

Stat 486 / 886 Survival Analysis

Chapter 4. Comparison of Failure Time Distributions

Professor: Wenyu Jiang

Department of Mathematics and Statistics
Queen's University

Acknowledgement: Profs. Jerry Lawless, David Matthews (Waterloo),
Profs. Bingshu Chen, Paul Peng, Dongsheng Tu (Queen's)

Sections of Chapter 4

1 Parametric Comparison

2 Nonparametric Comparison

- Two-Sample Log Rank Test
- Weighted Log Rank Test

Comparison of Failure Time Distributions

Informal comparisons, for two groups of subjects observed.

- Plot $\hat{S}_{KM}(t)$ for each group, put on the same graph; add confidence intervals if needed.
- Plot and compare $\hat{H}_{NA}(t)$, or other nonparametric estimates, of the groups.

4.1 Parametric Comparison

Aim: Test $H_0 : S_1(t) = S_2(t)$, equality of distribution between the 2 groups.

Assumption: S_1, S_2 come from the same parametric family.

Advantage of parametric comparison:

Being able to quantify differences between distributions.

- Log location-scale models.

Assume the failure time T_{ij} has a log location-scale distribution, and $Y_{ij} = \log T_{ij}$, for groups $i = 1, 2$.

Group 1: Log failure times Y_{11}, \dots, Y_{1n_1} .

Group 2: Log failure times Y_{21}, \dots, Y_{2n_2} .

Denote the cumulative distribution functions of the 2 groups by

$$\begin{cases} F_1(y) = G\left(\frac{y-\mu_1}{\sigma_1}\right), \\ F_2(y) = G\left(\frac{y-\mu_2}{\sigma_2}\right). \end{cases}$$

Steps for testing $H_0 : S_1(t) = S_2(t)$.

1. Test $H_0 : \sigma_1 = \sigma_2$. 
2. If $\sigma_1 = \sigma_2$ is not rejected, test $H_0 : \mu_1 = \mu_2$. 



Example 4.1.1. Leukemia Remission Times

Leukemia data with two treatment groups (control and drug 6-MP).

Continuation of Examples 1.2.2., 2.5.1.

Assume the failure time of each group has a Weibull distribution. Test if the two groups have the same distribution.

Test equality of distributions based on Weibull failure time models.

- Check on the equality of scale parameters of the 2 groups, $H_0 : \phi_1 = \phi_2$.

Likelihood ratio (LR) statistic $\Lambda \approx \chi_1^2$.

Observed value is $\lambda_{obs} = 2[-64.9 + (-41.7) - (-106.6)] = 0$.

p-value=1. No evidence against $H_0 : \phi_1 = \phi_2$.

- Test for equality of location parameters of the 2 groups, $H_0 : \mu_1 = \mu_2$.

LR statistic $\Lambda \approx \chi_1^2$.

Observed value is $\lambda_{obs} = 19.65 = 2[(-106.6 - (-116.4))]$.

p-value = 9.3×10^{-6} .

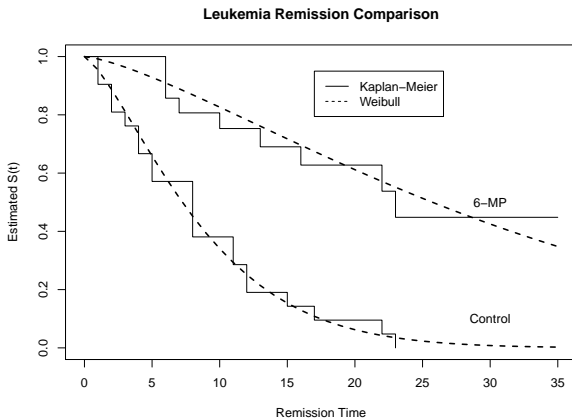
Or test based on **Wald** statistic.

p-value = 4.5×10^{-5} .

Very strong evidence against $H_0 : \mu_1 = \mu_2$.

Model checking: Are Weibull models good enough for the data?

Plot the estimated survivor function from Weibull fit for each group alone with the corresponding Kaplan-Meier estimate.



