Applied Survival Analysis Using R
Chapter 4: Nonparametric Comparison of Survival Distributions

Qi Guo

Department of Mathematical Sciences
The University of Texas at Dallas

April, 11 2019

Test Hypothesis commonly used

- Typically we are interested in testing a null hypothesis(H_0) that two population means are equal versus an alternative(H_a) that the means are not equal(for a two-sided test)
- And we compute a test statistic from the observed data, and *reject* the null hypothesis if the test statistic *exceeds* a particular constant.
- The significance level of the test(α) is the probability that we *reject* the null hypothesis when the null hypothesis is in fact *true*.
- A widely known test is the two-sample Students t-test for continuous observations, which requires the assumption that the observations are normally distributed.
- If the normal distribution assumption is in *doubt*, a rank-based test called the Mann-Whitney test may be used, which gives valid test results without making parametric assumptions

Nonparametric tests of equivalence of two survival functions

- Null Hypothesis:
  $H_0 : S_1(t) = S_0(t)$. S_1 and S_0 will represent the survival distributions for, respectively, an experimental and a control therapy.
- Alternative Hypothesis:
  - One-sided: $H_A : S_1(t) > S_0(t)$.
  - Two-sided: $H_A : S_1(t) \neq S_0(t)$.
- ⊗ But it will cause the alternative can take a wide range of forms(the only equal or not is not exact enough)
- ⊗ What if the survival distributions are similar for some values of t and differ for others?
- ⊗ What if the survival distributions cross?

Refine

- ⊙ Lehman alternative:
  - Survival function: $S_1(t) = [S_0(t)]^{\psi}$.
  - Proportional hazard function: $h_1(t) = \psi h_0(t)$.
- Hence the hypothesis test will be:
  - $H_0 : \psi = 1$.
  - $H_A : \psi < 1$(subjects in Group 1 will have longer survival times than subjects in Group 0).
- ⊙ Use rank-based test, and view the numbers of failure and numbers at risk at each distinct time as a two-by-two table.
- ⊙ Stratified Tests

Table

For the *i*'th failure time table showing:

	Control	Treatment	
Failure	d_{0i}	d_{1i}	d_i
Non-failures	$n_{0i} - d_{0i}$	$n_{1i} - d_{1i}$	$n_i - d_i$
	n_{0i}	n_{1i}	n_i

- The numbers *at risk* (n_{0i} and n_{1i} for the control and treatment arms)
- The number of *failures* (d_{0i} and d_{1i})

- Suppose that the numbers of failures in the control and treatment groups are independent, and all RVs except d_{0i} are fixed, then the distribution of d_{0i} follows what is known as a hypergeometric distribution.

Mean and Variance

- The mean and variance are given by:
  - Mean: $e_{0i} = E(d_{0i}) = \frac{n_{0i} d_i}{n_i}$.
  - Variance: $v_{0i} = var(d_{0i}) = \frac{n_{0i} n_{1i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)}$
- Sum up the differences between the observed and expected values to get a linear test statistic U_0, and the sum of the variances V_0, where D is the number of failure times:
  - $U_0 = \sum_{i=1}^{D} (d_{0i} - e_{0i}) = \sum d_{0i} - \sum e_{0i}$.
  - $var(U_0) = \sum v_{0i} = V_0$.
- Then construct a test statistic that is standard normal:
  - $\frac{U_0}{\sqrt{V_0}} \sim N(0, 1)$ or
  - $\frac{{U_0}^2}{V_0} \sim \chi_1^2$

Example

Table 4.1 Survival data

Patient	Survtime	Censor	Group
1	6	1	C
2	7	0	C
3	10	1	T
4	15	1	C
5	19	0	T
6	25	1	T

t_i	n_i	d_i	n_{0i}	d_{0i}	n_{1i}	d_{1i}	e_{0i}	e_{1i}	$v_{0i} = v_{1i}$
6	6	1	3	1	3	0	0.500	0.500	0.2500
10	4	1	1	0	3	1	0.250	0.750	0.1875
15	3	1	1	1	2	0	0.333	0.667	0.2222
25	1	1	0	0	1	1	0.000	1.000	0.0000
				2		2	1.083	2.917	0.6597

```
1 > tt <- c(6, 7, 10, 15, 19, 25)
2 > delta <- c(1, 0, 1, 1, 0, 1)
3 > trt <- c(0, 0, 1, 0, 1, 1)
4 > survdiff(Surv(tt, delta) ~ trt)
5 [1]          N   Observed   Expected   (O-E)^2/E   (O-E)^2/V
6 trt=0        3          2       1.08       0.776        1.27
7 trt=1        3          2       2.92       0.288        1.27
8
9 Chisq= 1.3 on 1 degrees of freedom, p= 0.259
```

- Conclusion: The p-value is 0.259, indicating that the group difference is *not statistically significant*.

