MIS6: Morbidity Studies (on Dragons)

Case 1: Complete Individual Data (Done)

Case 2: Complete Grouped Data

in intervals

In this case, we form the ogive by linearly interpolating between the \( F(a) \)-values where \( a \) is an endpoint of an interval.

*Note:* \( F(0) = \frac{\text{CDF at } t=0}{3} = \Pr(T \leq 0) = 0 \)

\[ F(\infty) = 1 \]

Example: LMS/soA

\[ ? F(12) \]

\[ F(10) = \frac{28 + 19}{100} = .47 \]

\[ F(20) = \frac{28 + 19 + 15}{100} = .62 \]

\[ \therefore F(12) = .8 F(10) + .2 F(20) = .5 \]

Case 3: Incomplete Individual Data

means we do not know when each dragon dies. Some dragons may fly away \( \& \) at a specific time, and so we only know the dragon died after that specific time (right-censored data).

Step 1: Organize the data (hard part, but it’s easy)
Notation:

1) $y_i$ denotes the times at which the dragons die in increasing order: $y_1 < y_2 < y_3 \ldots$.

2) $s_i$ denotes the number of dragons that die at time $y_i$.

3) $b_i$ denotes the number of dragons that flew away in the interval $[y_i, y_{i+1})$.

4) $r_i$ denotes the number of dragons at risk of being observed by me to die at time $y_i$.

Note: $r_1 = n = \#$ of dragon in the study if none flew away between time 0 and $y_1$.

For $i > 1$,

$$r_i = r_{i-1} - s_{i-1} - b_{i-1}$$

Example: (see notes) → next page

Estimating Survival Functions

Note: Recall:

$$s(x) = e^{-\int_0^x h(t) dt} = e^{-H(x)}$$

where $H(x)$ is called the cumulative hazard function.
For a mortality study on 40 dragons with censored data, fill in the missing values.

Assume $r_i = 40$.

<table>
<thead>
<tr>
<th>$i$</th>
<th>$y_i$</th>
<th>$s_i$</th>
<th>$b_i$</th>
<th>$r_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>-5</td>
<td>3</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>6</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>4</td>
<td>-6</td>
<td>-8</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>
Formulas: (to estimate $S_x(t)$)

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

$$S_n(t) = \prod_{j \leq t} \left(1 - \frac{S_j}{r_j}\right)$$

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

$$H_n(t) = \sum_{j \leq t} \frac{S_j}{r_j}$$

Then the estimate of $S_x(t)$ is

$$\hat{S}_x(t) = e^{-H_n(t)}$$

Formulas: (to estimate the variance of the estimator for $S_x(t)$)

1) Greenwood's Approximation

$$V_n = \text{Var} (S_n(t)) = \left[S_n(t)\right]^2 \sum_{j \leq t} \frac{S_j}{r_j (r_j - S_j)}$$

2) Klein's Approximation

$$V_n = \text{Var} (S_n(t)) = \left[S_n(t)\right]^2 \sum_{j \leq t} \frac{S_j (r_j - S_j)}{r_j^3}$$

3) Aalen's Approximation

$$V_n = \text{Var} (S_n(t)) = \left[S_n(t)\right]^2 \sum_{j \leq t} \frac{S_j}{r_j^2}$$

(replace $r_j - S_j$ in Klein's approximation by $r_j$)
A cohort of 100 newborns is observed from birth. During the first year, 10 drop out of the study and one dies at time 1. Eight more drop out during the next six months, then, at time 1.5, three deaths occur.

Compute the Nelson-Aalen estimator of the survival function, $\hat{S}(1.5)$.

Solution: Since 10 drop out before the first death, $n_i = 100 - 10 = 90$

<table>
<thead>
<tr>
<th>$i$</th>
<th>$y_i$</th>
<th>$s_i$</th>
<th>$b_i$</th>
<th>$r_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>3</td>
<td>81</td>
<td>$= 90 - 1 - 8$</td>
</tr>
</tbody>
</table>

\[ \hat{A}(1.5) = \sum_{y_i = 1.5}^{N_A} \frac{s_j}{r_j} = \frac{1}{90} + \frac{3}{81} = 0.0481 \]

\[ \Rightarrow \hat{S}(1.5) = e^{-0.0481} = 0.953 \]