

Analysis of life time data: Solutions to Exercises 1

1. There are two comparisons of the group of patients who do not receive a kidney transplant with the group of patients who do receive a kidney transplant which one naturally considers:
 - (a) Compare the time from starting dialysis to death for both groups. The disadvantage of this is that, in order to receive a kidney transplant, patient have to live until a suitable donor kidney becomes available. Therefore, the people in the transplant group will tend to have longer lifetimes than those in the no-transplant group.
 - (b) Compare the time from starting dialysis to death for the no-transplant group with the time from transplant to death for the other group. The disadvantage of this is that the length of time the transplant group survive from starting dialysis to transplant is ignored, so their survival will tend to be under-estimated.

There is no simple way to compare the survival. Models which take into account the age, severity of renal failure and general health are needed, so that one can attempt to adjust for the over and under-estimation problems in the two methods above. There are risks associated with the operation, which are part of what one seeks to assess.

2. It is not likely that most people died at about 37 years. The mean will be substantially affected by a large number of infants with lifetimes of only a few months, as infant mortality was very high in the middle ages. Those who survived to age five, say, would have a life expectancy of much nearer 70 years. The Jewish scriptures assume that the natural life span is seventy years, some centuries BC.

The risk of death would be high in early childhood. For men, there would be risks associated with warfare, and for women there would be risks associated with childbirth. In both cases, these risks would be highest in the age range, very roughly, 15 to 40 years. For boys in the UK around the second millennium AD, there is an increased risk of death at ages 16-20.

South African Life Tables for 2006 gave life expectancy at birth for women and men from high income families as 77.36 and 70.68 years respectively. For the lowest income families (<R42 000 per year), life expectancy at birth was 62.53 (female) and 54.76 (male) years. From age 1 year, the life expectancy was to age 66.11 (female) and 58.31 (male) years.

3. Mean - for roughly symmetrical data; Median - for skew data. Mode - for data dominated by one value, e.g. if a set of lifetimes has many zeroes, one would give the proportion of zeroes, and a separate summary of non-zero times.
4. Standard deviation - with mean, for roughly symmetrical data; Quartiles and interquartile range - for skew data. Frequencies - for categorical data, including binary data.

*Notation for censoring: use * or + or - after the lifetime. These are all common.*

5. Type I censoring for at 10 days. The mouse who escaped was randomly censored (discuss whether this might be informative). 3, 4, 5, 5*, 7, 7, 8, 9, 10*, 10*.

6. Type II censoring. Again, the mouse who escaped was randomly censored. 3, 4, 5, 5*, 6*, 6*, 6*, 6*, 6*, 6*.
7. This is intended to be Type I censoring: if all students were contacted and replied, then the data would give the time from graduation to employment, censored at a fixed data. There are various ways in which this will not happen. Some students will not be looking for work, or be studying a further degree: are these students censored, or not relevant? Some students will be travelling round the world, with the intention of looking for work when they return home. Other students will have job waiting for them on their return home. Some addresses will have changed, and those students might not receive the request. Other students will not bother to answer. The letters which are answered will be answered on different dates, so even replies will not all be censored at exactly the same date. Some-one who has a job interview might wait until after the interview to reply. So, it is not easy to answer the question 'What proportion of students are employed six months after graduating?'
8. People who died before 1968 will be truncated, so the people who are registered are only those who live long enough. Those who were registered and alive in 2003 would be censored (Type I). If a researcher is interested in long term survival, she could define the population as those who lived to 1968. In this case, those who register later than 1968 have left truncated survival times.
9. This is Type II censoring.
10. This is random censoring. Those patients who do not leave the trial enter at a random date, and are followed up until 1 January 2006, say. (Why 1 January, not 5 February? Because we should always allow some delay in reporting.) Those who move away are probably randomly censored. Patients who are censored when they stop taking the drugs might not be censored independently of the outcome of interest.
11. First, decide what time is of interest: time from finding the cancer to death, or time from a tumour of a certain size, such as 1cm, to death. Many people will not attend screening, so the times recorded are for a reduced population, or are truncated. If we consider time from finding the cancer, then we ignore issues of left and interval censoring of the starting point. If we consider a 1cm tumour, then people who were found to have a ≥ 1 cm tumour at first screening have left censored observations, and those who do not have a 1cm tumour at screen number i , but have one at screen number $i+1$ have interval censored observations. We have some information on people who attend for screening, but none on what happens to people who go to their doctor when they have trouble with their stomach, and have a tumour found that way. Such people might have larger tumours, and lower life expectancy.
12. Mouse foot-pad ulcers. See table T.1.
13. Aircraft component. See tabel T.2 The probability that a component lasts for an hour is 0.69. The median survival time is about 1.5 hours, and 23% of components were still working after three hours.

T.1 i	Failure times	No. at risk	No. failed	Number censored	Failure prob'ty	Cond'l survival prob.	Survival estimate
1	3	10	1	0	1/10	9/10	0.9
2	4	9	1	0	1/9	8/9	$8/9 \times 0.9=0.8$
3	5	8	1	0	1/8	7/8	0.7
4	7	6	2	1	2/6	4/6	0.47
5	8	4	1	0	1/4	3/4	0.35
6	9	3	1	0	1/3	2/3	0.23
7	10	2	0	0	01	0	0.23

T.2 i	Failure times	No. at risk	No. failed	Number censored	Conditional probability	Survival estimate
1	0.22	13	1	0	1/13=0.077	1-0.077=0.923
2	0.50	12	1	0	1/12=0.083	$0.923 \times 0.913= 0.846$
3	0.88	11	1	0	1/11	0.769
4	1.00	10	1	0	1/10	0.692
5	1.32	9	1	0	1/9	0.615
6	1.33	8	1	0	1/8	0.538
7	1.54	7	1	0	1/7	0.462
8	1.76	6	1	0	1/6	0.385
9	2.50	5	1	0	1/5	0.308
10	3.00	4	1	0	1/4	0.231

14. Aluminium

time	n.risk	n.event	survival
296	17	1	0.9412
660	16	1	0.8824
670	15	1	0.8235
728	14	1	0.7647
775	13	1	0.7059
797	12	1	0.6471
841	11	1	0.5882
869	10	1	0.5294
999	9	1	0.4706
1006	8	1	0.4118
1035	7	1	0.3529
1169	6	1	0.2941
1193	5	1	0.2353
1415	4	1	0.1765
1424	3	1	0.1176
1500	2	1	0.0588
1540	1	1	0.0000

15.

		Gehan		group= placebo			
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	
1	21	2	0.9048	0.0641	0.67005	0.975	
2	19	2	0.8095	0.0857	0.56891	0.924	
3	17	1	0.7619	0.0929	0.51939	0.893	
4	16	2	0.6667	0.1029	0.42535	0.825	
5	14	2	0.5714	0.1080	0.33798	0.749	
8	12	4	0.3810	0.1060	0.18307	0.578	
11	8	2	0.2857	0.0986	0.11656	0.482	
12	6	2	0.1905	0.0857	0.05948	0.377	
15	4	1	0.1429	0.0764	0.03566	0.321	
17	3	1	0.0952	0.0641	0.01626	0.261	
22	2	1	0.0476	0.0465	0.00332	0.197	
23	1	1	0.0000				

		Gehan		group= 6-MP			
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	
6	21	3	0.857	0.0764	0.620	0.952	
7	17	1	0.807	0.0869	0.563	0.923	
10	15	1	0.753	0.0963	0.503	0.889	
13	12	1	0.690	0.1068	0.432	0.849	
16	11	1	0.627	0.1141	0.368	0.805	
22	7	1	0.538	0.1282	0.268	0.747	
23	6	1	0.448	0.1346	0.188	0.680	

The quartiles and medians are 4, 8 and 12 weeks for placebo and 13, 23 and unknown for the 6-MP group. Half of the people on the drug 6-MP were without leukaemia for 23 weeks, but for the people on placebo, half had leukaemia again after only 8 weeks. (Rough linear interpolation for placebo gives 3.1, 6.1, 11.3; with a small data set, the estimates are very variable.)

16. logrank test R output:

```
survdif(formula = Surv(time, dead) ~ group)

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

Chisq= 16.8 on 1 degrees of freedom, p= 4.17e-05
```

As $p < 0.05$, we can reject the null hypothesis of no difference. The patients who were on the drug 6-MP were without leukaemia for much longer than those on placebo. After 22 weeks, 53% of the 6-MP group were without leukaemia, but only 5% of the placebo group were without leukaemia.