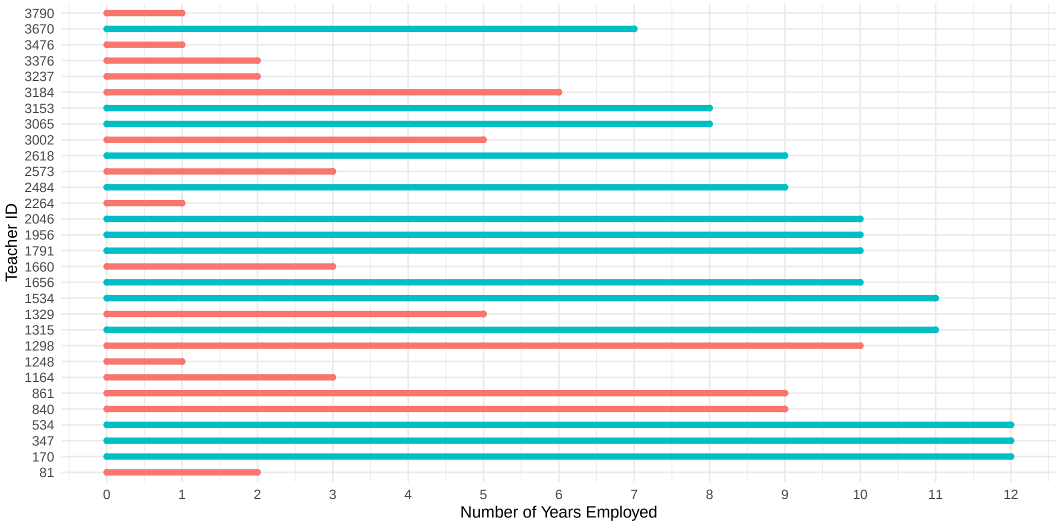
Introduction to Discrete time to Event Data

Teachers Employment Data



Discrete Time

Suppose we are interested in the event occurence time, T_i.

If $T_i \in [k, k+1)$ then we **discretize** and say $T_i = k$.

Goals of Discrete Time-to-Event Analysis

1. Estimate the **hazard function**, $h(k)$.

Goals of Discrete Time-to-Event Analysis

1. Estimate the **hazard function**, $h(k)$.
2. Estimate the **survivor function**, $S(k)$.

Goals of Discrete Time-to-Event Analysis

1. Estimate the **hazard function**, $h(k)$.
2. Estimate the **survivor function**, $S(k)$.
3. Characterize measures of **central tendency**.

Goals of Discrete Time-to-Event Analysis

1. Estimate the **hazard function**, $h(k)$.
2. Estimate the **survivor function**, $S(k)$.
3. Characterize measures of **central tendency**.
4. **Likelihood** type analyses.

The Hazard Function

Definition

Recall that the **discrete hazard function** is defined as

$$h(t) = P(T = t | T \geq t).$$

Intuitive Estimator - Hazard Function

Suppose that at time k there are r_k **individuals at risk**.

Intuitive Estimator - Hazard Function

Suppose that at time k there are r_k **individuals at risk**.

Of these, d_k **individuals experience the event**.

Intuitive Estimator - Hazard Function

Suppose that at time k there are r_k **individuals at risk**.

Of these, d_k **individuals experience the event**.

This gives an *obvious* **estimator** of

$$\widehat{h}(k) = \frac{d_k}{r_k}.$$

Interpretation and Life Tables

This is the **empirical proportion** of those experiencing the event out of **those under observation and at risk**. As a result, it intuitively accounts for censoring.

Interpretation and Life Tables

Year	Number At Risk	Number of Events	Number Censored	Hazard Function
0	3941	0	0	0.000
1	3941	456	0	0.116
2	3485	384	0	0.110
3	3101	359	0	0.116
4	2742	295	0	0.108
5	2447	218	0	0.089
6	2229	184	0	0.083
7	2045	123	280	0.060
8	1642	79	307	0.048
9	1256	53	255	0.042
10	948	35	265	0.037
11	648	16	241	0.025
12	391	5	386	0.013

Survivor Function

Intuitive Estimator - Survivor Function

Using a similar argument, we may be tempted to take

$$\widehat{S}(k) = \frac{r_k}{n}.$$

This **will not** work.

Relationship to Hazard

Recall that the **survivor function** can be expressed as

$$S(k) = P(T \geq k) = \prod_{j=1}^{k} \{1 - h(j)\}.$$

Estimation of the Survivor Function

This gives us the **plug-in estimator** given by

$$\widehat{S}(k) = \prod_{j=1}^{k} \left\{1 - \widehat{h}(j)\right\} = \prod_{j=1}^{k} \left\{1 - \frac{d_j}{r_j}\right\}.$$

Measures of Central Tendency

Characterizing the Distribution

- The **mean** will be biased downwards.

Characterizing the Distribution

- The **mean** will be biased downwards.
- In general, **nonparametric** mean estimation will not be possible.

Characterizing the Distribution

- The **mean** will be biased downwards.
- In general, **nonparametric** mean estimation will not be possible.
- Can instead use the **median**.