Weighted log-rank test

Cochran-Mantel-Haenzel Test

A test for *independence* of two factors (here, treatment and outcome) adjusted for a potential confounder(like log-rank test).

- An important generalization of this test makes use of a series of D weights w_i, with which we may define a weighted log-rank test by:
  - Mean: $U_0(w) = \sum w_i(d_{0i} - e_{0i})$
  - Variance: $var(U_0) = \sum w_i^2 v_{0i} = V_0(w)$
- The most common way of setting weights is:
  - $w_i = \{\hat{S}(t_i)\}^\rho$
  - $\rho = 0$: log-rank test
  - $\rho = 1$: prentice modification
- A log-rank test using these weights is called the Fleming-Harrington $G(\rho)$ test.

Example

- The effect of this test is then to place *higher weight* on earlier survival differences.

```
> head(pancreatic)
[1]   stage       onstudy      progression       death
1       MPC     12/16/2005       2/2/2006    10/19/2006
2       MPC      1/6/2006       2/26/2006     4/19/2006
3       LAPC     2/3/2006        8/2/2006     1/19/2007
4       MPC      3/30/2006            <NA>     5/11/2006
5       LAPC     4/27/2006      3/11/2007      5/29/2007
6       MPC      5/7/2006       6/25/2006    10/11/2006
```

- Died with no recorded, shown "NA", so **PFS** is time to death, otherwise **PFS** is time to the date of progression.

```
> attach(pancreatic)
> # convert the text dates into R dates
> Progression.d<-as.date(as.character(progression))
> OnStudy.d<-as.date(as.character(onstudy))
> Death.d<-as.date(as.character(death))
```

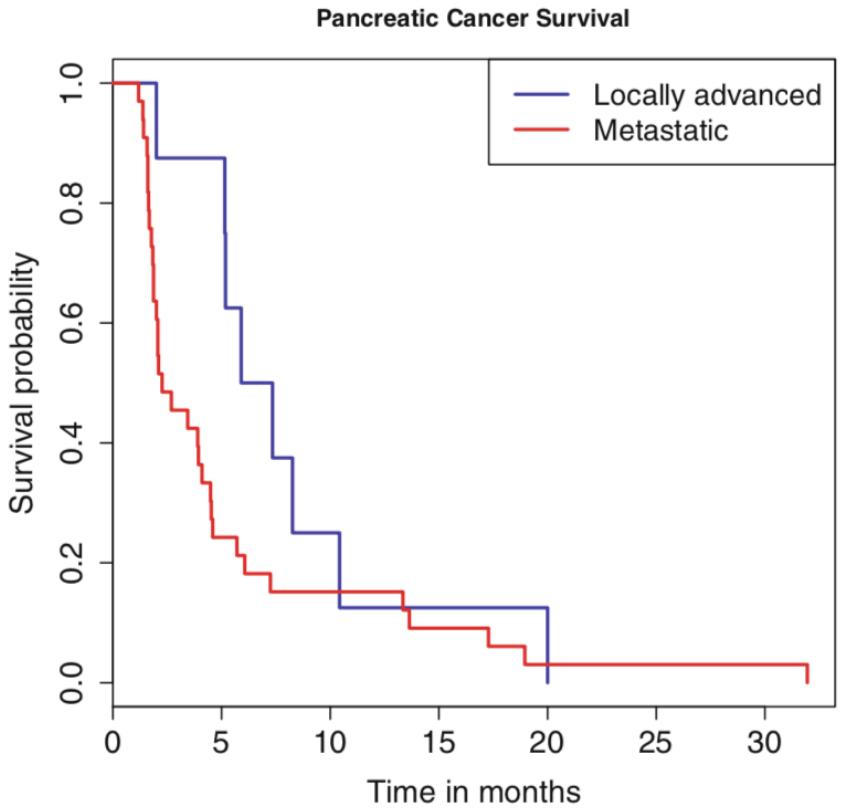
Example by log-rank test

```
> #compute progression free survival
> progressionOnly<-Progression.d - OnStudy.d
> overallSurvival<-Death.d-OnStudy.d
> pfs<-progressionOnly
> pfs[is.na(pfs)]<-overallSurvival[is.na(pfs)]
> #convert pfs to months
> pfs.month<-pfs/30.5
> #note that no observations are censored
> #plot
> plot(survfit(Surv(pfs.month) ~ stage), xlab="Time in months",
ylab="Survival probability", col=c("blue", "red"), lwd=2)
> legend("topright", legend=c("Locally advanced", "Metastatic"),
col=c("blue","red") , lwd=2)
```

- The log-rank test is when $\rho = 0$

```
> survdiff(Surv(pfs) ~ stage, rho=0)
[1]              N   Observed   Expected   (O-E)^2/E   (O-E)^2/V
stage=LA         8          8       12.3        1.49        2.25
stage=M         33         33       28.7        0.64        2.25

Chisq= 2.2 on 1 degrees of freedom, p= 0.134
```

Example by log-rank test



- Conclusion: The p-value is 0.134, indicating that the **LA** and **M** difference is *not statistically significant*.

Example by Prentice modification

```
> survdiff(Surv(pfs) ~ stage, rho=1)
[1]             N   Observed      Expected   (O-E)^2/E   (O-E)^2/V
stage=LA        8       2.34          5.88       2.128        4.71
stage=M        33      18.76         15.22       0.822        4.71

Chisq= 4.7 on 1 degrees of freedom, p= 0.0299
```

- Conclusion: The p-value is 0.0299, indicating that the **LA** and **M** difference is *statistically significant*.
- Explain:This version of the test places *higher weight on earlier* survival times. From Fig last slide we see that indeed the **M** group shows an *early survival advantage over* the locally advanced group, but the survival curves converge after about 10 months. The reason for the difference is that these two tests, with ρ 0 or 1, are optimized for different alternatives.

Hypothesis Test

- Compare two groups while adjusting for another *covariate*, and if the covariate we are adjusting for is categorical with a small number of levels G, we may construct a *stratified log-rank test*.
- Null hypothesis: $H_0 : h_{0j}(t) = h_{1j}(t)$ for $j = 1, 2, ..., G$.
- Each level of the second variable, compute a *score statistic* U_{0g} and variance V_{0g}, where $g = 1, 2, ..., G$ is the group indicator.
- The test statistic is: $X^2 = \frac{(\sum_{g=1}^{G} U_{0g})^2}{\sum_{g=1}^{G} {V_{0g}}^2}$
- Compared to a χ_1^2.

Example by unstratified groups

- **`Treatment center`**, **`age group`**, or **`gender`** are examples of variables need to stratify.
- The primary goal is to compare the *time to relapse* (defined in this study as return to smoking) between two treatment groups:

```
> attach (pharmacoSmoking)
> survdiff(Surv(ttr, relapse) ~ grp)
[1]                    N   Observed    Expected   (O-E)^2/E   (O-E)^2/V
grp=combination   61      37          49.9         3.36        8.03
grp=patchOnly     64      52          39.1         4.29        8.03

Chisq=8 on 1 degrees of freedom, p= 0.00461
```

- If we are concerned that the group comparison may differ by age, we may define a categorical variable

```
> table(ageGroup2))
[1] ageGroup2
     21-49   50+
     66      59
```

Example by stratified groups

- The log-rank test stratified on "`ageGroup2`":

```
> survdiff(Surv(ttr, relapse) ~ grp + strata(ageGroup2))
[1]                  N   Observed    Expected   (O-E)^2/E   (O-E)^2/V
grp=combination     61         37        49.1        2.99        7.03
grp=patchOnly       64         52        39.9        3.68        7.03

Chisq=7 on 1 degrees of freedom, p= 0.008
```

- Conclusion: The chi-square test in this case *differs only slightly* from the unadjusted value, indicating that it was not necessary to stratify on this variable.

