10/1/19

M156: Mortality Studies using Seriatim Data

Case 1: Complete Individual Data
Case 2: Complete Grouped Data
} See Notes from Last Thursday

Case 3: $\underbrace{\text{Incomplete}}$ Individual Data (today)

means we don't know when every dragon dies. Some dragons may fly away (quit the study) at a specific time, and so we only know the dragon died after that specific time (right-censored data)

In this case we first organize the data using the following standard notation:

x_i: denotes the time dragon i dies

u_i^+: denotes the time dragon i flys away (censored)

y_j: denotes the times at which the dragons die in increasing order (the x_i but not u_i^+ values)

s_j: denotes the number of dragons that die at time y_j

b_j: denotes the number of dragons that fly away in the interval $[y_j, y_{j+1})$

r_j: denotes the number of dragons at risk of being observed by us to die at time y_j

Generally, $r_j = r_{j-1} - s_{j-1} - b_{j-1}$

Examples

M1S6 Example 1:

In a mortality study on a sample of 40 dragons with censored data, fill in the missing values.

i	y_i	s_i	b_i	r_i
1	4	3	-	-
2	6	-	3	31
3	9	6	4	23
4	13	4	-	-
5	15	2	4	6

M6S1 Example 1:

In a mortality study on a sample of 40 dragons with censored data, fill in the missing values.

i	y_i	s_i	b_i	r_i
1	4	3	- 6	- 40
2	6	- 5	3	31
3	9	6	4	23
4	13	4	- 3	- 13
5	15	2	4	6

M1S6 Example 2:

In a mortality study of a cohort of twelve 80-year olds, you are given the following observed exit times:

1+	2	2	2+	4	4+	5	5+	7	8	9+	9+

Enter values in the following table for all that can be done.

i	y_i	s_i	b_i	r_i

Now suppose there is a 13th observation that is truncated at time 3 (entered at time 3) and is right censored before time 4 (exits before time 4). How does that affect the r and s values in the table above?

M1S6 Example 2:

In a mortality study of a cohort of twelve 80-year olds, you are given the following observed exit times:

1+ 2 2 2+ 4 4+ 5 5+ 7 8 9+ 9+

Enter values in the following table for all that can be done.

i	y_i	s_i	b_i	r_i
1	2	2	1	11
2	4	1	1	8
3	5	1	1	6
4	7	1	0	4
5	8	1	2	3

don't include the one exit at time 1

Now suppose there is a 13th observation that is truncated at time 3 (entered at time 3) and is right censored before time 4 (exits before time 4). How does that affect the values in the table above?

r & s

No change in r & s values.

Estimating Survival & Cumulative Hazard Functions

Recall $\quad {}_xp_0 = e^{-\int_0^x \mu_t \, dt}$ (actuaries)

$$S_X(x) = e^{-\int_0^x h_X(t)\,dt} \quad \text{(stats)} \quad h\text{ - hazard function}$$

Define $H_X(x) = \int_0^x h_X(t)\,dt$ = cumulative hazard function

Punchline: $S_X(x) = e^{-H_X(x)}$

$$\Longleftrightarrow \quad H_X(x) = -\ln(S_X(x))$$

Formulas:

1) Kaplan-Meier (Product-Limit) Estimate of $S_X(t)$ (based on a sample of size n)

$$S_X(t) \approx S_n(t) = \prod_{y_j \le t} \left(1 - \frac{s_j}{r_j}\right)$$

Then $H_X(t) \approx H_n(t) = -\ln(S_n(t))$

2) Nelson-Aalen Estimate of $H_X(t)$

$$H_X(t) \approx \hat{H}(t) = \sum_{y_j \le t} \frac{s_j}{r_j}$$

Then $S_X(t) \approx \hat{S}(t) = e^{-\hat{H}(t)}$