Characterizing the Distribution

- The **mean** will be biased downwards.
- In general, **nonparametric** mean estimation will not be possible.
- Can instead use the **median**.
- Take m to be such that $\widehat{S}(m) > 0.5 > \widehat{S}(m+1)$, then

$$\text{median} = m + \left[\frac{\widehat{S}(m) - 0.5}{\widehat{S}(m) - \widehat{S}(m+1)} \right].$$

Teaching Example

Year	Hazard Function	Survivor
0	0.000	1.000
1	0.116	0.884
2	0.110	0.787
3	0.116	0.695
4	0.108	0.620
5	0.089	0.565
6	0.083	0.518
7	0.060	0.487
8	0.048	0.464
9	0.042	0.444
10	0.037	0.428
11	0.025	0.417
12	0.013	0.412

Likelihood Based Analyses

Time-to-Event as Stochastic Process

Just like with **transition models**, we can take

$$\{\, Y(s) \in \{0, 1\} \mid s = 0, 1, 2, \ldots \,\},$$

to be a **stochastic process** representing the outcome process, with $Y(s) = 0$ when the event has not been observed and $Y(s) = 1$, otherwise.

Time-to-Event as Stochastic Process

Just like with **transition models**, we can take

$$\{Y(s) \in \{0, 1\} \mid s = 0, 1, 2, \ldots\},$$

to be a **stochastic process** representing the outcome process, with $Y(s) = 0$ when the event has not been observed and $Y(s) = 1$, otherwise.

Further define

$$\{Z(s) \in \{0, 1\} \mid s = 0, 1, 2, \ldots\},$$

to be the **censoring process** with $Z(s) = 1$ when the individual is **still under observation** and $Z(s) = 0$ otherwise.

History Vectors

Also as before take

$$\mathcal{H}^Y(s) = (Y(0), Y(1), \ldots, Y(s-1)),$$

in addition to the corresponding vectors for the **observation process**, $\mathcal{H}^Z(s)$, and the **joint process** $\mathcal{H}^{Y,Z}(s)$.

Complete Likelihood

For any individual, we are interested in the **likelihood contribution** given by

$$L_i = \prod_{s=1}^{\infty} P(Y(s), Z(s)|\mathcal{H}^{Y,Z}(s)) = \prod_{s=1}^{\infty} P(Y(s)|Z(s), \mathcal{H}^{Y,Z}(s)) \times P(Z(s)|\mathcal{H}^{Y,Z}(s)).$$

Simplifying Assumptions

We typically make **two simplifying assumptions**

1. **Conditionally Independent Censoring:** Under conditionally independent censoring we have

$$P(Y(s)|Z(s), \mathcal{H}^{Y,Z}(s)) = P(Y(s)|\mathcal{H}^{Y}(s)).$$

Simplifying Assumptions

We typically make **two simplifying assumptions**

1. **Conditionally Independent Censoring:** Under conditionally independent censoring we have

$$P(Y(s)|Z(s), \mathcal{H}^{Y,Z}(s)) = P(Y(s)|\mathcal{H}^{Y}(s)).$$

2. **Non-Informative Censoring:** Under non-informative censoring we assume that $P(Z(s)|\mathcal{H}^{Y,Z}(s))$ and $P(Y(s)|\mathcal{H}^{Y}(s))$ are **functionally independent**.

Simplified Likelihood

Under these assumptions, the **partial likelihood** becomes

$$L_i = \prod_{s=1}^{\infty} P(Y(s)|\mathcal{H}^Y(s)) = \prod_{s=1}^{C_i} h_i(s;\theta)^{Y_i(s)}(1 - h_i(s;\theta))^{1-Y_i(s)}.$$

Here $h_i(s;\theta)$ is **some parametric model** for the hazard!

Does this look familiar?

Summary

- When we **discretize** time, we can use **discrete-time survival methods**.

Summary

- When we **discretize** time, we can use **discrete-time survival methods**.
- Using **life tables** we can use the **empirical estimator** for the hazard function.

Summary

- When we **discretize** time, we can use **discrete-time survival methods**.
- Using **life tables** we can use the **empirical estimator** for the hazard function.
- The **survivor function** can be estimated using the form of the **hazard function**.

Summary

- When we **discretize** time, we can use **discrete-time survival methods**.
- Using **life tables** we can use the **empirical estimator** for the hazard function.
- The **survivor function** can be estimated using the form of the **hazard function**.
- The **median** can be estimated as a measure of central tendency.

Summary

- When we **discretize** time, we can use **discrete-time survival methods**.
- Using **life tables** we can use the **empirical estimator** for the hazard function.
- The **survivor function** can be estimated using the form of the **hazard function**.
- The **median** can be estimated as a measure of central tendency.
- We can use **stochastic process** notation to write down the likelihood expression.

Summary

- When we **discretize** time, we can use **discrete-time survival methods**.
- Using **life tables** we can use the **empirical estimator** for the hazard function.
- The **survivor function** can be estimated using the form of the **hazard function**.
- The **median** can be estimated as a measure of central tendency.
- We can use **stochastic process** notation to write down the likelihood expression.
- Under the assumptions of **conditionally independent** and **non-informative** censoring, this becomes a **binomial likelihood**.