5.3 Comparison of two groups of survival data:

Suppose $t_{(1)} < t_{(2)} < \cdots < t_{(m)}$ are m distinct death times across two groups and that at time $t_{(j)}$, d_{1j} individuals in Group I die and d_{2j} individuals in Group II die, for $j = 1, \cdots, m$. Also, suppose n_{1j} individuals in Group I are still alive just before time $t_{(j)}$ while n_{2j} individuals in Group II are still alive just before time $t_{(j)}$. Let $n_j = n_{1j} + n_{2j}$ and $d_j = d_{1j} + d_{2j}$. Then, we have the following table,

Group	Number of death at $t_{(j)}$	Number surviving beyond $t_{(j)}$	Number alive just before $t_{(j)}$
I	d_{1j}	$n_{1j}-d_{1j}$	n_{1j}
II	d_{2j}	$n_{2j}-d_{2j}$	n_{2j}
Total	d_j	n_j-d_j	$n_j = n_{1j} + n_{2j}$

Then, we can regard $d_{1j} \sim B(n_{1j}, \pi_{1j})$ and $d_{2j} \sim B(n_{2j}, \pi_{2j})$ with

$$logit(\pi_{1j}) = \lambda_j + \Delta,$$

$$logit(\pi_{2j}) = \lambda_j.$$

Thus, conditionally on d_j, d_{1j} has a hypergeometric distribution. Based on the conditional distribution, the score statistic

$$U_L = \left[\frac{dl_c(\Delta)}{d\Delta}\right]_{\Delta=0} = \sum_{j=1}^{m} (d_{1j} - e_{1j}) = \sum_{j=1}^{m} \left(d_{1j} - \frac{n_{1j} d_j}{n_j}\right),$$

where $l_c(\Delta)$ is the conditional log-likelihood function and

$$e_{1j} = \frac{n_{1j} d_j}{n_j}$$

is the conditional mean of d_{1j}. Further,

$$V_L = Var(U_L) = \sum_{j=1}^{m} \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}.$$

As the number of death time is not too small, then

$$Z = \frac{U_L}{\sqrt{V_L}} \approx N(0, 1).$$

Note:

We can also use the continuity-correct value

$$Z^- = \frac{U_L - {}^1/_2}{\sqrt{V_L}}$$

We can use Z or Z^- to test $H_0: \Delta = 0$. The test using

$$Z^2 = \frac{U_L^2}{V_L}.$$

is referred to as Log-rank test or Mantel-Haenzel procedure. Note that

$$Z^2 \sim \chi_1^2$$

under $H_0: \Delta = 0$.

The other test is based on

$$U_W = \sum_{j=1}^{m} n_j (d_{1j} - e_{1j}) = \sum_{j=1}^{m} n_j \left(d_{1j} - \frac{n_{1j} d_j}{n_j} \right)$$

which is referred to as the Wilcoxon test. The difference between U_L and U_W is that the Wilcoxon test, each difference $d_{1j} - e_{1j}$ is weighted by n_j. The effect of this is to give less weight to difference between d_{1j} and e_{1j} at those times when the total number of individuals who are still alive is small, that is, at the longest survival time. The variance of U_W is

$$V_W = Var(U_W) = \sum_{j=1}^{m} n_j^2 \left[\frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)} \right].$$

Thus, the Wilcoxon statistic to test $H_0: \Delta = 0$ is

$$\frac{U_W^2}{V_W}$$

and

$$\frac{U_W^2}{V_W} \sim \chi_1^2$$

under $H_0: \Delta = 0$.

Example (continue):

In the motivating example, we have data

Group	T	C	T	C	T	T	C	C
Survival time	6	6^*	7	7	7^*	9.5^*	10^*	11

Then, we have the following tables:

$t_{(1)} = 6$:

Group	# of death	# alive beyond	# alive
T	$d_{11} = 1$	3	$n_{11} = 4$
C	$d_{21} = 0$	4	$n_{21} = 4$
	$d_1 = 1$	7	$n_1 = 8$

$t_{(2)} = 7$:

Group	# of death	# alive beyond	# alive
T	$d_{12} = 1$	2	$n_{12} = 3$
C	$d_{22} = 1$	2	$n_{22} = 3$
	$d_2 = 2$	4	$n_2 = 6$

$t_{(3)} = 11$:

Group	# of death	# alive beyond	# alive
T	$d_{13} = 0$	0	$n_{13} = 0$
C	$d_{23} = 1$	0	$n_{23} = 1$
	$d_3 = 1$	0	$n_3 = 1$

Thus,

$$\begin{aligned} U_L &= (d_{11} - e_{11}) + (d_{12} - e_{12}) + (d_{13} - e_{13}) \\ &= \left(1 - \frac{4 \cdot 1}{8}\right) + \left(1 - \frac{3 \cdot 2}{6}\right) + \left(0 - \frac{0 \cdot 1}{1}\right) \\ &= 0.5 \end{aligned}$$

and

$$\begin{aligned} U_W &= n_1(d_{11} - e_{11}) + n_2(d_{12} - e_{12}) + n_3(d_{13} - e_{13}) \\ &= 8\left(1 - \frac{4 \cdot 1}{8}\right) + 6\left(1 - \frac{3 \cdot 2}{6}\right) + 1\left(0 - \frac{0 \cdot 1}{1}\right) \\ &= 4. \end{aligned}$$

Similarly, V_L and V_W can be obtained.