.091	-2.35
13	2.565	.818	.182	-1.605

Plot the points $(2.197, -2.35)$, $(2.565, -1.605)$, etc.

43

Maximum Likelihood Estimation (MLE)

- **A generic likelihood function:**

$$L(\theta) = L(\theta|t_1, \cdots, t_n) = \prod_{i=1}^{n} f(t_i|\theta)$$

- **maximum likelihood estimator** (MLE), denoted by $\widehat{\theta}$, is the value of θ in Ω that maximizes $L(\theta)$ or, equivalently, maximizes the log-likelihood

$$\log L(\theta) = \sum_{i=1}^{n} \log f(t_i|\theta).$$

- MLE's possess the **invariance property**; that is,

$$\widehat{\tau(\theta)} = \tau(\widehat{\theta}).$$

- **Fisher information matrix $I(\theta)$:**

Let θ be a $d \times 1$ vector parameter.

$$I(\theta) = \left(\left(-E(\frac{\partial^2}{\partial\theta_j \partial\theta_k} \log L(\theta)) \right) \right),$$

where E denotes expectation. $I(\theta)$ is a $d \times d$ matrix.

- The MLE $\widehat{\theta}$ has the following **large sample distribution**:

$$\widehat{\theta} \overset{a}{\sim} \textbf{MVN}(\theta, I^{-1}(\theta)),$$

where **MVN** denotes multivariate normal and $\overset{a}{\sim}$ is read "is asymptotically distributed."

- When θ is a scalar,

$$\text{var}_a(\widehat{\theta}) = \frac{1}{I(\theta)},$$

where $I(\theta) = -E(\partial^2 \log L(\theta)/\partial\theta^2)$.

- **observed information matrix $i(\theta)$:**

$$i(\theta) = \left(\left(-\frac{\partial^2}{\partial\theta_j \partial\theta_k} \log L(\theta) \right) \right)$$

evaluated at the MLE $\widehat{\theta}$ approximates $I(\theta)$

- For the univariate case,

$$i(\theta) = -\frac{\partial^2 \log L(\theta)}{\partial\theta^2} \, .$$

- Hence, $\text{var}_a(\widehat{\theta})$ is approximated by $(i(\widehat{\theta}))^{-1}$.

The **delta method** is useful for
1. obtaining limiting distributions of smooth functions of an MLE.
2. removing the parameter of interest from the variance of an estimator. This removes this source of variation which can yield more efficient estimators; i.e., narrower confidence intervals. This is called variance-stabilization.

We describe it for the univariate case.

Delta method:

Suppose a random variable Z has a mean μ and variance σ^2 and suppose we want to approximate the distribution of some function $g(Z)$. Take a first order Taylor expansion of $g(Z)$ about μ and ignore the higher order terms to get

$$g(Z) \approx g(\mu) + (Z - \mu)g'(\mu).$$

Then the mean$(g(Z)) \approx g(\mu)$ and the var$(g(Z)) \approx (g'(\mu))^2 \sigma^2$.

Furthermore, if

$$Z \overset{a}{\sim} \text{normal}(\mu, \sigma^2),$$

then

$$g(Z) \overset{a}{\sim} \text{normal}(g(\mu), (g'(\mu))^2 \sigma^2).$$

Confidence Intervals and Tests

An approximate $(1 - \alpha) \times 100\%$ confidence interval for the parameter θ is given by

$$\widehat{\theta} \pm z_{\frac{\alpha}{2}} \text{s.e.}(\widehat{\theta})\,,$$

where $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ quantile of the standard normal distribution and

$$\text{s.e.}(\widehat{\theta}) = \sqrt{\text{var}_a(\widehat{\theta})} \approx \sqrt{-\left(\partial^2 \log L(\theta)/\partial\theta^2\right)^{-1}} = \sqrt{(i(\widehat{\theta}))^{-1}}.$$

Likelihood ratio test:

To test $H_0 : \theta \in \Omega_0$ (reduced model) against $H_A : \theta \in \Omega_0^c$ (full model), use

$$r(\underline{x}) = \frac{\sup_{\Omega_0} L(\theta)}{\sup_{\Omega} L(\theta)}.$$

Note that $r(\underline{x}) \leq 1$.

• This handles hypotheses with nuisance parameters. Suppose $\theta = (\theta_1, \theta_2, \theta_3)$.

For example $H_0 : (\theta_1 = 0, \theta_2, \theta_3)$ against $H_A : (\theta_1 \neq 0, \theta_2, \theta_3)$. Here θ_2 and θ_3 are nuisance parameters.

• Most often, finding the sup amounts to finding the MLE's and then evaluating $L(\theta)$ at the MLE.

• Reject H_0 for small values. Or, equivalently, we reject H_0 for large values of

$$r^*(\underline{x}) = -2 \log r(\underline{x}).$$

• $\quad r^*(\underline{x}) \quad \overset{a}{\sim} \quad \chi^2_{(df)}.$

One-Sample Problem

Fitting data to the exponential model:

Let u, c, and n_u denote uncensored, censored, and number of un-censored observations, respectively. The n observed values are now represented by the vectors $\underline{y}$ and $\underline{\delta}$, where $\underline{y}' = (y_1, \ldots, y_n)$ and $\underline{\delta}' = (\delta_1, \ldots, \delta_n)$. Then

- **Likelihood:**

$$
\begin{aligned}
L(\lambda) &= \prod_u f(y_i|\lambda) \cdot \prod_c S_f(y_i|\lambda) \\
&= \prod_u \lambda \exp(-\lambda y_i) \prod_c \exp(-\lambda y_i) \\
&= \lambda^{n_u} \exp\left(-\lambda \sum_u y_i\right) \exp\left(-\lambda \sum_c y_i\right) \\
&= \lambda^{n_u} \exp\left(-\lambda \sum_{i=1}^n y_i\right)
\end{aligned}
$$

- **Log-likelihood:**

$$
\log L(\lambda) = n_u \log(\lambda) - \lambda \sum_{i=1}^n y_i
$$

The MLE $\hat{\lambda}$ solves
$$
\frac{\partial \log L(\lambda)}{\partial \lambda} = \frac{n_u}{\lambda} - \sum_{i=1}^n y_i = 0
$$

$$
\frac{\partial^2 \log L(\lambda)}{\partial \lambda^2} = -\frac{n_u}{\lambda^2} = -i(\lambda)
$$

- **MLE:**

$$\widehat{\lambda} = \frac{n_u}{\sum_{i=1}^n y_i} \qquad \text{and} \qquad \text{var}_a(\widehat{\lambda}) = \left(-E\left(-\frac{n_u}{\lambda^2}\right)\right)^{-1} = \frac{\lambda^2}{E(n_u)},$$

where $E(n_u) = n \cdot P(T \leq C)$.

$$\frac{\widehat{\lambda} - \lambda}{\sqrt{\lambda^2/E(n_u)}} \overset{a}{\sim} N(0, 1).$$

Since $E(n_u)$ depends on G, **use the observed information** $i(\lambda)$ evaluated at $\widehat{\lambda}$ for the variance.

$$\text{var}_a(\widehat{\lambda}) \approx \frac{1}{i(\widehat{\lambda})} = \frac{\widehat{\lambda}^2}{n_u},$$

where $i(\lambda) = n_u/\lambda^2$.

By **invariance property**, the MLE for the mean $\theta = 1/\lambda$ is $\widehat{\theta} = 1/\widehat{\lambda} = \sum_{i=1}^n y_i/n_u$.

On the AML data, $n_u = 7$,

$$\widehat{\lambda} = \frac{7}{423} = 0.0165, \qquad \text{and} \qquad \text{var}_a(\widehat{\lambda}) \approx \frac{\widehat{\lambda}^2}{7} = \frac{0.0165^2}{7}.$$

- **A 95% C.I. for λ** is given by

$$\widehat{\lambda} \pm z_{0.025} \cdot \text{se}(\widehat{\lambda}) =: 0.0165 \pm 1.96 \cdot \frac{0.0165}{\sqrt{7}} =: [0.004277 \, , \, 0.0287].$$

- Let's use the **delta method**. Take $g(\lambda) = \log(\lambda)$. Then $g'(\lambda) = 1/\lambda$. The delta method tells us

$$\text{var}_a(\log(\widehat{\lambda})) = (g'(\lambda))^2 \times \text{var}_a(\widehat{\lambda})$$

$$\approx \frac{1}{\lambda^2} \times \frac{1}{i(\lambda)}$$

$$= \frac{1}{\lambda^2} \times \frac{\lambda^2}{n_u}$$

$$= \frac{1}{n_u}, \quad \text{which is free of } \lambda.$$

and

$$\log(\widehat{\lambda}) \overset{a}{\sim} N\left(\log(\lambda), \frac{1}{n_u}\right).$$

A 95% C.I. for $\log(\lambda)$ is given by

$$\log(\widehat{\lambda}) \pm 1.96 \cdot \frac{1}{\sqrt{n_u}}$$

$$\log\left(\frac{7}{423}\right) \pm 1.96 \cdot \frac{1}{\sqrt{7}}$$

$$[-4.84, -3.36].$$

Transform back by taking exp(endpoints). This second 95% C.I. for λ is

$$[.0079, .0347],$$

which is slightly wider than the previous interval for λ.

Invert and reverse endpoints to obtain a **95% C.I. for the mean θ**. This yields $[28.81, 126.76]$ weeks.

Analogously,

$$\log(\widehat{\theta}) \quad \overset{a}{\sim} \quad N\left(\log(\theta), \frac{1}{n_u}\right)$$

$$\log(\widehat{t_p}) \quad \overset{a}{\sim} \quad N\left(\log(t_p), \frac{1}{n_u}\right).$$

Analogously, we first construct C.I.'s for the log(parameter), then take exp(endpoints) to obtain C.I.'s for the parameter.

Most statisticians prefer this approach.

Using the AML data, 95% C.I.'s are summarized:

parameter	point estimate	log(parameter)	parameter	
mean	60.43 weeks	[3.361, 4.84]	[28.82, 126.76]	weeks
median	41.88 weeks	[2.994, 4.4756]	[19.965, 87.85]	weeks

The S/R functions we use compute these preferred intervals.

- **The likelihood ratio test:**

  Test $H_0 : \theta = 1/\lambda = 30$ against $H_A : \theta = 1/\lambda \neq 30$:

$$
\begin{aligned}
r^*(\underline{y}) &= -2 \log L(\lambda_0) + 2 \log L(\widehat{\lambda}) \\
&= -2n_u \log(\lambda_0) + 2\lambda_0 \sum_{i=1}^{n} y_i + 2n_u \log \left(\frac{n_u}{\sum_{i=1}^{n} y_i} \right) - 2n_u \\
&= -2 \cdot 7 \cdot \log(\frac{1}{30}) + \frac{2}{30} \cdot 423 + 2 \cdot 7 \cdot \log \left(\frac{7}{423} \right) - 2 \cdot 7 \\
&= 4.396.
\end{aligned}
$$

  The tail area under the $\chi^2_{(1)}$ density curve approximates the p-value. The p-value $= P(r^*(\underline{y}) \geq 4.396) \approx 0.036$. Therefore, here we reject $H_0 : \theta = 1/\lambda = 30$ and conclude that mean survival is > 30 weeks.

  In S/R, use the following code to obtain the p-value:
  ```
  > 1-pchisq(4.396,1)
  ```

53

S/R Application

The S/R function **survreg** fits parametric models with the MLE approach.

• By default **survreg** uses a **log link** function which transforms the problem into estimating location $\mu = -\log(\lambda)$ and scale $\sigma = 1/\alpha$.

• In the output, $\widehat{\mu}$ is the **intercept value** and **Scale** $= \widehat{\sigma}$.

• Once the parameters are estimated via **survreg**, we can use S/R functions to compute estimated survival probabilities and quantiles. These functions are summarized below:

	Weibull	logistic ($Y = \log(T)$)	normal ($Y = \log(T)$)
$F(t)$	`pweibull(`q`, `α`, `λ^{-1}`)`	`plogis(`q`, `μ`, `σ`)`	`pnorm(`q`, `μ`, `σ`)`
t_p	`qweibull(`p`, `α`, `λ^{-1}` )`	`qlogis(`p`, `μ`, `σ`)`	`qnorm(`p`, `μ`, `σ`)`

Weibull fit

```
> aml1 <- aml[group==1, ]
> attach(aml1)
> weib.fit <- survreg(Surv(weeks,status)~1,dist="weib")
> summary(weib.fit)
              Value Std. Error       z         p
(Intercept)  4.0997      0.366  11.187 4.74e-029
 Log(scale) -0.0314      0.277  -0.113 9.10e-001
Scale= 0.969
```

$\hat{\lambda} = \exp(-\hat{\mu}) = \exp(-4.0997) = .0166.$
$\hat{\alpha} = 1/\hat{\sigma} = 1/(.969) = 1.032.$
# Estimated median along with a 95% C.I. (in weeks).

```
> medhat <- predict(weib.fit,type="uquantile",p=0.5,se.fit=T)
> medhat1 <- medhat$fit[1]
> medhat1.se <- medhat$se.fit[1]
> exp(medhat1)
[1] 42.28842
> C.I.median1 <- c(exp(medhat1),exp(medhat1-1.96*medhat1.se),
                   exp(medhat1+1.96*medhat1.se))
> names(C.I.median1) <- c("median1","LCL","UCL")
> C.I.median1
   median1      LCL       UCL
  42.28842  20.22064  88.43986
> qq.reg.resid.r(aml1,weeks,status,weib.fit,"qweibull",
          "standard extreme value quantile")  #Draws a Q-Q plot
```

Log-logistic fit

```
> loglogis.fit<-survreg(Surv(weeks,status)~1,dist="loglogistic")
> summary(loglogis.fit)
             Value Std. Error      z       p
(Intercept)  3.515      0.306  11.48 1.65e-030
 Log(scale) -0.612      0.318  -1.93 5.39e-002
Scale= 0.542
```

Estimated median along with a 95% C.I. (in weeks).

```
> medhat <- predict(loglogis.fit,type="uquantile",p=0.5,se.fit=T)
> medhat1 <- medhat$fit[1]
> medhat1.se <- medhat$se.fit[1]
> exp(medhat1)
[1] 33.60127
> C.I.median1 <- c(exp(medhat1),exp(medhat1-1.96*medhat1.se),
                   exp(medhat1+1.96*medhat1.se))
> names(C.I.median1) <- c("median1","LCL","UCL")
> C.I.median1
  median1       LCL       UCL
 33.60127  18.44077  61.22549
> qq.reg.resid.r(aml1,weeks,status,loglogis.fit,"qlogis",
                "standard logistic quantile")  #Draws a Q-Q plot
> detach()
```
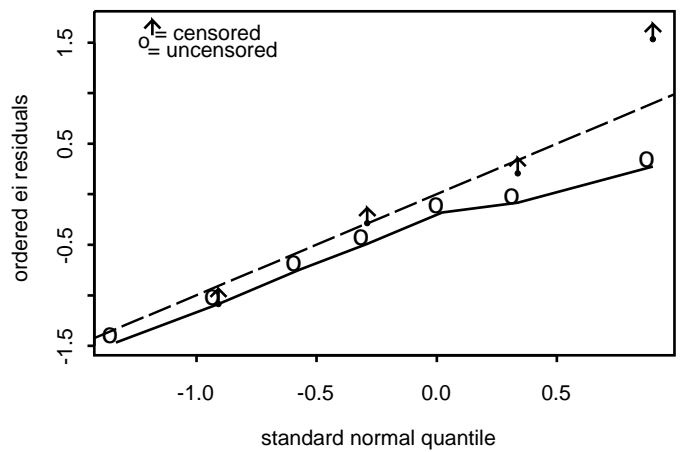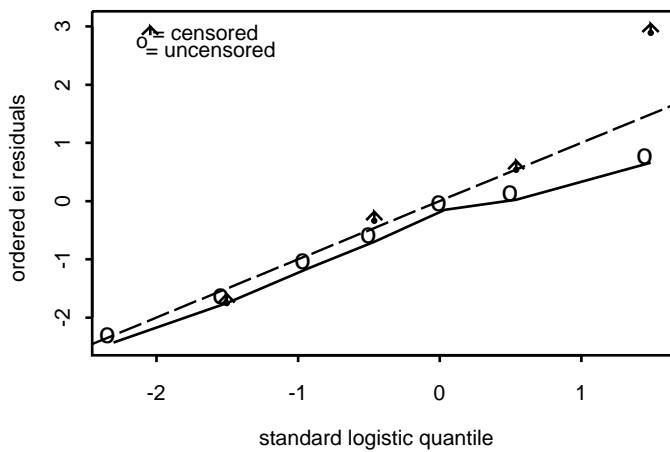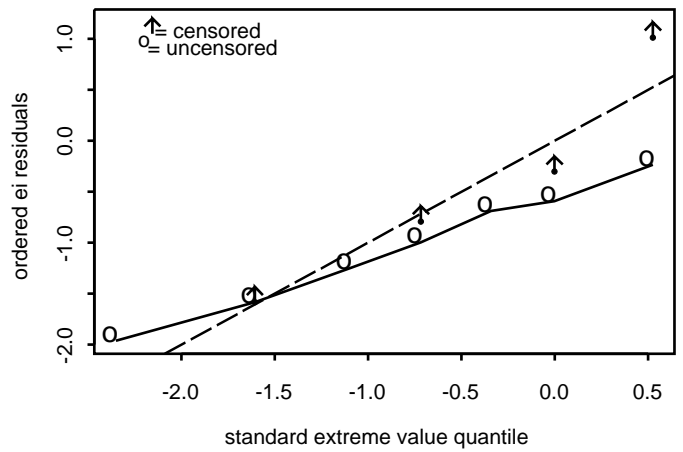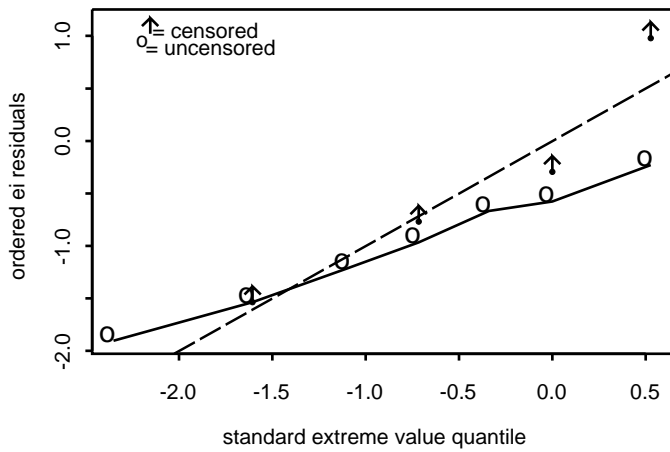
Q-Q plots of the ordered residuals $e_i = (y_i - \hat{\mu})/\hat{\sigma}$ where y_i denotes the log-data. Dashed line is the 45°-line through the origin. AML1 data.

Discussion

Summary table:

MLE's fit to AML1 data at the models:

model	$\widehat{\mu}$	median1	95% C.I.	$\widehat{\sigma}$
exponential	4.1	41.88	[19.97, 87.86] weeks	1
Weibull	4.1	42.29	[20.22, 88.44] weeks	.969
log-logistic	3.52	33.60	[18.44, 61.23] weeks	.542

The log-logistic gives the narrowest C.I. among the three. Its estimated median of 33.60 weeks is the smallest and very close to the K-M estimated median of 31 weeks.

The Q-Q plots are useful for distributional assessment. It "appears" that a log-logistic model fits adequately.

The estimated log-logistic survival curve is overlayed on the K-M curve for the AML1 data in the last figure.
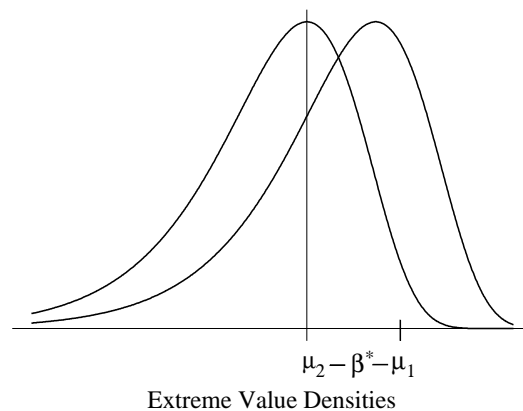

