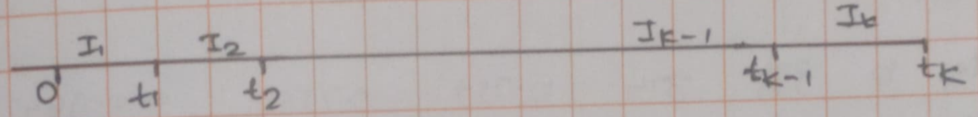
Practical 07 : Acturial and Kaplan Meier Estimation of Survival Curve

(1) Acturial Method

Let the time interval $[0, t_j]$ are under consideration and this interval are partioned as $I_1, I_2, \ldots I_K$



Some basic notation

$n_i \rightarrow$ no of individual at risk in beginning interval I_i

$d_i \rightarrow$ no of death during the interval I_i

$w_i \rightarrow$ no of withdrawn during the interval I_i

$l_i \rightarrow$ no of ~~least~~ lost to followup during the interval I_i

Formula

Estimation Survival function $\hat{S}(t)$ under acturial method as follows

$$\hat{S}(t) = \prod_{i=1}^{k} \hat{p}_i$$

where, $\hat{p}_i = 1 - \frac{d_i}{n_i}$ [if there is no existence of l_i's and w_i]

$\hat{p}_i = 1 - \frac{d_i}{n_i'}$ [if there is existence]

$n_i' = n_i = \frac{1}{2}[l_i + w_i]$

$\hat{q}_i = \frac{d_i}{n_i'}$

$1 - p_i = 1 - \frac{d_i}{n_i'}$

I_i	d_i	w_i	n_i	n_i'	$\hat{p}_i$	$\hat{S}(t_i)$	
[0,1)	90	0	374	374	0.759	0.759	0.00084
[1,2)	76	0	284	254	0.732	0.55	0.00128
[2,3)	51	0	208	208	0.754	0.469	0.00156
[3,4)	25	12	159	151	0.834	0.3502	0.00131
[4,5)	20	5	120	117.5	0.929	0.2406	0.0974
[5,6)	7	9	95	90.5	0.922	0.260	0.00092
[6,7)	4	9	79	74.5	0.946	0.252	0.00076
[7,8)	1	3	66	64.5	0.984	0.248	0.00024
[8,9)	3	5	62	59.5	0.949	0.235	0.00089
[9,10)	2	5	54	51.5	0.961	0.226	0.00098
[10,∞)	47	0	47	47	0	0	

To estimate the probability of survival $\hat{p}_i$ beyond interval

variance of estimate P_s is given by

$$Var(P_s) = (\hat{P}_s)^2 \Rightarrow \hat{S}[t_s]^2 \Rightarrow \sum_{i=1}^{s} \frac{d_i}{n_i'(n_i' - d_i)}$$

$$Var[\hat{S}(t_s)] = 0.00056$$

(2) <u>objective</u> :

To estimate the Survival function at t, $3(t)$ · both groups using K–M estimation

plot K-M Survival curves for both groups on same graph

To estimate the median of survival time and variance of $\hat{S}(t_{100})$ for ~~both~~ the groups

obtain 95% Confidence interval for survival function at 1.02 months for both groups

Some basic notation are :

$$\hat{S}(t) = \prod_{T(i) \le t} (\hat{p}_i)^{\delta(i)}$$

where

$$\hat{p}(i) = \begin{cases} 1 - d_i/n_i & \text{if } \delta(i) = 1 \\ 1 & \text{if } \delta(i) = 0 \end{cases}$$

Calculation in following table

Subject	Time	Status	d_i	c_i	n_i	$\hat{p}_i$	$\hat{S}(t)$	$\hat{q}_i$
1	6	NED	0	1	25	1	1	0
2	8	NED	0	1	24	1	1	0
3	11	~~DOD~~	1	0	23	0.956	0.986	–
4	15	DOD	1	0	22	1	0.956	0.044
5	15	DOD	2	0	22	0.909	0.869	0
6	21	DOD	1	0	20	0.95	0.826	0.091
7	26	NED	0	1	19	1	0.826	0.05

8	29	DOD	1	0	18	0.944	0.780	0
9	33	DOD	1	0	17	0.941	0.734	0.056
10	34	NED	0	1	16	1	0.734	0.059
11	35	NED	0	1	15	1	0.734	0
12	79	DOD	1	0	12	0.928	0.682	0
13	82	NED	0	1	13	1	0.682	0.071
14	95	NED	0	1	12	1	0.682	0
15	102	DOD	1	0	11	0.904	0.620	0
16	109	NED	0	1	10	1	0.62	0.091
17	109	NED	0	1	10	1	0.62	0
18	117	NED	0	1	8	1	0.62	0
19	127	NED	0	1	7	1	0.62	0
20	129	NED	0	1	6	1	0.62	0
21	137	NED	0	1	5	1	0.62	0
22	138	NED	0	1	4	1	0.62	0
23	155	NED	0	1	3	1	0.62	0
24	212	DOD	1	0	2	0.5	0.62	0
25	337	NED	0	1	1	1	0.31	0.5
							0.31	0

$q_i/n_i \hat{p}_i$

0 0.0093 0.0035 0 0 0 0 0.0045
0.0037 0 0 0 0 0.0026 0 0 0
0 0.0019 0 0 0.0091 0 1

$$\text{var}\left[\hat{S}(t)\right]^2 \sum_{T(i) \le k} \frac{q_i}{n_i \hat{p}_i}$$

(iii) Estimate of median time of Survival (0,1)