Simulated Dataset

Clinical trial comparison

Set up a simulated dataset from a clinical trial comparing a standard therapy (control = 0) to an experimental therapy (treated = 1). Suppose that the survival times are exponentially distributed, and that the disease is rapidly fatal, so that there is no censoring. We also suppose that there is a confounding variable, "`genotype`", which can either be wild type (i.e. normal) or mutant, the data will shown below:

```
> lambda.mutant.0 <- 0.03
> lambda.mutant.1 <- 0.03*0.55
> lambda.wt.0 <- 0.03*0.2
> lambda.wt.1 <- 0.03*0.2*0.55
```

Simulated Dataset

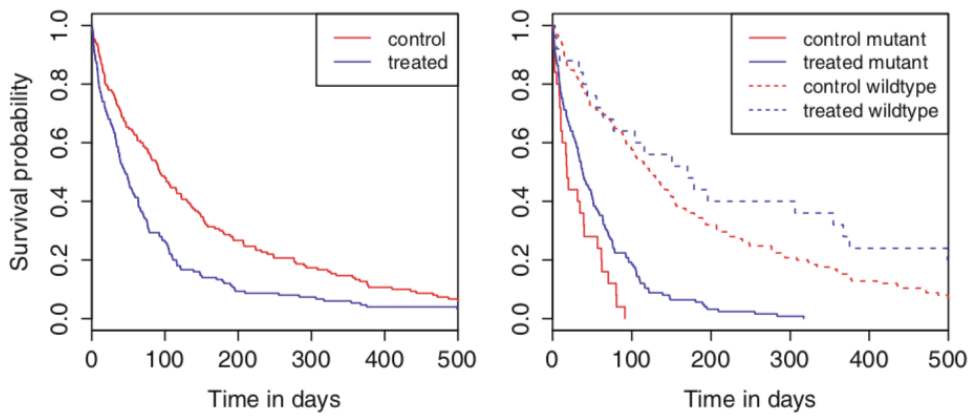
- (1) set a "`seed`" for the random variable generator, so that this example may be reproduced exactly.
- (2) generate exponential random variables and string them together into the variable "`ttAll`".
- (3) create the censoring variable "`status`".
- (4) create the treatment variable "`trt`" and `genotype`, as follows:

```
> set.seed(4321)

> tt.control.mutant <- rexp(25, rate=lambda.mutant.0)
> tt.treat.mutant <- rexp(125, rate=lambda.mutant.1)
> tt.control.wt <- rexp(125, rate=lambda.wt.0)
> tt.treat.wt <- rexp(25, rate=lambda.wt.1)
> ttAll <- c(tt.control.mutant, tt.treat.mutant,
 tt.control.wt, tt.treat.wt)
> status <- rep(1, length(ttAll))
> genotype <- c(rep("mutant", 150), rep("wt", 150))
> trt <- c(rep(0, 25), rep(1, 125), rep(0, 125), rep(1, 25))
```

Simulated Dataset

- The survival plots comparing the two treatments is shown below(left).
- The log-rank test appears to confirm this with a very *strong* p-value.

```
> survdiff(Surv(ttAll, status) ~ trt))
[1]          N   Observed    Expected    (O-E)^2/E    (O-E)^2/V
trt=0      150      150         183          6.00         15.9
trt=1      150      150         117          9.41         15.9

Chisq=15.9 on 1 degrees of freedom, p= 6.66e-05
```



Simulated Dataset

- But plot the survival curves comparing treatment to control separately for the mutant and wild type patients(right of the Figure last slide), we see that *within each genotype* the treatment is actually superior to the control.

```
> survdiff(Surv(ttAll, status) ~ trt + strata(genotype))
[1]          N   Observed    Expected    (O-E)^2/E   (O-E)^2/V
trt=0      150      150          133         2.17        7.57
trt=1      150      150          167         1.73        7.57

Chisq=7.6 on 1 degrees of freedom, p= 0.00595
```

- Explain:
  - The treatment improves survival compared to the control.
  - Patients carrying the wild type form of the gene have better survival than do patients carrying the mutation.
  - There are more mutation-carrying patients in the treatment group than in the control group, whereas the reverse is true for wild type patients.