Survival curves for AML data

58
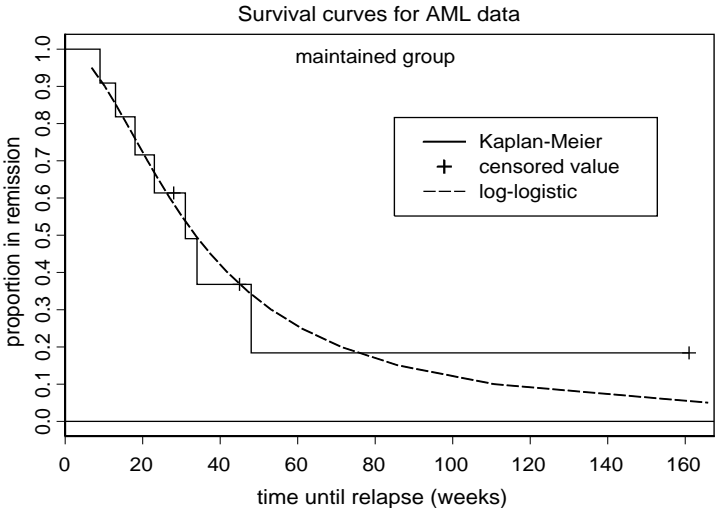
Two-Sample Problem

We compare two survival curves from the same parametric family.
- Focus on comparing scale parameters λ_1 and λ_2.
- In the log-transformed problem, this compares the two locations,

$$\mu_1 = -\log(\lambda_1) \text{ and } \mu_2 = -\log(\lambda_2).$$

$$\mu_2 - \beta^* - \mu_1$$

Extreme Value Densities

We explore if any of the log-transform distributions, which belong to the location and scale family, fit this data adequately.

- The **full model** can be expressed as a log-linear model as follows:

$$
\begin{aligned}
Y &= \log(T) \\
&= \widetilde{\mu} + \text{error} \\
&= \theta + \beta^* \text{group} + \text{error} \\
&= \begin{cases} \theta + \beta^* + \text{error} & \text{if group} = 1 \text{ (maintained)} \\ \theta + \text{error} & \text{if group} = 0 \text{ (nonmaintained)}. \end{cases}
\end{aligned}
$$

The $\widetilde{\mu}$ is called the **linear predictor**. In this two groups model, it has two values $\mu_1 = \theta + \beta^*$ and $\mu_2 = \theta$.

- We know $\widetilde{\mu} = -\log(\widetilde{\lambda})$, where $\widetilde{\lambda}$ denotes the scale parameter values of the distribution of the target variable T.

- Then $\widetilde{\lambda} = \exp(-\theta - \beta^* \text{group})$.

The two values are

$$\lambda_1 = \exp(-\theta - \beta^*) \ \text{ and } \ \lambda_2 = \exp(-\theta)$$

.

The **null hypothesis** is:

$H_0 : \lambda_1 = \lambda_2 \quad$ if and only if $\quad \mu_1 = \mu_2 \quad$ if and only if $\quad \beta^* = 0$.

Recall that the scale parameter in the log-transform model is the reciprocal of the shape parameter in the original model; that is, $\sigma = 1/\alpha$.

We test H_0 under each of the following cases:

> **Case 1:** Assume equal shapes (α); that is, we assume equal scales $\sigma_1 = \sigma_2 = \sigma$. Hence, error $= \sigma Z$, where the random variable Z has either the standard extreme value, standard logistic, or the standard normal distribution. Recall by standard, we mean $\mu = 0$ and $\sigma = 1$.

> **Case 2:** Assume different shapes; that is, $\sigma_1 \neq \sigma_2$.

S/R Application

Weibull fit:

Model 1: Data come from same distribution. **The Null Model** is $Y = \log(T) = \theta + \sigma Z$, where Z is a standard extreme value random variable.

```
> attach(aml)
> weib.fit0 <- survreg(Surv(weeks,status) ~ 1,dist="weib")
> summary(weib.fit0)
              Value Std. Error      z        p
(Intercept)  3.6425      0.217  16.780 3.43e-063
Scale= 0.912
Loglik(model)= -83.2   Loglik(intercept only)= -83.2
```

Model 2: Case 1: With different locations and equal scales σ, we express this model by

$$Y = \log(T) = \theta + \beta^* \text{group} + \sigma Z.$$

```
> weib.fit1 <- survreg(Surv(weeks,status) ~ group,dist="weib")
> summary(weib.fit1)
             Value Std. Error      z        p
(Intercept)  3.180      0.241  13.22 6.89e-040
      group  0.929      0.383   2.43 1.51e-002
Scale= 0.791
Loglik(model)= -80.5 Loglik(intercept only)= -83.2

    Chisq= 5.31 on 1 degrees of freedom, p= 0.021

> weib.fit1$linear.predictors # Extracts the estimated mutildes.

 4.1091 4.1091 4.1091 4.1091 4.1091 4.1091 4.1091 4.1091
 4.1091 4.1091 4.1091 3.1797 3.1797 3.1797 3.1797 3.1797
 3.1797 3.1797 3.1797 3.1797 3.1797 3.1797 3.1797

 # muhat1=4.109 and muhat2=3.18 for maintained and
 # nonmaintained groups respectively.
```

Model 3: Case 2: $Y = \log(T) = \theta + \beta^*\mathrm{group} + \mathrm{error}$, different locations, different scales.