$\frac{1}{2} = 0.5.$

median of $\hat{S}(t) = 0.5.$

$155 \le t \le 212$ where,

$$\hat{S}(t_{155}) = 0.6199$$
$$\hat{S}(t_{212}) = 0.3099$$

$$\frac{155 + 212}{2} = 183.5$$
$$= 184 \text{ months.}$$

median of Survival time may be 183 month

$$\widehat{Var}\,\hat{S}(t_{102}) = [S(t_{102})]^2 \sum_{T(i) \le t_{102}} \frac{q_i}{n_i(\hat{p}_i)}$$

$$= (0.6199)^2 \cdot (0.0307) = 0.012$$

Variance for lower grade Tumour group is 0.012

(iv) The $(1-\alpha)100\%$ Confidence interval for $S(t)$ is given by

$$\hat{S}(t) \pm \sqrt{\tfrac{1}{2n} \log (2/\alpha)}$$

95% Confidence Interval for lower grade Tumour group is

$$\hat{S}(t_{102}) \pm \sqrt{\frac{1}{2 \times 25} \log\left(\frac{2}{0.05}\right)} \Rightarrow \sqrt{\frac{1}{50} \log\left(\frac{2}{0.05}\right)}$$

$$= \pm 0.179$$
$$\Rightarrow 0.6199 \pm 0.179$$
$$(0.4409,\ 0.798)$$

Lower bound is 0.4409
upper bound is 0.798

Calculation for high grade tumour group

Subject	Time	Status	d_i	c_i	n_i	p_i	$\hat{S}(t)$	$\hat{q}_i$	$\frac{\hat{q}_i}{n_i \hat{p}_i}$
13	0	DOD	1	0	14	0.929	0.929	0.071	0.005
5	2	DOD	1	0	13	0.923	0.857	0.071	0.006
14	3	DOD	1	0	12	0.917	0.786	0.085	0.007
12	4	DOD	1	0	11	0.909	0.714	0.091	0.0091
3	6	DOD	1	0	10	0.9	0.692	0.1	0.011
4	7	DOD	1	0	9	0.809	0.511	0.111	0.014
6	9	DOD	1	0	8	1	0.411	0	0
10	19	DOD	1	0	7	0.75	0.408	0.25	0.043
11	12	DOD	1	0	6	0.833	0.357	0.167	0.033
8	17	DOD	1	0	5	0.8	0.286	0.2	0.05
7	17	DOD	1	0	4	0.75	0.214	0.25	0.083
9	23	DOD	1	0	3	0.669	0.143	0.33	0.166
2	27	DOD	1	0	2	0.5	0.0715	0.5	0.5
1	102	NED	0	1	1	1	0.0715	0	0

Median of Occuring at 9 months

$$\text{Var}\left(\hat{S}(t_{102})\right) = \left[\hat{S}(t_{102})\right]^2 \sum_{T(i) \le t_{102}} \frac{q_i}{n_i \hat{p}_i}$$

$$= (0.0715)^2 \, (0.4193)$$

$$= 0.0047$$

(ii) $\hat{S}(t) \pm \sqrt{\tfrac{1}{2n} \cdot \log \tfrac{2}{\alpha}}$

$\alpha = 0.05$

$$\hat{S}(t_{102}) \pm \sqrt{\tfrac{1}{2n} \log\left(\frac{2}{0.05}\right)}$$

$$= 0.0715 \pm 0.239$$

Confidence Interval are (0.0715 ± 0.239)

Codes:

```
# Lower Grade Tumour Group
data_lower = {
    "Subject": [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 9, 10, 11, 12,
                13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25],
    "Time": [29, 129, 79, 138, 21, 95, 137, 6, 212, 11, 212, 11, 15, 337,
             82, 33, 75, 109, 26, 117, 8, 127, 155, 102, 34, 109, 15],
    "Status": ["dod", "ned", "dod", "ned", "dod", "ned", "ned", "ned", "dod",
               "dod", "dod", "dod", "dod", "ned", "ned", "dod", "ned", "ned", "ned",
               "ned", "ned", "ned", "ned", "dod", "ned", "ned", "dod"]

}
df_lower = pd.DataFrame(data_lower)
df_lower['Event'] = df_lower['Status'].apply(lambda x: 1 if x == 'dod' else 0)

# High Grade Tumour Group
data_high = {
    "Subject": [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14],
    "Time": [102, 27, 6, 7, 2, 9, 17, 16, 23, 9, 12, 4, 0, 3],
    "Status": ["ned", "dod", "dod", "dod", "dod", "dod", "dod",
               "dod", "dod", "dod", "dod", "dod", "dpo", "dod"]

}
df_high = pd.DataFrame(data_high)
df_high['Event'] = df_high['Status'].apply(lambda x: 1 if x == 'dod' else 0)

# Kaplan-Meier Estimator
kmf_lower = KaplanMeierFitter()
kmf_high = KaplanMeierFitter()

plt.figure(figsize=(10, 6))


# Fit for Lower Grade Tumour Group
kmf_lower.fit(durations=df_lower['Time'], event_observed=df_lower['Event'], label='Lower Grade Tumour')
kmf_lower.plot_survival_function()

# Fit for High Grade Tumour Group
kmf_high.fit(durations=df_high['Time'], event_observed=df_high['Event'], label='High Grade Tumour')
kmf_high.plot_survival_function()

# Plot formatting
plt.title('Kaplan-Meier Survival Curves for Tumour Groups')
plt.xlabel('Time (Days)')
plt.ylabel('Survival Probability')
plt.legend()
plt.grid()
plt.show()
```

Output: