

## Chapter 5. Testing

Two samples tests:

When there are two groups, group 1 and group 2, one of the interested question of survival analysis is that whether the survival experience of the 2 groups are equal to each other. Let  $S_1(t)$  and  $S_2(t)$  be the survival functions of the 2 groups respectively. Thus, one wants to know whether  $H_0 : S_1(t) = S_2(t)$  for all  $t$  or  $H_a : S_1(t) \neq S_2(t)$  for some  $t$ .

Suppose at the time origin, there are  $n_1$  individuals belong to group 1 and  $n_2$  individuals belong to group 2. After a period of time, we have two set of survival data. Based on these two data set, we can have the following table where  $d_j$  is the number of death within the monitoring period among group  $j$ ,  $j = 1, 2$ . In addition,  $d = d_1 + d_2$  and  $n = n_1 + n_2$ .

	# of death	# of survive	
Group 1	$d_1$	$n_1 - d_1$	$n_1$
Group 2	$d_2$	$n_2 - d_2$	$n_2$
	$d$	$n - d$	$n$

Assume group 1 and group 2 are independent. If no censoring, we then can consider the Fisher's (chi-square) test when  $d$  is large. Specifically, assume the marginal  $n, n_1, d$  are fixed, then  $d_1$  follows hypergeometric distribution. We have

$$P(D_1 = d_1) = \binom{n_1}{d_1} \binom{n_2}{d - d_1} / \binom{n}{d}.$$

Thus,  $E(D_1) = dn_1/n$  and  $Var(D_1) = [d(n - d)n_1n_2]/[(n - 1)n^2]$ . This implies

$$[d_1 - \frac{dn_1}{n}] / \sqrt{\frac{d(n - d)n_1n_2}{(n - 1)n^2}} \rightarrow N(0, 1)$$

if  $n \rightarrow \infty$  and  $d \rightarrow \infty$ . Furthermore, if  $d$  is small, we can use hypergeometric distribution to get critical point directly. This is what we call Fisher's exact test.

What else we can do?

- (1) Compare mean — t tests
- (2) Compare median — sign tests
- (3)  $S_1(t^*)$  vs  $S_2(t^*)$ ,  $t^* = 5$  years for example.

(Note:  $\frac{\hat{S}_1(5) - \hat{S}_2(5)}{s.e} \sim N(0, 1)$  where  $s.e^2 = S_1^2(5) \sum \frac{d_{1i}}{n_{1i}(n_{1i} - d_{1i})} + S_2^2(5) \sum \frac{d_{2i}}{n_{2i}(n_{2i} - d_{2i})}$ .)

(4)  $P(T_1 > T_2) = 0.5$  — Wilcoxon test (rank sum test)  $\equiv$  Mann-Whitney test.

(Note:  $MW = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{ij}$  where

$$w_{ij} = \begin{cases} 1, & \text{if } T_{1i} > T_{2i} \\ 0, & \text{if } T_{1i} = T_{2i} \\ -1, & \text{if } T_{1i} < T_{2i} \end{cases} .$$

)

Intuitively, in knowing whether two survival curves are equal to each other, (1) to (3) have less power. Moreover, if there are censoring, the first example and (4) can not be used directly. Later, we will discuss Gehan's Generalized Wilcoxon test which is an extension of (4), and also log rank test which can be look as an extension of the first example.

Suppose the failure time and censoring time of group 1 and group 2 are

Group 1:  $T_{11}, \dots, T_{1,n1} \sim S_1; C_{11}, \dots, C_{1,n1} \sim G_1$   
and

Group 2:  $T_{21}, \dots, T_{2,n2} \sim S_2; C_{21}, \dots, C_{2,n2} \sim G_2,$

respectively. However, we can only observed  $Y_{ji} = \min(T_{ji}, C_{ji})$  and  $\delta_{ji} = I(T_{ji} = Y_{ji}), j = 1, 2.$

Gehan's Generalized Wilcoxon test:

Redefinne

$$w_{ij}^* = \begin{cases} 1, & \text{if } \delta_{2i} = 1 \text{ and } Y_{1i} > Y_{2i} \\ 0, & \text{o.w} \\ -1, & \text{if } \delta_{1i} = 1 \text{ and } Y_{1i} < Y_{2i} \end{cases} .$$

Similarly as Wilcoxon test, the asymptotic distribution of  $W_g = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{ij}^*$  is normal distribution.

Log rank test:

If there is only one death time, say  $t_{(1)}$ , then testing  $H_0 : S_1 = S_2 \Leftrightarrow h_1 = h_2$ .  
However

	$t_{(1)}$	sample size
Group 1	$d_1$	$n_1$
Group 2	$d_2$	$n_2$

$\hat{h}_i = d_i/n_i$  ( $i = 1, 2$ ). Here,  $n_i$  is the number at risk for group  $i$  at time  $t_{(1)}$ .  
Therefore, the test statistic becomes

$$(\hat{h}_1 - \hat{h}_2) / \sqrt{(\frac{d_1}{n_1} \frac{n_1 - d_1}{n_1}) / n_1 + (\frac{d_2}{n_2} \frac{n_2 - d_2}{n_2}) / n_2}$$

which converges to standard normal distribution. On the other hand, if  $d = d_1 + d_2$  is small, we then can consider Fisher's test statistic

$$(d_1 - \frac{dn_1}{n}) / \sqrt{\frac{d(n-d)n_1n_2}{n^2}}.$$

How about if there are more than one death time?

Let  $t_{(1)} < t_{(2)} < \dots < t_{(r)}$  be  $r$  ordered distinct failure time across two groups. At each death time  $t_{(j)}$  construct a  $2 \times 2$  table

Group	# of death at $t_{(j)}$	# of surviving beyond $t_{(j)}$	# at risk just before $t_{(j)}$
1	$d_{1j}$	$n_{1j} - d_{1j}$	$n_{1j}$
2	$d_{2j}$	$n_{2j} - d_{2j}$	$n_{2j}$
total	$d_j$	$n_j - d_j$	$n_j$

where  $d_{ij}$  = # of death at  $t_{(j)}$  for group  $i$  and  $n_{ij}$  = # at risk at  $t_{(j)}$  for group  $i$ . Given the marginal, that is given  $d_j, n_j, n_{1j}, n_{2j}$ ,  $D_{1j}$  follows hypergeometry distribution. Specifically,

$$P(D_{1j} = d) = \binom{n_{1j}}{d} \binom{n_{2j}}{d_j - d} / \binom{n_j}{d_j},$$

$$E(D_{1j}) = d_j n_{1j} / n_j \text{ and } Var(D_{1j}) = [d_j(n_j - d_j)n_{1j}n_{2j}] / [n_j^2(n_j - 1)].$$

Define  $O_j = d_{1j}$ ,  $E_j = E(D_{1j})$  and  $V_j = Var(D_{1j})$ , we have

$$(O_j - E_j) / \sqrt{V_j} \rightarrow N(0, 1) \text{ when } d_j \rightarrow \infty.$$

Consider

$$T_w = \sum w_j (O_j - E_j) / \sqrt{V_w}$$

where  $V_w = \sum w_j^2 Var(O_j - E_j) = \sum w_j^2 V_j \approx Var(\sum w_j (O_j - E_j))$ . Under  $H_0$ ,  $T_w$  converges to standard normal or equivalently  $T_w^2$  converges to  $X_1^2$ .

There are many choice for  $w_j$ .

- (1)  $w_j = 1 \Rightarrow$  Ordinary log rank test
- (2)  $w_j = n_j \Rightarrow$  Gehan's test
- (3)  $w_j = n\hat{S}(t_{(j)}) \Rightarrow$  Prentice's test. Here,  $\hat{S}(t)$  is K-M estimate of  $S(t)$  using combined sample
- (4)  $w_j = n[\hat{S}(t_j)]^k \Rightarrow$  Haorington and Fleming test with  $0 \leq k \leq 1$
- (5)  $w_j = \sqrt{n_j} \Rightarrow$  Tarone and Ware test

Note that

- (i)  $h_1(t)/h_2(t) = \text{constant}$  implies (1);
- (ii) (1) and (3) are most useful;
- (iii) (2) depends on censoring distribution.

For an example. Suppose at the beginning, each group have 1000 units. Because of no money, each group throw away 950 units after some failure times occur. If we consider Gehan's test, the result will dominate by the earlier failure time.

Left truncation:

Data:  $(Y_i, L_i, X_i)$ ,  $i = 1, \dots, n$ ,  $L_i < Y_i$ ,  $X_i = 1$  or  $2$ .

Let  $t_{(1)} < \dots < t_{(r)}$  be ordered distinct death time  $Y_i$ . For each death time  $t_{(j)}$  construct a  $2 \times 2$  table

Group	# of death at $t_{(j)}$	# of surviving beyond $t_{(j)}$	# at risk just before $t_{(j)}$
1	$d_{1j}$	$n_{1j} - d_{1j}$	$n_{1j}$
2	$d_{2j}$	$n_{2j} - d_{2j}$	$n_{2j}$
total	$d_j$	$n_j - d_j$	$n_j$

where  $d_{ij} = \#$  of death at  $t_{(j)}$  for group  $i$  and  $n_{ij} = \sum_{k=1}^n I(L_k \leq t_j \leq Y_k, X_k = i)$ . Everything hold the same as when only right censoring. A slightly different for left truncation, we need to redefine the number at risk at  $t_{(j)}$  for group  $i$ .

k-samples test:

$H_0 : S_1(t) = \dots = S_k(t)$  vs  $H_a$  : at least 1 different.

