Analysis of Survival data

11.1 Data Collection in Follow-up Studies

11.2 The Life-Table Method

11.3 The Product-Limit Method

1

- Survival time/Time to event/Failure data
  - Subject ID
  - Entry date (event 1: Dx)
  - Latest date (event 2: dead)
  - Status at latest date

- Will have partial info for some subjects
  - Study stopped before event occurs for some subjects
  - Subjects lost to follow-up
  - Right censoring: subject known to have survived to certain time, but it is unknown when beyond that time died (or will die)

- Main assumption: censoring independent of event of interest (time at censoring has no information on potentoial time at failure)

- Raw survival data is a list of subjects with values for survival time, status and other covariates collected at the beginning of the study or over time
- Tabulated data would look like Table 11.1 (p. 299)

- DOI: 10.1007/978-3-540-26577-1_18

2

11.2 The Life-Table Method

- Cumulative Distribution Function $F(x) = P[X \le x]$
  Probability that a subject will not survive beyond time x

- Survival Function $S(x) = P[X > x] = 1 - F(x)$
  Probability that a subject survives beyond time x

- Hazard Function $h(x) \approx P[\text{event occurs in } x + \delta]$
  Probability of event occurring in a small unit of time just after x

- Hazard Ratio/Risk Ratio/Relative Risk $RR = \dfrac{h(x \mid \text{Group A})}{h(x \mid \text{Group B})}$
  How much more/less likely are those in group A to experience the event by time x, compared to those in group B

3

Table 11.2

Years after Diagnosis	Midpoint	Death d_i	Lost l_i	Withdrawn w_i	Number Entering Interval n_i	Exposed to Risk n_i'	Conditional Probability: Dying q_i	Conditional Probability: Surviving $1-q_i$	Cumulative Probability Surviving P_i	$\frac{q_j}{n_j'(1-q_j)}$	Standard Error $SE(P_i)$
(1)		(2)				(3)	(4)	(5)	(6)		(7)
0-1	0.5	1421	68	0	2711	2677.0	0.5308	0.4692	1.0000	0.00042263	0.0000
1-2	1.5	335	19	37	1222	1194.0	0.2806	0.7194	0.4692	0.00032662	0.0096
2-3	2.5	132	17	84	831	780.5	0.1691	0.8309	0.3375	0.00026079	0.0092
3-4	3.5	64	10	47	598	569.5	0.1124	0.8876	0.2805	0.00022231	0.0089
4-5	4.5	44	12	48	477	447.0	0.0984	0.9016	0.2489	0.00024425	0.0087
5-6	5.5	20	12	39	373	347.5	0.0576	0.9424	0.2244	0.00017574	0.0086
6-7	6.5	19	10	35	302	279.5	0.0680	0.9320	0.2115	0.00026095	0.0086
7-8	7.5	14	14	19	238	221.5	0.0632	0.9368	0.1971	0.00030460	0.0086
8-9	8.5	7	10	25	191	173.5	0.0403	0.9597	0.1847	0.00024232	0.0087
9-10	9.5	7	9	19	149	135.0	0.0519	0.9481	0.1772	0.00040509	0.0088
10-11	10.5	5	4	14	114	105.0	0.0476	0.9524	0.1680	0.00047619	0.0090
11-12	11.5	5	4	17	91	80.5	0.0621	0.9379	0.1600	0.00082267	0.0093
12-13	12.5	1	4	11	65	57.5	0.0174	0.9826	0.1501	0.00030781	0.0097
13-14	13.5	3	1	15	49	41.0	0.0732	0.9268	0.1475	0.00192555	0.0099
14-15	14.5	1	0	13	30	23.5	0.0426	0.9574	0.1367	0.00189125	0.0109
15-16	15.5	0	0	7	16	12.5	0.0000	1.0000	0.1309	0.00000000	0.0119
16-17	16.5	0	0	9	9	4.5	0.0000	1.0000	0.1309	0.00000000	0.0119

$$q_i = \frac{d_i}{n_i - \frac{l_i + w_i}{2}} = \frac{d_i}{n_i'}$$

$$\widehat{\text{var}}(P_i) = P_i^2 \sum_{j=1}^{i-1} \frac{q_j}{n_j'(1-q_j)}$$

$$P_1 \equiv 1$$

$$P_2 = (1 - q_1)$$

$$P_3 = (1 - q_2)(1 - q_1) = (1 - q_2)P_2$$

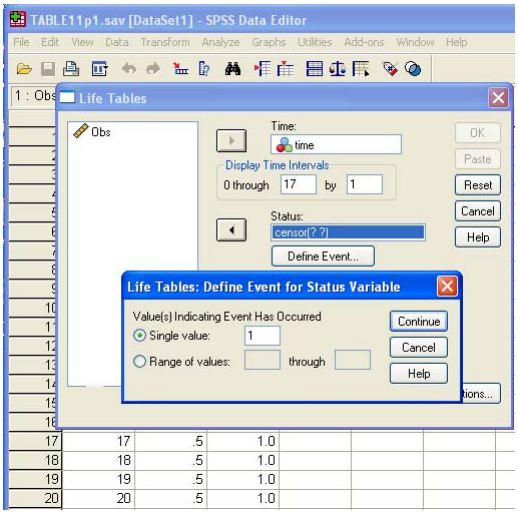
4

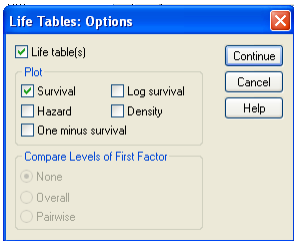
- To use tabulated survival times in SPPS, you need to 'expand' the data: e.g. for 8-9 interval create 7 observations with times in range [8,9) and status dead, and 35 observations with status censored (**for SAS usage see website**)

- **Editing a Table (**see slide #8**)**
  - Double click on Table to be modified
  - Pivot Table opens
- Change width of a column
  - Place cursor on right side of column to be modified
  - Cursor will change to ↔
  - Left click shows column width
  - Drag ↔ cursor to desired width
- Change number of decimals to display
  - Click on top cell, hold Shift key down, left click on bottom cell
  - Release Shift key, right click inside selected cells, select "Cell Properties"
  - Change "Decimals" tab

5

- Load data
- 'Analyze' → 'Survival' → 'Life Tables'
- Assign 'Time' variable; enter values in 'Display Time Intervals'
- Assign 'Status' variable; set values in 'Define events' tab
- Select from 'Options' tab: leave 'Life table(s)' checked (default); check "Survival' and 'Hazard'





6

Interval Start Time	Number Entering Interval	Number Withdrawing during Interval	Number Exposed to Risk	Number of Terminal Events	Proportion Terminating	Proportion Surviving	Cumulative Proportion Surviving at End of Interval	Std. Error of Cumulative Proportion Surviving at End of Interval
.0	2711	68	2677.0	1421	.53	.47	.4692	.009645
1.0	1222	56	1194.0	335	.28	.72	.3375	.009239
2.0	831	101	780.5	132	.17	.83	.2805	.008913
3.0	598	57	569.5	64	.11	.89	.2489	.008739
4.0	477	60	447.0	44	.10	.90	.2244	.008624
5.0	373	51	347.5	20	.06	.94	.2115	.008598
6.0	302	45	279.5	19	.07	.93	.1971	.008623
7.0	238	33	221.5	14	.06	.94	.1847	.008697
8.0	191	35	173.5	7	.04	.96	.1772	.008791
9.0	149	28	135.0	7	.05	.95	.1680	.008995
10.0	114	18	105.0	5	.05	.95	.1600	.009251
11.0	91	21	80.5	5	.06	.94	.1501	.009686
12.0	65	15	57.5	1	.02	.98	.1475	.009863
13.0	49	16	41.0	3	.07	.93	.1367	.010933
14.0	30	13	23.5	1	.04	.96	.1309	.011915
15.0	16	7	12.5	0	.00	1.00	.1309	.011915
16.0	9	9	4.5	0	.00	1.00	.1309	.011915

a. The median survival time is .942

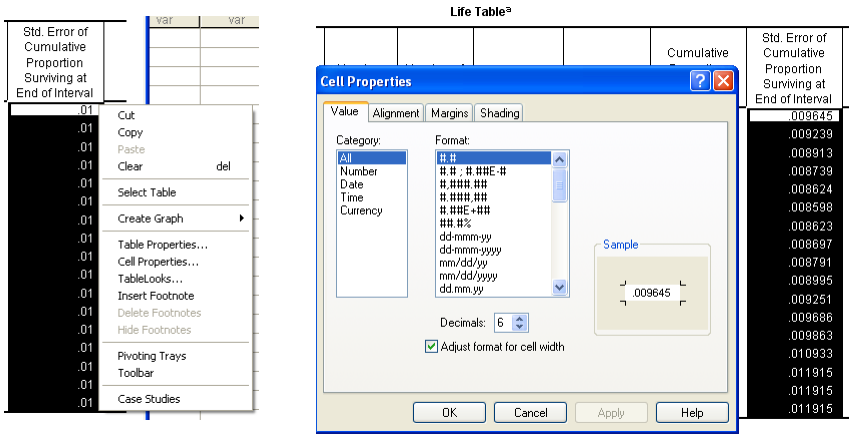
Approximate 95% percent CI for the 5-year survival probability

$$P_i \pm (z_{1-\alpha/2}) s.e.(P_i) \rightarrow 0.224 \pm 1.96(0.008624) \rightarrow (0.207, 0.241)$$

If there were 2 groups can compare survival probabilities using at certain time

7

Editing a Table



8

Descriptives of the survival function are: mean and median

- Mean affected by skewed data (especially long survivors) and if last data are censored observation
  - See pages 302-304
- Median preferred
  - See pages 304-305

Hazard rate/function

$$h_i = \frac{d_i}{a_i(n_i' - d_i/2)}$$

$$= \frac{2q_i}{a_i(2-q_i)}$$

$$\widehat{\text{var}}(h_i) = h_i^2\left(\frac{1-(h_i a_i/2)^2}{n_i' q_i}\right)$$

Cumulative Proportion Surviving at End of Interval	Std. Error of Cumulative Proportion Surviving at End of Interval	Probability Density	Std. Error of Probability Density	Hazard Rate	Std. Error of Hazard Rate
.4692	.009645	.5308	.009645	.7226	.017874
.3375	.009239	.1316	.006674	.3264	.017592
.2805	.008913	.0571	.004791	.1847	.016011
.2489	.008739	.0315	.003845	.1191	.014857
.2244	.008624	.0245	.003612	.1035	.015587
.2115	.008598	.0129	.002848	.0593	.013245
.1971	.008623	.0144	.003238	.0704	.016134
.1847	.008697	.0125	.003269	.0653	.017434
.1772	.008791	.0075	.002781	.0412	.015560
.1680	.008995	.0092	.003413	.0532	.020113
.1600	.009251	.0080	.003518	.0488	.021809
.1501	.009686	.0099	.004343	.0641	.028653
.1475	.009863	.0026	.002593	.0175	.017543
.1367	.010933	.0108	.006042	.0759	.043818
.1309	.011915	.0058	.005711	.0435	.043468
.1309	.011915	.0000	.000000	.0000	.000000
.1309	.011915	.0000	.000000	.0000	.000000

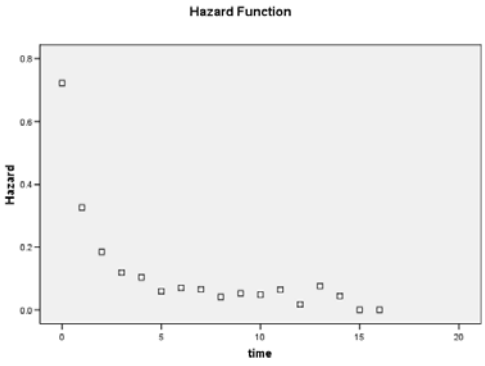
9

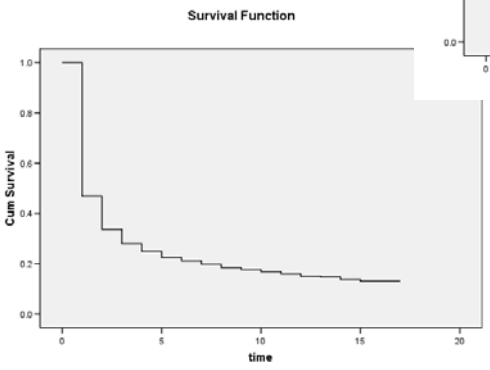
A 95% CI for the mortality rate in the first year is