Summary and Conclusions:

- Model assumptions

Survival times for the control and drug 6-MP groups follow Weibull distributions. That is, log survival times of the 2 groups have EV (μ_1, σ_1) and EV (μ_2, σ_2) distributions respectively.

- Model comparison

We test the equality of survival distributions for control vs. drug 6-MP groups in two steps.

1. Test for the equality of scale parameters ($H_0 : \sigma_1 = \sigma_2$).

A LR test results in a p-value of 1, indicating no difference on the scale parameters between the two models for the two groups.

2. Test for equality of location parameters ($H_0 : \mu_1 = \mu_2$).

The LR test results in a p-value of 9.3×10^{-6} . There is very strong evidence against $H_0 : \mu_1 = \mu_2$. (The Wald test gives the same conclusion.)

- The tests suggest the control and 6-MP groups have different survival (remission) time distributions;
for describing the survival times of the 2 groups, we should consider a Weibull model with $\phi_1 = \phi_2$ ($\sigma_1 = \sigma_2$), but $\mu_1 \neq \mu_2$.
- Model checking
The survivor functions estimated from the above Weibull model are plotted along with the Kaplan-Meier estimates for the 2 groups. The Weibull fit agrees reasonably well with the nonparametric fit. The Weibull model is appropriate for the data.
- Conclusions:
 - ▶ The remission time distributions are significantly different between the patients receiving control treatment and patients receiving drug 6-MP treatment.
 - ▶ A Weibull model with $\phi_1 = \phi_2$, but $\mu_1 \neq \mu_2$ is suitable for describing the survival distributions of the two groups.

```

> library(survival)
> leukdata<-read.table("eg122.txt", header=T)
> # Weibull fit for the data under null hypothesis.
> fitall<-survreg(Surv(time,status)~as.factor(grp),
data=leukdata)
> print(summary(fitall))
Call:
survreg(formula = Surv(time, status) ~ as.factor(grp),
data = leukdata)

              Value Std. Error      z      p
(Intercept)    2.248      0.166 13.55 8.30e-42
as.factor(grp)1  1.267      0.311  4.08 4.51e-05
Log(scale)     -0.312      0.147 -2.12 3.43e-02
Scale= 0.732

Weibull distribution
Loglik(model)= -106.6   Loglik(intercept only)= -116.4
      Chisq= 19.65 on 1 degrees of freedom, p= 9.3e-06
n= 42

```

```
> # Weibull fit for control group.
> fit1<-survreg(Surv(time,status)~1,data=leukdata,
subset=grp==0)
> print(summary(fit1))
Call:
survreg(formula = Surv(time, status) ~ 1,
data = leukdata, subset = grp == 0)

              Value Std. Error      z      p
(Intercept)  2.249      0.168 13.40 5.72e-41
Log(scale)   -0.315      0.174 -1.82 6.94e-02
Scale= 0.73
Weibull distribution
Loglik(model)= -64.9   Loglik(intercept only)= -64.9
n= 21
```

```

> # Weibull fit for drug 6-MP group.
> fit2<-survreg(Surv(time,status)~1,data=leukdata,
subset=grp==1)
> print(summary(fit2))
Call:
survreg(formula = Surv(time, status) ~ 1,
data = leukdata, subset = grp == 1)

              Value Std. Error      z      p
(Intercept)  3.519      0.273 12.87 6.28e-38
Log(scale)   -0.303      0.278 -1.09 2.77e-01
Scale= 0.739

Weibull distribution
Loglik(model)= -41.7   Loglik(intercept only)= -41.7
n= 21

```

```

> mu1<-fitall$coeff[1]
> mu1
(Intercept)
  2.248352
> mu2<-fitall$coeff[2]+mu1
> mu2
as.factor(grp)1
  3.515687
> sig<-fitall$scale
> sig
[1] 0.7321944
>
> # Model checking: plot estimated survival functions
> # by KM method, and Weibull model.
>
> fit.km<-survfit(Surv(time,status)~as.factor(grp),
data=leukdata,conf.int=F)

```

```

> # KM plot.
> plot(fit.km,xlab="Remission Time",ylab="Estimated
S(t)",main="Leukemia Remission Comparison")
>
> # Plot estimated survivor functions by Weibull fit.
> t<-0:35
> St1<-exp(-exp((log(t)-mu1)/sig))
> St2<-exp(-exp((log(t)-mu2)/sig))
> text(30,0.1,"Control")
> text(30,0.5,"6-MP")
> lines(t,St1,lty=2,lwd=2)
> lines(t,St2,lty=2,lwd=2)
> legend(17,0.95,c("Kaplan-Meier","Weibull"),lty=1:2)

```

4.2 Nonparametric Comparison

4.2.1 Two-Sample Log Rank Test

Compare the failure time distributions of 2 groups by testing

$H_0 : S_1(t) = S_2(t)$ for all t , versus

$H_a : S_1(t) \neq S_2(t)$ for some t .

Intuition about log rank statistic:

Let $t_1 < t_2 < \dots < t_K$ be the distinct failure times for the **pooled sample**.

At time t_j	# failures	# did not fail	# at risk
Group 1	d_{1j}		n_{1j}
Group 2			
	d_j		n_j

Let D_{ij} be the number of failures in group i at time t_j .

Under H_0 that the 2 groups have the same distribution, D_{ij} can be modeled by a **hypergeometric distribution**, with

$$P(D_{ij} = d_{ij}) = \frac{\binom{d_j}{d_{1j}} \binom{n_j - d_j}{n_{1j} - d_{1j}}}{\binom{n_j}{n_{1j}}}.$$

The **log rank statistic** takes the form

$$U = \sum_{j=1}^K \left(\frac{d_{1j}}{n_j} - \frac{e_{1j}}{n_j} \right), \text{ with } e_{1j} = E(D_{1j}) = n_{1j} \cdot \frac{d_j}{n_j}.$$

Its variance is estimated by

$$V = \sum_{j=1}^K v_{1j}, \text{ with } v_{1j} =$$

The **log rank test** is based on the result that, under H_0 ,

$$W^2 = \frac{U^2}{V} \approx \chi_1^2, \text{ or } W = \frac{U}{\sqrt{V}} \approx N(0, 1).$$

Example:

4.2.2 Weighted Log Rank Test

- Counting process notation for describing lifetime
 - ▶ More general.
 - ▶ More convenient for theoretical development of methods.
 - ▶ More modern.

Describe the survival time of 1 subject:

$N(t)$ = number of failures up to time t . **Right-continuous** by convention.

$Y(t)$ = number at risk at time t . **Left-continuous** by convention.

For subjects $l = 1, \dots, n$ under study, at time t , describe their lifetime by **failure processes** $N_1(t), \dots, N_n(t)$, and **at risk processes** $Y_1(t), \dots, Y_n(t)$.

Further define

$$N_{\cdot}(t) = \sum_{l=1}^n N_l(t)$$

$$Y_{\cdot}(t) = \sum_{l=1}^n Y_l(t)$$

$$dN_{\cdot}(t) = \sum_{l=1}^n dN_l(t)$$

Recall that in general, if T is continuous,

$$H(t) =$$

$$\text{and } \hat{H}_{NA}(t) =$$

Using the counting process notation

$$\hat{H}_{NA}(t) = \int_0^t d\hat{H}(u) = \int_0^t \frac{dN_{\cdot}(t)}{Y_{\cdot}(t)}.$$

- Idea for weighted log rank test

Group 1: subjects $l = 1, \dots, n_1$.

Counting processes: $N_{11}(t), \dots, N_{1n_1}(t)$; and $Y_{11}(t), \dots, Y_{1n_1}(t)$.

Group 2: subjects $l = 1, \dots, n_2$.

Counting processes: $N_{21}(t), \dots, N_{2n_2}(t)$; and $Y_{21}(t), \dots, Y_{2n_2}(t)$.

Test for $H_0 : S_1(t) = S_2(t)$ through a weighted comparison of \hat{S}_1 and \hat{S}_2 .

