

Survival Analysis

Lecture 7

Marta Fiocco^{1,2} & Hein Putter¹

(1) Department of Medical Statistics and Bioinformatics
Leiden University Medical Center

(2) Mathematical Institute Leiden University

Outline

One-sample tests

Tests for two or more samples

Theory

Illustration

Tests for trend

Stratified tests

Local tests

Outline

One-sample tests

Tests for two or more samples

Theory

Illustration

Tests for trend

Stratified tests

Local tests

One-sample problem

Question

- ▶ Is observed data compatible with pre-specified distribution?
- ▶ For instance, do patients with a particular type of disease have a higher risk of dying than (age/sex-matched) population?

Hypothesis

$$H_0 : \quad h(t) = h_0(t) , \text{ for all } t \leq \tau$$

$$H_1 : \quad h(t) \neq h_0(t) , \text{ for some } t \leq \tau$$

Basic idea

Use the hazards

- ▶ We have an estimator of the (cumulative) hazard (Nelson-Aalen)

$$d\hat{H}(t_i) = \frac{d_i}{Y(t_i)}$$

- ▶ Under H_0 , we would expect that $d\hat{H}(t_i) \approx h_0(t_i)$
- ▶ Compare these over all time points t_i

Test statistic

Use the hazards

- ▶ Test statistic takes the form

$$\begin{aligned} Z(\tau) &= \int_0^{\tau} W(s)(d\hat{H}(s) - dH_0(s)) \\ &= O(\tau) - E(\tau) \end{aligned}$$

- ▶ Under H_0 :

$$\text{var}(Z(\tau)) = \int_0^{\tau} W^2(s) \frac{h_0(s)}{Y(s)} ds$$

- ▶ Again under H_0 , asymptotically $Z(\tau)/\sqrt{\text{var}(Z(\tau))}$ has a standard normal distribution

Special cases

Choice of weight function

- ▶ Choice $W(t) = Y(t)$ with τ is end-of-study (largest time point) gives **one-sample log-rank test**
- ▶ In that case, it is not difficult to see that

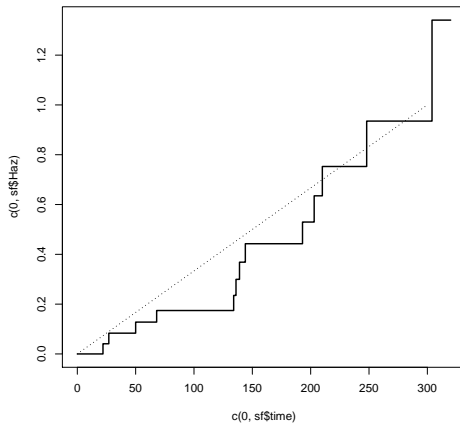
$$O(\tau) = \# \text{deaths} , \quad E(\tau) = \text{var}(Z(\tau)) = \sum_{i=1}^n H_0(t_i)$$

Example

Lung data

```
> sf <- survfit(Surv(time2, death2) ~ 1, data = lung)
> sf <- data.frame(time = sf$time, surv = sf$surv, atrisk = sf$n.risk)
> sf$Haz <- -log(sf$surv)
> plot(c(0, sf$time), c(0, sf$Haz), type = "s", lwd = 2)
> lines(c(0, 300), c(0, 1), type = "l", lwd = 1, lty = 3)
```


Plot



Example

Lung data; compatible with $\exp(1/300)$?

```
> lung$H0ti <- lung$time2/300
> lung$S0ti <- exp(-lung$H0ti)
> Obs <- sum(lung$death2)
> Exp <- sum(lung$H0ti)
> Z <- Obs - Exp
> varZ <- Exp
> z <- Z/sqrt(varZ)
> p <- 2*(1-pnorm(abs(z)))
> data.frame(Observed=Obs, Expected=Exp, z=z, p=p)
```

	Observed	Expected	z	p
1	13	14.53667	-0.4030392	0.6869194

Software

- ▶ We can also use the function *survdiff* from the *survival* package
- ▶ Then we need to define for each patient the "expected" survival probability
- ▶ Already done that (*S0ti* in *lung*)
- ▶ Use that as offset

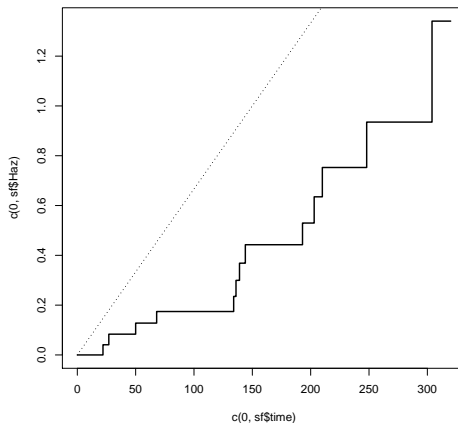
```
> survdiff(Surv(time2,death2) ~ offset(S0ti), data=lung)
```

Call:

```
survdiff(formula = Surv(time2, death2) ~ offset(S0ti), data = lung)
```

Observed	Expected	Z	p
13.000	14.537	0.403	0.687

Compatible with $\exp(1/150)$?



Example

Lung data; compatible with $\exp(1/150)$?

```
> lung$H0ti <- lung$time2/150
> Obs <- sum(lung$death2)
> Exp <- sum(lung$H0ti)
> Z <- Obs - Exp
> varZ <- Exp
> z <- Z/sqrt(varZ)
> p <- 2*(1-pnorm(abs(z)))
> data.frame(Observed=Obs, Expected=Exp, z=z, p=p)
```

	Observed	Expected	z	p
1	13	29.07333	-2.980976	0.002873309

```
> lung$S0ti <- exp(-lung$H0ti)
> survdiff(Surv(time2, death2) ~ offset(S0ti), data=lung)
```

Call:

```
survdiff(formula = Surv(time2, death2) ~ offset(S0ti), data = lung)
```

	Observed	Expected	Z	p
	13.00000	29.07333	2.98098	0.00287

Outline

One-sample tests

Tests for two or more samples

Theory

Illustration

Tests for trend

Stratified tests

Local tests

Two-sample tests

- ▶ Compare hazard rates of $K \geq 2$ populations
- ▶ Hypotheses:

$$H_0 : h_1(t) = h_2(t) = \dots = h_K(t) , \text{ for all } t \leq \tau$$

$$H_1 : \text{at least one of the } h_j(t)'s \text{ is different for some } t \leq \tau$$

Data

- ▶ $t_1 < t_2 < \dots < t_D$: distinct death times in **pooled** sample
- ▶ At time t_i :
 - ▶ In sample j : d_{ij} events out of Y_{ij} at risk
 - ▶ In pooled sample: $d_i = \sum_{j=1}^K d_{ij}$ events out of $Y_i = \sum_{j=1}^K Y_{ij}$ at risk

A class of test functions

Basic idea

- ▶ Expected hazard rate in j^{th} sample under H_0 is estimated as Nelson-Aalen estimator of the pooled sample
- ▶ Compare Nelson-Aalen estimator of j^{th} sample with overall Nelson-Aalen
- ▶ Observed versus expected
- ▶ For positive weight function $W_j(t)$:

$$Z_j(\tau) = \sum_{i=1}^D W_j(t_i) \left\{ \frac{d_{ij}}{Y_{ij}} - \frac{d_i}{Y_i} \right\}$$

- ▶ Technical detail: need $W_j(t_i) = 0$ whenever Y_{ij} is zero

A special class of test functions

Common weight function

- ▶ In practice all the commonly used tests have

$$W_j(t_i) = Y_{ij} W(t_i)$$

- ▶ Then

$$Z_j(\tau) = \sum_{i=1}^D W(t_i) \left\{ d_{ij} - Y_{ij} \frac{d_i}{Y_i} \right\}$$

- ▶ Test statistic is sum of weighted difference between observed number of deaths and expected under H_0

Distribution

Distribution of $Z_j(\tau)$

- ▶ Asymptotically joint normal
- ▶ Variance of $Z_j(\tau)$ is estimated by

$$\sum_{i=1}^D W(t_i)^2 \frac{Y_{ij}}{Y_i} \left(1 - \frac{Y_{ij}}{Y_i}\right) \left(\frac{Y_i - d_i}{Y_i - 1}\right) d_i$$

- ▶ Covariance of $Z_j(\tau), Z_g(\tau)$ is estimated by

$$- \sum_{i=1}^D W(t_i)^2 \frac{Y_{ij}}{Y_i} \frac{Y_{ig}}{Y_i} \left(\frac{Y_i - d_i}{Y_i - 1}\right) d_i$$

