Advanced Statistics Demo3

Michael Williams

Part 1: The table below contains the survival time for 10 patients in a clinical trial. The censored patients are defined as those who reach the endpoint of the study without dying or lost to follow-up.

Patient	Length of time in trials (years)	Outcome	Status
1	6	Survived	0
2	5	Died	1
3	3	Died	1
4	5	Lost to follow-up	0
5	5	Died	1
6	5	Died	1
7	3	Lost to follow-up	0
8	2	Died	1
9	6	Died	1
10	2	Died	1

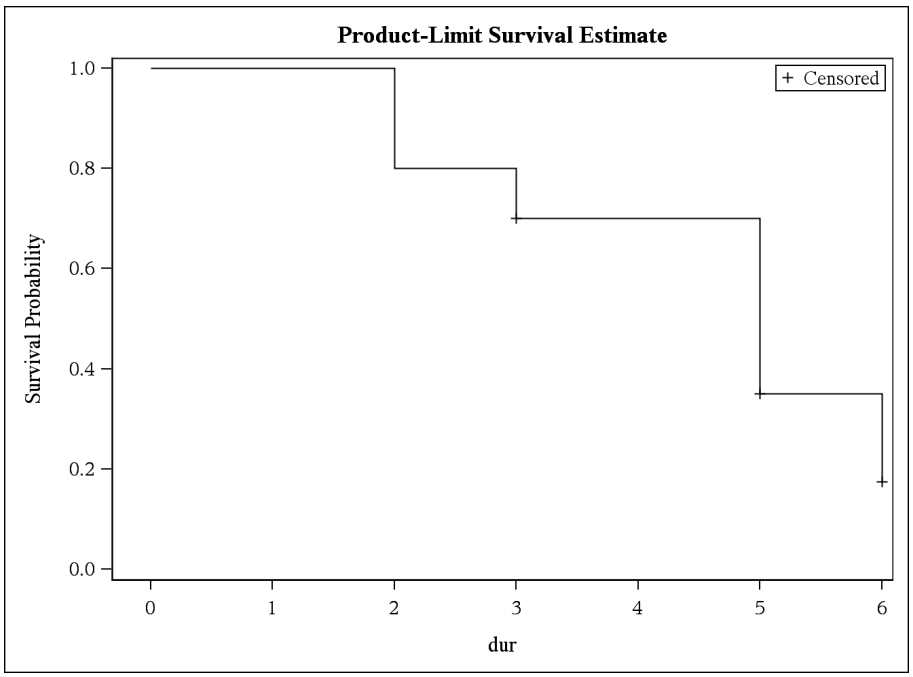
Part 1A: The following table is based on the original. The new table is used to calculate the cumulative survival S(t) according to the Kaplan-Meier Estimator

$$S(t) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{r_j}\right)$$

where d_j is the number of events at time j, and r_j is the number at risk at time j.

Subject	Time	d_j	#Censored	r_j	Cumulative Survival S(t)
	0	0	0	10	1
8,10	2	2	0	10	(1)(0.8)=0.80
3,7	3	1	1	8	(1)(0.8)(0.875)=0.70
2,4,5,6	5	3	1	6	(1)(0.8)(0.875)(0.5)=0.35
1,6	6	1	1	2	(1)(0.8)(0.875)(0.5)(0.5)=0.175

Part 1B: The survival plot for the data in Part 1A is given below.