The **weighted** log rank test consider the statistic

$$U = \int_0^{\infty} w(t) \frac{Y_{1\cdot}(t)Y_{2\cdot}(t)}{Y_{..\cdot}(t)} [d\hat{H}_1(t) - d\hat{H}_2(t)] , \quad (1)$$

where $w(t)$ is the pre-specified weight,

$$Y_{1\cdot}(t) = \sum_{l=1}^{n_1} Y_{1l}(t)$$

$$Y_{2\cdot}(t) = \sum_{l=1}^{n_2} Y_{2l}(t)$$

$$Y_{..\cdot}(t) = Y_{1\cdot}(t) + Y_{2\cdot}(t) =$$

$\hat{H}_i(t)$ is the NA estimate for group i , $i = 1, 2$;

for group i , $d\hat{H}_i(t) = \frac{dN_{i\cdot}(t)}{Y_{i\cdot}(t)}$,

$$N_{i\cdot}(t) = \sum_{l=1}^{n_i} N_{il}(t)$$

$$dN_{i\cdot}(t)$$

Denote $dN_{i.}(t)$ at time t_j by $di.(t_j)$, $i = 1, 2$.

At time t_j	# failures	# did not fail	# at risk
Group 1	$d_{1.}(t_j)$		$Y_{1.}(t_j)$
Group 2			

The **weighted log rank statistic** has the form

$$U = \sum_{j=1}^K w(t_j) \left[d_{1.}(t_j) - \frac{Y_{1.}(t_j)}{Y_{..}(t_j)} d_{..}(t_j) \right] \quad (2)$$

Some results for U , the weighted log rank statistic (2).

- $E(U) = 0$.
- The variance of U can be estimated by

$$V = \sum_{j=1}^K [w(t_j)]^2 Y_{1\cdot}(t_j) \frac{d_{\cdot\cdot}(t_j)}{Y_{\cdot\cdot}(t_j)} \left[1 - \frac{d_{\cdot\cdot}(t_j)}{Y_{\cdot\cdot}(t_j)} \right] \frac{Y_{2\cdot}(t_j)}{Y_{\cdot\cdot}(t_j) - 1}.$$

The **weighted log rank test** is based on the result that, under H_0 ,

$$W^2 = \frac{U^2}{V} \approx \chi_1^2, \text{ or } W = \frac{U}{\sqrt{V}} \approx N(0, 1).$$

Remarks:

- The log rank test and the Wilcoxon test are the most popular among the weighted log rank tests.
- Taking $w(t_j) = 1$ in (2) gives the **log rank statistic**;
taking $w(t_j) = \hat{S}(t)$ gives the **Wilcoxon statistic** for testing H_0 .
- Advantage of nonparametric methods: No distributional assumptions.

Example 4.2.1 Leukemia Remission Time

Leukemia data with two treatment groups (control and drug 6-MP).

Continuation of Examples 1.2.2., 2.5.1, 4.1.1.

Test equality of distributions through nonparametric methods.

Methods:

Log rank test and Wilcoxon test for $H_0 : S_1(t) = S_2(t)$,
(group 1 = those given control treatment, group 2 = those given drug 6-MP).

Results and conclusions:

- Log rank test and Wilcoxon test both give very small p-values (4×10^{-5} and 1×10^{-4}).
- There are very strong evidences against the null hypothesis.
- The distributions of remission time are **significantly different** between the drug 6-MP group and the control group.

```

> library(survival)
> leuk<-read.table("eg122.txt", header=T)
> # Log rank test.
> logrk<-survdifff(Surv(time,status)~as.factor(grp),
data=leuk)
> logrk
Call:
survdifff(formula = Surv(time, status) ~ as.factor(grp),
data = leuk)

```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
as.factor(grp)=0	21	21	10.7	9.77	16.8
as.factor(grp)=1	21	9	19.3	5.46	16.8

```

  Chisq= 16.8  on 1 degrees of freedom, p= 4e-05
>

```

```

> # Wilcoxon test.
> wilcx<-survdif(Surv(time,status)~as.factor(grp),
data=leuk,rho=1)
> wilcx
Call:
survdif(formula = Surv(time, status) ~ as.factor(grp),
data = leuk, rho = 1)

```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
as.factor(grp)=0	21	14.55	7.68	6.16	14.5
as.factor(grp)=1	21	5.12	12.00	3.94	14.5

Chisq= 14.5 on 1 degrees of freedom, p= 1e-04

[Notes](#) on survdiff() in R