Research Methods 2

Week 14: Exercise Sheet 1

Solution sheet

Question 1

a)
There are several things you might try. If there were no censored times then matters would be easy. You would count how many remission times exceeded 15 and divide this by 21, the total number of patients.

In fact, if there were censored times but all these values were over 15 weeks, then the same calculation would also present no problems, because we know that the patients with censored times have already been in remission more than 15 weeks.

The problem is that there are patients who have not yet relapsed but have not yet been in remission for 15 weeks. E.g., the patient with time 10*: he or she might relapse in the following week, i.e. relapse at week 11 or stay in remission for another year – we just do not know.

You could assume that the patients with censored times relapsed straight after they were last seen, so you essentially ignore the *s. There would then be 11 patients known to be in remission for 15 weeks out of 21 in total, giving a 15 week remission proportion of 11/21 = 0.52. However, this seems pessimistic and does indeed give an answer that is too low.

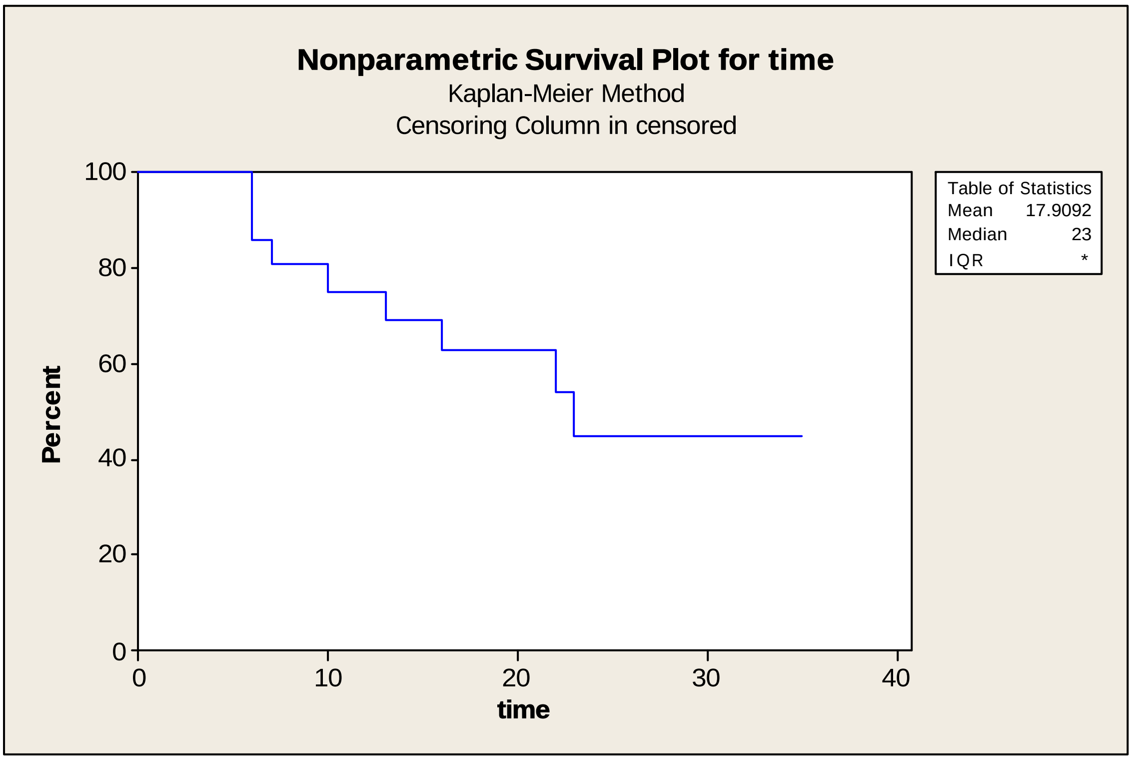
You might just ignore patients still in remission who have not been observed for at least 15 weeks. So all patients with *ed times less than 15, i.e. patients with times 6*, 9*, 10* and 11* are omitted, leaving 11 patients who are known to have stayed in remission for more than 15 weeks out of 17 patients, giving a 15 remission proportion of 11/17 = 0.65. However this is clearly wasting data.

b)

First you need to enter the data into Minitab. The first thing is to enter the times into a column, say column C1. Enter all 21 times, ignoring the *s. Now enter a second column, say C2, which contains just the values 1 and 0. Next to any fully observed remission time, enter a 1 and next to a censored time enter a 0. So, e.g. the first six rows of the Data Window will be

C1	C2
6	0
6	1
6	1
6	1
7	1
9	0

Now Click on **<u>S</u>tat** - > **Re<u>l</u>iability/Survival** - > **Di<u>s</u>tribution Analysis (Right Censoring)** -> **<u>N</u>onparametric Distribution Analysis**. Now follow the instructions given in the study document for the analysis of the lymphoma data. The graph you obtain is shown below and the survival probabilities are also shown.



Kaplan-Meier Estimates

Time	Number at Risk	Number Failed	Survival Probability	Standard Error	95.0% Normal CI Lower	95.0% Normal CI Upper
6	21	3	0.857143	0.076360	0.707479	1.00000
7	17	1	0.806723	0.086935	0.636333	0.97711
10	15	1	0.752941	0.096350	0.564099	0.94178
13	12	1	0.690196	0.106815	0.480843	0.89955
16	11	1	0.627451	0.114054	0.403910	0.85099
22	7	1	0.537815	0.128234	0.286482	0.78915
23	6	1	0.448179	0.134591	0.184385	0.71197

The estimated proportion of patients in remission for at least 15 weeks is from the plot, or from the table, 0.6902.

This differs from the results of part a). While the difference in this instance is practically unimportant, there is no guarantee that this will be the case and, furthermore, this method is based on a logical approach to dealing with censored observations.

Notice that the vertical axis in the above plot starts at 0: this looks better than the default which started at the minimum computed survival probability, namely 0.448. To effect this change, double-click on the vertical axis of the default plot and select **Position of ticks** in the **Major Tick Positions** box (and select where the major ticks should be - I chose 0, 20 , 40, 60, 80 , 100). Also remove the tick from **Auto** next to **Minimum** and **Maximum** in the **Scale Range** box and specify values of 0 and 100 respectively. Sadly Minitab will not let you specify a maximum of, say, 110, to allow some more space at the top of the plot.

End of solution sheet