$$h_i \pm (z_{1-\alpha/2})s.e.(h_i)$$

$$0.723 \pm 1.96(0.0179)$$

$$(0.688, 0.758)$$





10

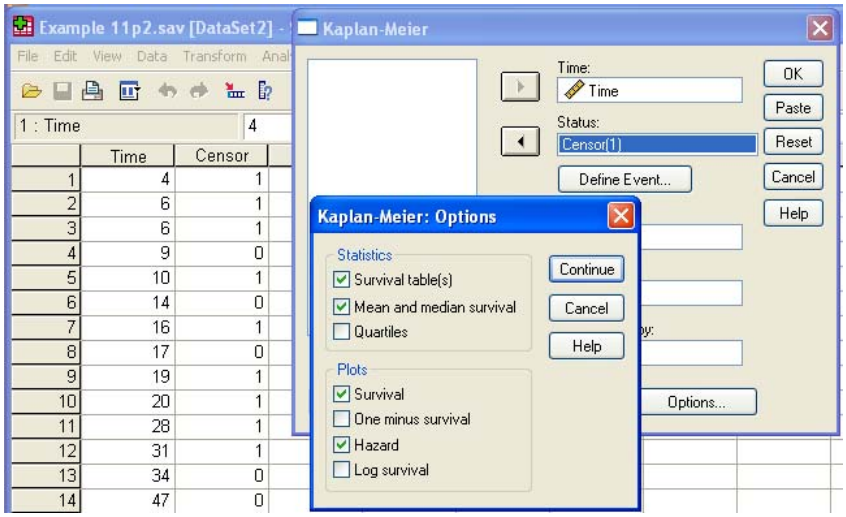
11.3 The Product-Limit Method

- For small sample sizes
- Also known as the Kaplan-Meier method
- Goal: to estimate survival function without grouping data
- Principles for estimation are similar to Life-Table method

- See Example 11.2 (p. 306-307)
  - Table 11.4 is data from clinic records
  - Calculate survival times
    - Subject 1: Discharge January 1990; Termination December 1993; Status Still sober (withdrawn) → time=47, status=0 (censored) → 47*
  - Procedure on page 307-309

- Using SPSS
  - Type-in all data (Times and corresponding status)
  - Analyze → Survival → Kaplan-Meier
  - Assign 'Time' variable
  - Assign 'Status' variable → Set status codes in 'Define Event' tab
  - Leave default checks ('Survival table' and 'Mean and median survival')
  - Check 'Survival' and 'Hazard' plots

11



12

Case Processing Summary

Total N	N of Events	Censored	
		N	Percent
14	9	5	35.7%

Survival Table

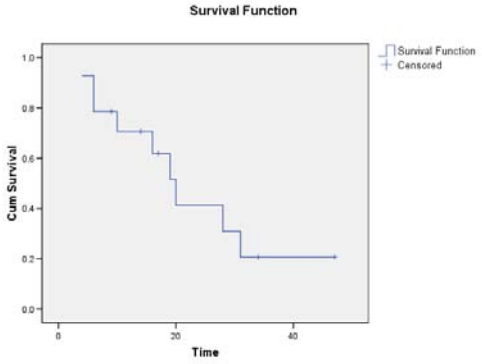
	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
			Estimate	Std. Error		
1	4	1	.9286	.068830	1	13
2	6	1	.	.	2	12
3	6	1	.7857	.109664	3	11
4	9	0	.	.	3	10
5	10	1	.7071	.123683	4	9
6	14	0	.	.	4	8
7	16	1	.6188	.136194	5	7
8	17	0	.	.	5	6
9	19	1	.5156	.147456	6	5
10	20	1	.4125	.149745	7	4
11	28	1	.3094	.143490	8	3
12	31	1	.2063	.127439	9	2
13	34	0	.	.	9	1
14	47	0	.	.	9	0

13

Means and Medians for Survival Time

Mean[a]				Median			
		95% Confidence Interval				95% Confidence Interval	
Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
23.143	4.297	14.722	31.564	20.000	2.904	14.308	25.692

a. Estimation is limited to the largest survival time if it is censored.



14