- ▶ Term $\frac{Y_i - d_i}{Y_i - 1}$ is a correction for ties

Test statistic

- ▶ $Z_1(\tau), \dots, Z_K(\tau)$ are linearly dependent, because $\sum_{j=1}^K Z_j(\tau) = 0$
- ▶ Terms $\frac{Y_{ij}}{Y_i}(1 - \frac{Y_{ij}}{Y_i})d_i$ and $-\frac{Y_{ij}}{Y_i}\frac{Y_{ig}}{Y_i}d_i$ come from multinomial distribution with parameters d_i and $p_j = \frac{Y_{ij}}{Y_i}$
- ▶ Overall test statistic is constructed by selecting any $K - 1$ of the Z_j 's
- ▶ Put corresponding variances and covariances in Σ
- ▶ Test statistic is given by

$$\chi^2 = (Z_1(\tau), \dots, Z_K(\tau))\Sigma^{-1}(Z_1(\tau), \dots, Z_K(\tau))^T$$

- ▶ Under H_0 it has a chi-squared distribution, asymptotically, with $K - 1$ degrees of freedom

Two samples

Special (common) case $K = 2$

- ▶ Two-sample test

$$Z = \frac{\sum_{i=1}^D W(t_i) \left\{ d_{i1} - Y_{i1} \frac{d_i}{Y_i} \right\}}{\sqrt{\sum_{i=1}^D W(t_i)^2 \frac{Y_{i1}}{Y_i} \left(1 - \frac{Y_{i1}}{Y_i} \right) \frac{Y_i - d_i}{Y_i - 1} d_i}}$$

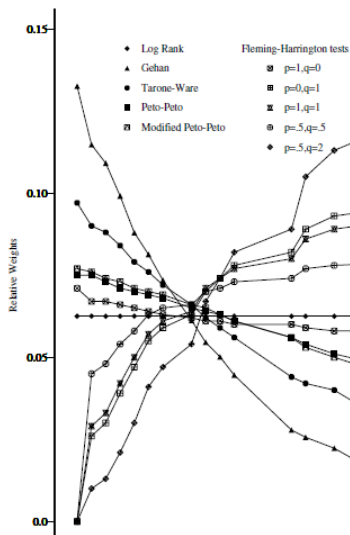
- ▶ Under H_0 it has a standard normal distribution for large samples

Choice of weight functions

- ▶ **Log-rank test** (easiest and most common): $W(t) \equiv 1$
- ▶ **Tarone-Ware** class: $W(t_i) = f(Y_i)$ for some fixed function f
 - ▶ $f(y) = y$ quite common
 - ▶ $f(y) = \sqrt{y}$ suggested by Tarone & Ware (1977)
- ▶ Peto-Peto (1972) and Kalbfleisch & Prentice (1980) suggest $W(t_i) = \tilde{S}(t_i)$ (close to Kaplan-Meier estimate)
- ▶ **Fleming-Harrington** (1981) propose a very general class of tests that includes log-rank

$$W_{p,q}(t_i) = \hat{S}(t_{i-1})^p (1 - \hat{S}(t_{i-1}))^q, \quad p \geq 0, \quad q \geq 0$$

Weight functions illustrated



Example

- ▶ Data set `kidney` from the *KMSurv* package
- ▶ Compares two methods for placing catheters in kidney dialysis patients (column `type`)
- ▶ Interested in time to cutaneous exit-site infection (`time`)

```
> table(kidney$type)
```

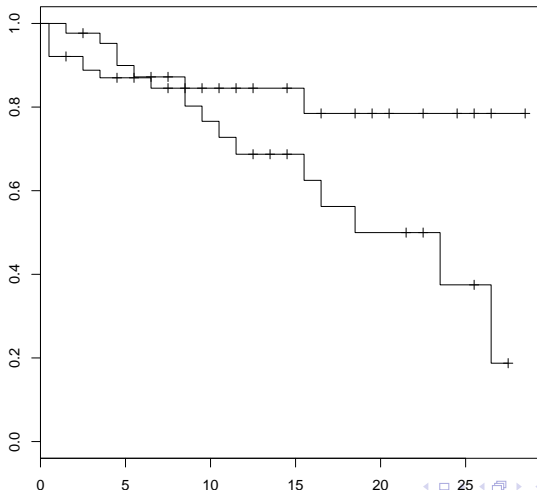
```
 1  2
43 76
```

```
> table(kidney$delta, kidney$type)
```

```
      1  2
0 28 65
1 15 11
```