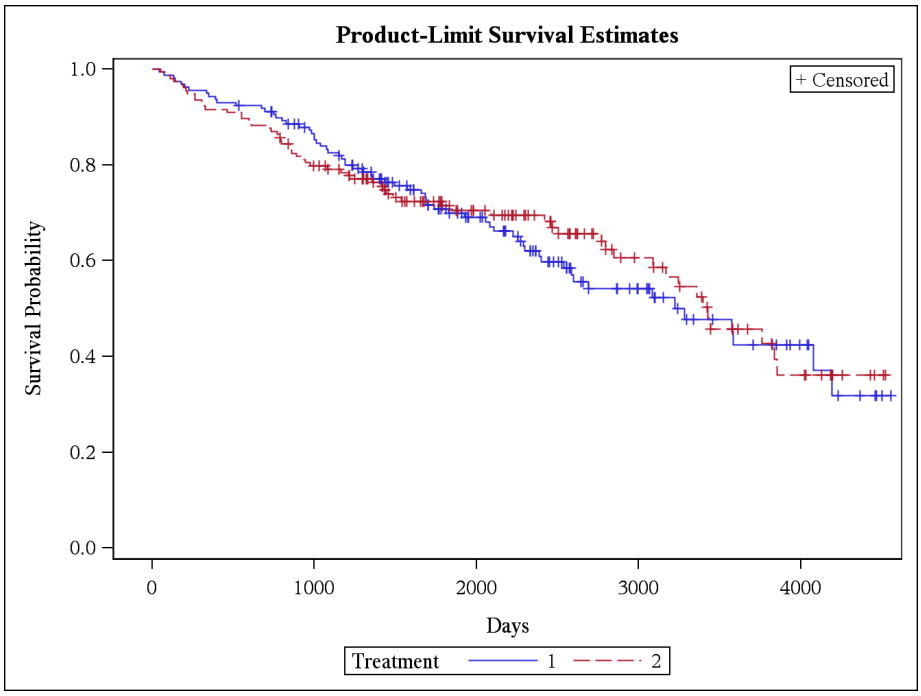


Data for Parts 2 and 3: We refer to the PBC data set that contains patient data from a clinical trial concerning primary biliary cirrhosis (PBC) of the liver. The clinical trial was conducted over a 10 year period. The clinical trial was a randomized placebo controlled trial of the drug D-penicillamine.

The variables in the PBC data set are listed below with descriptions.

Variable Name	**Description**
ID	Subject ID
DAYS	The number of days between registration and the earlier of death or termination of the follow-up (non-death event)
EVENT	1 if time to death 0 if time to censoring
TREATMENT	1 = D-pencillamine 2 = placebo
ALBUMIN	Albumin, in gm/dl
ALBUMIN_HI	1 if ALBUMIN>3.55 0 otherwise
STAGE	Histologic stage of disease, graded 1,2,3, or 4

Part 2: We use the log-rank test to compare the survival between the D-penicillamine group (Treatment 1) and the Placebo group (Treatment 2). The survival plot is given below.



The plot suggests that the survival rates of the two patient groups are close. The log rank test reveals a p-value of p=0.7498 > 0.10, so we conclude that the survival rates between the two patient groups are not significantly different.

Part 3A: We estimate the relative risk of death with the ALBUMIN variable by using Cox regression. Note that the ALBUMIN variable is continuous.

The Cox model is

$$h(t)=h_0(t)\exp(-1.79572X_{albumin}).$$

The hazard ratio is exp(-1.79572) = 0.166. In terms of relative risk, for one unit increase in albumin level, the percent change in hazard is 100(0.166-1)=-83.4 percent; so there is a 83.4 percent decrease in hazard for one unit increase in albumin. The results here are statistically significant with a p-value of <.0001.

Part 3B: We estimate the relative risk of death with the ALBUMIN_HI variable by using Cox regression. Note that the ALBUMIN_HI variable is categorical.

The Cox model is

$$h(t)=h_0(t)\exp(-1.22762X_{albumin\_hi}).$$

The hazard ratio is exp(-1.22762) = 0.293. Patients will a high level of albumin have 0.293 times the hazard of those with a lower level of albumin. In terms of relative risk, the risk of death is 70.7% lower for those with high albumin levels. The results here are statistically significant with a p-value of <.0001.

Part 3C: We determine whether or not the STAGE variable is a confounder for the association between ALBUMIN_HI and the risk of death.

When we include the various stages into the analysis, the hazard ratios are dramatically affected.

1. In stage 1, the hazard ratio is 37% larger.
2. In stage 2, the hazard ratio is 11 times larger.
3. In stage 3, the hazard ratio is 21 times larger.
4. In stage 4, the hazard ratio is 42 times larger.

We conclude that STAGE is a confounder.