Fit each group separately. On each group run a `survreg` to fit the data. This gives the MLE's of the two locations μ_1 and μ_2, and the two scales σ_1 and σ_2.

```
> weib.fit20 <- survreg(Surv(weeks,status) ~ 1,
               data=aml[group==0, ],dist="weib")

> weib.fit21 <- survreg(Surv(weeks,status) ~ 1,
               data=aml[group==1, ],dist="weib")


> summary(weib.fit20)
            Value  Std.Error      z          p
(Intercept)  3.222     0.198  16.25 2.31e-059
Scale=0.635

> summary(weib.fit21)
            Value  Std.Error      z          p
(Intercept)  4.1       0.366  11.19 4.74e-029
Scale=0.969
```

To test the reduced model against the full Model 2, we use the **LRT**. The `anova` function is appropriate for hierarchical models.

```
> anova(weib.fit0,weib.fit1,test="Chisq")
Analysis of Deviance Table Response: Surv(weeks, status)

   Terms Resid. Df    -2*LL Test Df  Deviance      Pr(Chi)
1      1        21   166.3573
2 group         20   161.0433    1  5.314048  0.02115415


 # Model 2 is a significant improvement over the null
 # model (Model 1).
```

To construct the appropriate likelihood function for Model 3 to be used in the LRT:

```
> loglik3 <- weib.fit20$loglik[2]+weib.fit21$loglik[2]
> loglik3
[1] -79.84817

> lrt23 <- -2*(weib.fit1$loglik[2]-loglik3)
> lrt23
[1] 1.346954

> 1 - pchisq(lrt23,1)
[1] 0.2458114  # Retain Model 2.
```

The following table summarizes the **three models weib.fit0, 1, and 2**:

Model	Calculated Parameters	The Picture
1 (0)	θ, σ	same location, same scale
2 (1)	θ, β^*, $\sigma \equiv \mu_1$, μ_2, σ	different locations, same scale
3 (2)	μ_1, μ_2, σ_1, σ_2	different locations, different scales

We now use the log-logistic and log-normal distribution to estimate Model 2. **The form of the log-linear model is the same**. **The distribution of error terms is what changes**.

$$Y = \log(T) = \theta + \beta^*\text{group} + \sigma Z,$$

where $Z \sim$ standard logistic or standard normal.

```
> loglogis.fit1 <- survreg(Surv(weeks,status) ~ group,
                           dist="loglogistic")
> summary(loglogis.fit1)


            Value Std. Error      z         p
(Intercept) 2.899      0.267  10.84 2.11e-027
      group 0.604      0.393   1.54 1.24e-001
Scale= 0.513

Loglik(model)= -79.4   Loglik(intercept only)= -80.6

Chisq= 2.41 on 1 degrees of freedom, p= 0.12 # p-value of LRT.

   # The LRT is test for overall model adequacy. It is not
   # significant.

> lognorm.fit1 <- survreg(Surv(weeks,status) ~ group,
                          dist="lognormal")
> summary(lognorm.fit1)


            Value Std. Error       z         p
(Intercept) 2.854      0.254  11.242 2.55e-029
      group 0.724      0.380   1.905 5.68e-002
Scale= 0.865

Loglik(model)= -78.9   Loglik(intercept only)= -80.7

Chisq= 3.49 on 1 degrees of freedom, p= 0.062 # p-value of LRT.

    # Here there is mild evidence of the model adequacy.
```

Summary:

Summary of the distributional fits to Model 2:

distribution	max. log-likeli $L(\widehat{\theta}, \widehat{\beta^*})$	p-value for model adequacy	$\widehat{\theta}$	$\widehat{\beta^*}$	p-value for group effect
Weibull	-80.5	0.021	3.180	0.929	0.0151
log-logistic	-79.4	0.12	2.899	0.604	0.124
log-normal	-78.9	0.062	2.854	0.724	0.0568

```
  # The Q-Q plots are next.
> t.s0 <- Surv(weeks[group==0],status[group==0])
> t.s1 <- Surv(weeks[group==1],status[group==1])
> qq.weibreg(list(t.s0,t.s1),weib.fit1)
> qq.loglogisreg(list(t.s0,t.s1),loglogis.fit1)
> qq.lognormreg(list(t.s0,t.s1),lognorm.fit1)
```
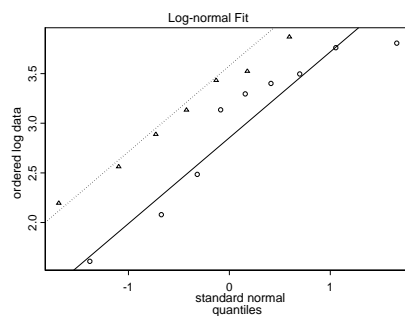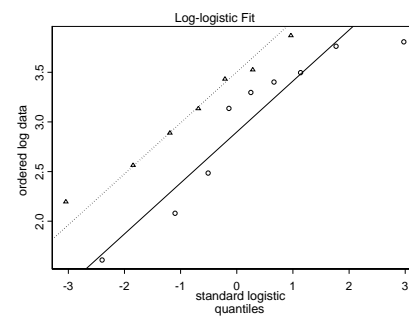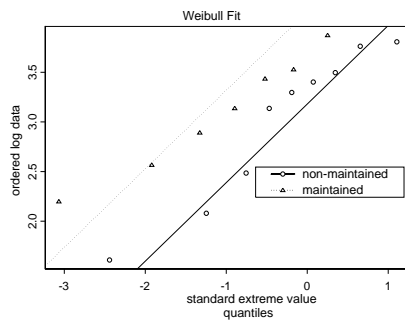

Model 2: $\widehat{y}_p = \widehat{\theta} + \widehat{\beta^*}\text{group} + \widehat{\sigma}Z_p.$

Prelude to Parametric Regression Models

Let's explore Model 2 under the assumption that $T \sim$ Weibull.

$$
\begin{aligned}
Y &= \log(T) \\
&= \theta + \beta^* \text{group} + \sigma Z \\
&= \widetilde{\mu} + \sigma Z \,,
\end{aligned}
$$

where Z is a standard extreme minimum value random variable.

• The linear predictor $\widetilde{\mu} = -\log(\widetilde{\lambda})$

• $\sigma = 1/\alpha$.

• The hazard function for the Weibull in this context is expressed as

$$
\begin{aligned}
h(t|\text{group}) &= \alpha \widetilde{\lambda}^{\alpha} t^{\alpha-1} \\
&= \alpha \lambda^{\alpha} t^{\alpha-1} \exp(\beta \text{group}) \\
&= h_0(t) \exp(\beta \text{group}) \,,
\end{aligned}
$$

when we set $\lambda = \exp(-\theta)$ and $\beta = -\beta^*/\sigma$. WHY!

• $h_0(t)$ denotes the baseline hazard, which is free of the covariate group.

The hazard ratio (HR) of group 1 to group 0 is

$$
\text{HR} = \frac{h(t|1)}{h(t|0)} = \frac{\exp(\beta)}{\exp(0)} = \exp(\beta) \,.
$$

If we believe the Weibull model is appropriate, the HR is **constant** over follow-up time t; that is, the graph of HR is a horizontal line with height $\exp(\beta)$. We say the Weibull enjoys the **proportional hazards property**.

On the AML data,

$$\widehat{\beta} = \frac{-\widehat{\beta^*}}{\widehat{\sigma}} = \frac{-0.929}{0.791} = -1.1745 \ .$$

Therefore, the estimated HR is

$$\widehat{\text{HR}} = \frac{\widehat{h}(t|1)}{\widehat{h}(t|0)} = \exp(-1.1745) \approx 0.31 \ .$$

The maintained group has 31% of the risk of the control group's risk of relapse. Or, the control group has (1/0.31)=3.23 times the risk of the maintained group of relapse at any given time t.

The HR is a measure of effect that describes the relationship between time to relapse and group.

If we consider the ratio of the estimated survival probabilities, say at $t = 31$ weeks, since $\widehat{\lambda} = \exp(-\widehat{\mu})$, we obtain

$$\frac{\widehat{S}(31|1)}{\widehat{S}(31|0)} = \frac{0.652}{0.252} \approx 2.59 \ .$$

The maintained group is 2.59 times more likely to stay in remission

at least 31 weeks.