Let  $t_{(1)} < \dots < t_{(r)}$  be ordered distinct death time from all samples. Construct a  $k \times 2$  table:

Group	# of death at $t_{(j)}$	# of surviving beyond $t_{(j)}$	# at risk at $t_{(j)}$
1	$d_{1j}$	$n_{1j} - d_{1j}$	$n_{1j}$
2	$d_{2j}$	$n_{2j} - d_{2j}$	$n_{2j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$k$	$d_{kj}$	$n_{kj} - d_{kj}$	$n_{kj}$
total	$d_j$	$n_j - d_j$	$n_j$

Define two  $(k-1) \times 1$  vectors,  $O_j = (d_{2j}, \dots, d_{kj})$ ,  $E_j = (E_{2j}, \dots, E_{kj})$  and a  $(k-1) \times (k-1)$  matrix  $V_j = [\text{cov}(D_{ij}, D_{i'j})]$  where  $E_{ij} = d_j n_{ij} / n_j$ ,  $\text{Var}(D_{ij}) = [d_j(n_j -$

$d_j)n_{ij}(n_j - n_{ij})]/[n_j^2(n_j - 1)]$  and  $cov(D_{ij}, D_{i'j}) = [-d_j(n_j - d_j)n_{ij}n_{i'j}]/[n_j^2(n_j - 1)]$ .  
Let  $O = \sum O_j, E = \sum E_j$  and  $V = \sum V_j$ , then

$$T = [O - E]^T V^{-1} [O - E] \sim X_{k-1}^2.$$

$T$  is called  $k$  samples log rank statistic. Intuitively, we may think to consider  $\sum w_j O_j$  where  $w_j$  is a  $(k-1) \times (k-1)$  weight matrix. However, it is too complicate. Most of the time we just use  $T$ .

Test for trend:

$$H_0 : S_1 = S_2 = S_3 \text{ vs } H_a : S_1 < S_2 < S_3$$

Let  $c = (c_2, c_3, \dots, c_k)^T$  has a  $(k-1)$  vectors of constants such that  $c_2 < c_3 < \dots < c_k$ .

Since  $(O - E) \sim MN(0, V)$ ,  $c^T(O - E)/\sqrt{c^T V c} \sim N(0, 1)$ .

How to define  $c$ ?

Decide  $c$  before looking at the data, just based on other knowledge.

- (i)  $(c_2, c_3, \dots, c_k) = (1, 2, \dots, k-1)$
- (ii)  $(c_2, c_3, \dots, c_k) = \text{dose level}, (\text{dose level})^2, \ln(\text{dose}), \text{ etc.}$

Stratify test:

Treatment vs Control

$$H_0 : S_1 = S_2 \text{ vs } H_a : S_1 \neq S_2$$

Example:

Female group:  $\sum(O_{fj} - E_{fj}) = T_f$  and  $V_f = Var(T_f)$

Male group:  $\sum(O_{mj} - E_{mj}) = T_m$  and  $V_m = Var(T_m)$

Consider  $T = [T_f + T_m]/\sqrt{V_f + V_m} \sim N(0, 1)$  or equivalently  $T^2 \sim X_1^2$ . We can also consider  $T^{*2} = T_f^2/V_f + T_m^2/V_m \sim X_2^2$ . In general, lower df is better.

One sample:

$$H_0 : S(t) = S_0(t) \text{ vs } H_a : S(t) \neq S_0(t).$$

Define  $d_j = \#$  of death at  $t_{(j)}$ ,  $N(t) = \#$  at risk at time  $t$ . Let  $T_j = d_j - \int_0^\infty h_0(t)N(t)dt$  and  $V = Var(d_j - \int_0^\infty h_0(t)N(t)dt) = \int_0^\infty h_0(t)N(t)dt$ . Consider

$$T = \sum T_j / \sqrt{\sum V_j} \sim N(0, 1).$$

Other way:

- (i)  $\sup |\hat{S}(t) - S_0(t)|$
- (ii)  $\int [\hat{S}(t) - S_0(t)]^2 w(t) dt$
- (iii)  $Y_i^* = H_0(Y_i) = \min\{H_0(T_i), H_0(C_i)\}$ ,  $\delta_i^* = I(H_0(T_i) \leq H_0(C_i)) = \delta_i$ , thne  $(Y_i^*, \delta_i^*)$  are right censored data from  $Exp(1)$  under null hypothesis. One way to do the test is compare  $[\sum Y_i^* - d]/\sqrt{d}$  with  $N(0, 1)$  where  $d = \sum \delta_i^*$ . This is true since the MLE of  $\lambda$ ,  $\hat{\lambda} = d/\sum Y_i^*$ , this implies  $(\frac{\sum Y_i^*}{d} - 1) \sim N(0, d)$ .