Survival curves

```
> plot(survfit(Surv(time,delta) ~ type, data=kidney))
```



Log-rank test

► Implemented in the function *survfit*

```
> survdiff(Surv(time,delta) ~ type, data=kidney)
```

Call:

```
survdiff(formula = Surv(time, delta) ~ type, data = kidney)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
type=1	43	15	11.0	1.42	2.53
type=2	76	11	15.0	1.05	2.53

Chisq= 2.5 on 1 degrees of freedom, p= 0.112

Other tests

- *survfit* implements the Fleming-Harrington family with $q = 0$

```
> survdiff(Surv(time,delta) ~ type, data=kidney, rho=1)
```

Call:

```
survdiff(formula = Surv(time, delta) ~ type, data = kidney, rho = 1)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
type=1	43	12.0	9.48	0.686	1.39
type=2	76	10.4	12.98	0.501	1.39

Chisq= 1.4 on 1 degrees of freedom, p= 0.239

```
> survdiff(Surv(time,delta) ~ type, data=kidney, rho=0.5)
```

Call:

```
survdiff(formula = Surv(time, delta) ~ type, data = kidney, rho = 0.5)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
type=1	43	13.4	10.2	1.002	1.91
type=2	76	10.7	13.9	0.735	1.91

Chisq= 1.9 on 1 degrees of freedom, p= 0.167

Outline

One-sample tests

Tests for two or more samples

Theory

Illustration

Tests for trend

Stratified tests

Local tests

Tests for trend

- ▶ We want a test that has power against ordered alternatives, i.e. against

$$H_1 : h_1(t) \leq h_2(t) \leq \dots \leq h_K(t) ,$$

for $t \leq \tau$, with at least one strict inequality

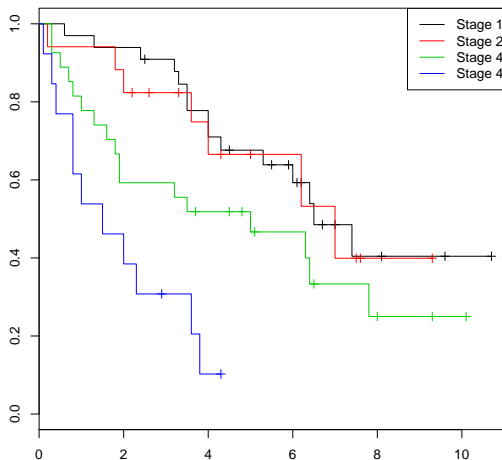
- ▶ Equivalent to $H_1 : S_1 \geq S_2(t) \geq \dots \geq S_K(t)$
- ▶ Ingredients are the $Z_j(\tau)$ from before
- ▶ Let $\hat{\Sigma}$ be the $K \times K$ covariance matrix of $Z_j(\tau)$
- ▶ Test: for scores $a_1 < a_2 < \dots < a_K$,

$$Z = \frac{\sum_{j=1}^K a_j Z_j(\tau)}{\sqrt{\sum_{j=1}^K \sum_{k=1}^K a_j a_k \hat{\Sigma}_{jk}}}$$

- ▶ Under H_0 of no difference, Z has asymptotic normal distribution

Survival curves

```
> plot(survfit(Surv(time,delta) ~ stage, data=larynx), col=1:4)
> legend("topright",c("Stage 1", "Stage 2", "Stage 4", "Stage 4"),lwd=1,c
```



Log-rank test

► The log-rank test gives

```
> survdiff(Surv(time,delta) ~ stage, data=larynx)
```

Call:

```
survdiff(formula = Surv(time, delta) ~ stage, data = larynx)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
stage=1	33	15	22.57	2.537	4.741
stage=2	17	7	10.01	0.906	1.152
stage=3	27	17	14.08	0.603	0.856
stage=4	13	11	3.34	17.590	19.827

Chisq= 22.8 on 3 degrees of freedom, p= 4.53e-05

Test for trend

- From the book (p. 217)

Using the log-rank weights,

$$\mathbf{Z}(10.7) = (-7.5660, -3.0117, 2.9155, 7.6623) \text{ and}$$

$$\hat{\Sigma} = \begin{pmatrix} 12.0740 & -4.4516 & -6.2465 & -1.3759 \\ -4.4516 & 7.8730 & -2.7599 & -0.6614 \\ -6.2465 & -2.7599 & 9.9302 & -0.9238 \\ -1.3759 & -0.6614 & -0.9238 & 2.9612 \end{pmatrix}.$$

The value of the test statistic (7.4.2) is 3.72 and the p -value of the test is less than 0.0001.

- As far as I know, *survdifff* does not allow for trend test
- This afternoon we will implement it ourselves!

Outline

One-sample tests

Tests for two or more samples

Theory

Illustration

Tests for trend

Stratified tests

Local tests

Stratified tests

- ▶ Suppose we want to adjust for other covariates that may affect the event rates in the K populations
- ▶ Then we can **stratify** on these covariates
- ▶ It means that we test the null hypothesis

$$H_0 : h_{1s}(t) = h_{2s}(t) = \dots = h_{Ks}(t) ,$$

for $s = 1, \dots, M, t \leq \tau$

- ▶ So the null hypothesis is that the K hazard rates are the same in each of M strata
- ▶ The strata are defined by the values of the covariates
- ▶ Should not be too many; if you stratify for gender and age (e.g. < 40 , $40-60$, > 60), then you have six strata
- ▶ Stratifying for continuous covariates is usually not feasible

Stratified test

Construction of the test

- ▶ The building blocks are the $Z_{js}(\tau)$, calculated for each group j within each stratum s
- ▶ Likewise we have covariance matrices Σ_s for each stratum
- ▶ Define

$$Z_{j\bullet}(\tau) = \sum_{s=1}^M Z_{js}(\tau) , \quad \Sigma_{\bullet} = \sum_{s=1}^M \Sigma_s$$

- ▶ Then the test statistic is defined as in the two-sample test, but then based on $Z_{j\bullet}(\tau)$ and Σ_{\bullet}

Example

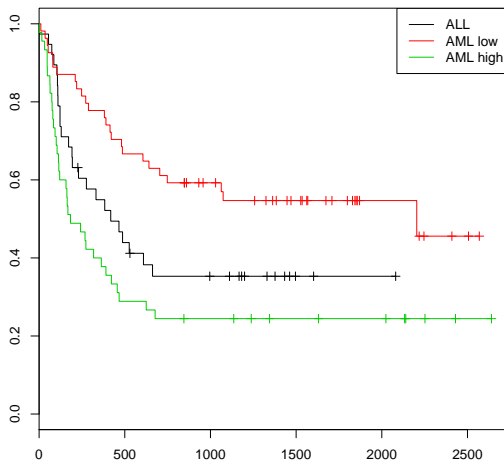
- ▶ BMT data (`bmt` from *KMsurv*)
- ▶ Interested in relapse-free survival (`t2`, time in days, `d3`, status)
- ▶ Differences between three groups of patients (`group`, 1=ALL, 2=AML low-risk, 3=AML, high-risk)
- ▶ But we want to adjust for the use of a graft-versus-host prophylactic MTX (`z10`, 0=no MTX, 1=MTX)

```
> data(bmt)
> table(bmt$group, bmt$z10)
```

	0	1
1	21	17
2	42	12
3	34	11

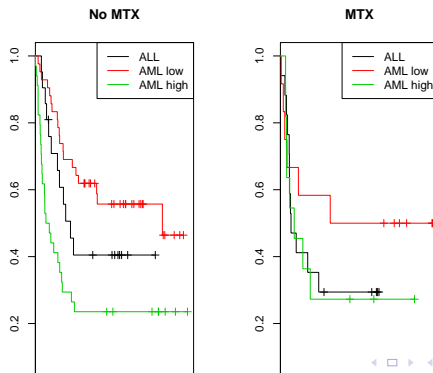
Survival curves overall

```
> plot(survfit(Surv(t2,d3) ~ group, data=bmt), col=1:3)
> legend("topright",c("ALL", "AML low", "AML high"),lwd=1,col=1:3)
```



Survival curves per MTX

```
> bmt0 <- bmt[bmt$z10==0,]  
> bmt1 <- bmt[bmt$z10==1,]  
> par(mfrow=c(1,2))  
> plot(survfit(Surv(t2,d3) ~ group, data=bmt0), col=1:3, main="No MTX")  
> legend("topright",c("ALL", "AML low", "AML high"),lwd=1,col=1:3)  
> plot(survfit(Surv(t2,d3) ~ group, data=bmt1), col=1:3, main="MTX")  
> legend("topright",c("ALL", "AML low", "AML high"),lwd=1,col=1:3)
```



Stratified log-rank test

```
> survdiff(Surv(t2,d3) ~ group + strata(z10), data=bmt)
```

Call:

```
survdiff(formula = Surv(t2, d3) ~ group + strata(z10), data = bmt)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
group=1	38	24	23.2	0.0261	0.038
group=2	54	25	38.7	4.8663	9.621
group=3	45	34	21.0	7.9673	10.796

Chisq= 13.2 on 2 degrees of freedom, p= 0.00136

SKIP

- ▶ Renyi type tests (7.6) and other two-sample tests (7.7) hardly ever used
- ▶ SKIP

Outline

One-sample tests

Tests for two or more samples

Theory

Illustration

Tests for trend

Stratified tests

Local tests

Local test

At t_0

- ▶ For treatment A :
 - ▶ Kaplan-Meier estimate $\hat{S}_A(t_0)$
 - ▶ Its standard error $se(\hat{S}_A(t_0))$
- ▶ For treatment B :
 - ▶ Kaplan-Meier estimate $\hat{S}_B(t_0)$
 - ▶ Its standard error $se(\hat{S}_B(t_0))$
- ▶ Basic idea:
 - ▶ Look at difference $\hat{S}_A(t_0) - \hat{S}_B(t_0)$
 - ▶ If this is large reject H_0

Local test

- ▶ If two groups are independent, then

$$se(\hat{S}_A(t_0) - \hat{S}_B(t_0)) = \sqrt{se^2(\hat{S}_A(t_0)) + se^2(\hat{S}_B(t_0))}$$

- ▶ Test statistic

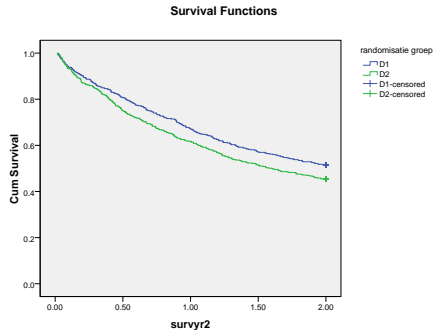
$$Z = \frac{\hat{S}_A(t_0) - \hat{S}_B(t_0)}{\sqrt{se^2(\hat{S}_A(t_0)) + se^2(\hat{S}_B(t_0))}}$$

- ▶ P-value from standard-normal distribution
- ▶ Confidence interval for $S_A(t_0) - S_B(t_0)$ can be constructed

$$\hat{S}_A(t_0) - \hat{S}_B(t_0) \pm 1.96 \cdot \sqrt{se^2(\hat{S}_A(t_0)) + se^2(\hat{S}_B(t_0))}$$

Example

D1D2 study



Example

- ▶ Let us look at two years
- ▶ D1-dissection: $\hat{S}_A(t_0) = 0.516$, with standard error of 0.022
- ▶ D2-dissection: $\hat{S}_B(t_0) = 0.453$, with standard error of 0.021
- ▶ $\hat{S}_A(t_0) - \hat{S}_B(t_0) = 0.516 - 0.453 = 0.063$
- ▶ Its standard error is $\sqrt{0.022^2 + 0.021^2} = 0.0304$
- ▶ Leading to $z = 0.063/0.0304 = 2.072$, with a P-value of 0.038
- ▶ 95% confidence interval of the difference is given by $0.063 \pm 1.96 \cdot 0.0304$ which runs from 0.003 to 0.123

More than two groups

- ▶ These tests can be extended to more than two groups
- ▶ Simultaneous tests of $S_1(t_0) - S_2(t_0)$ and $S_1(t_0) - S_3(t_0)$ for instance
- ▶ Using contrasts
- ▶ Read Section 7.8 of Klein & Moeschberger