Introduction

What is survival analysis?

- The analysis of "time-to-event" (or "failure time") data.
- Examples:
  - (*Biomedical science*) Time to death; to recurrent of disease
  - (*Engineering*) Time for machine to break down
  - (*Social science* Time to events such as, job change, marriage
  - (*Financial industry*) Time to default on a loan; to bankruptcy.

Challenges

- Main difficulty of analyzing survival data is that some observations are **censored**.
- That is, the event is not observed during the period of study.
- What is censoring?

Censoring: a toy example

Four lung cancer patients are recruited to study the effect of one drug on their survival from lung cancer. The variable of interest is "time to death due to lung cancer".

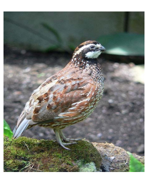
- Patient 1 participated the study until his death at time T_1. (No censoring)
- Patient 2 participated the study until he withdrew from the study at time T_2. (Censored)
- Patient 3 participated the study until time T_3, when he died from a car accident. (Censored)
- Patient 4 participated the whole study and was still alive at the end of study T_4. (Censored)

Challenges

- In survival analysis, the observations are subject to censoring.
- No clear choice of a parametric family of distribution.
  - The main outcome of the survival experiment is the time-to-event, which is almost always nonnegative.
  - The distribution of time-to-event is typically skewed.
- Nonparametric and semiparametric methods are more popular.
  - Making inference of survival data without any distribution assumptions.
- Unlike other statistics courses, survival analysis focuses on different parameters like survival and hazard functions.

Example 1: Northern Bobwhite

- 18 birds were radio-tagged and followed for 13 weeks from March to May in 1985.
- Pollock et al. (1989). J. Wildlife Man. 53:7-15
- Goal: investigate the spring survival



Example 1: Northern Bobwhites

Table 1. Kaplan-Meier survival estimates for northern bobwhite quail radiotagged in North Carolina, spring 1985.

Week (t)	Dates	No. at risk (r_j)	No. deaths (d_j)	No. censored	Survival ($\hat{S}[t]$)	95% CI
1	3 Mar–9 Mar	18	0	0	1.0000	1.0000–1.0000
2	10 Mar–16 Mar	18	0	0	1.0000	1.0000–1.0000
3	17 Mar–23 Mar	18	2	0	0.8889	0.7520–1.0258
4	24 Mar–30 Mar	16	0	0	0.8889	0.7437–1.0341
5	31 Mar–6 Apr	16	0	0	0.8889	0.7437–1.0341
6	7 Apr–13 Apr	16	1	0	0.8333	0.6667–1.0000
7	14 Apr–20 Apr	15	0	0	0.8333	0.6612–1.0055
8	21 Apr–27 Apr	15	1	1	0.7778	0.5922–0.9633
9	28 Apr–4 May	13	1	2	0.7179	0.5107–0.9252
10	5 May–11 May	10	1	1	0.6462	0.4079–0.8844
11	12 May–18 May	8	0	0	0.6462	0.3798–0.9125
12	19 May–25 May	8	0	1	0.6462	0.3798–0.9125
13	26 May–1 Jun	7	0	0	0.6462	0.3614–0.9309

Example 1: Northern Bobwhites

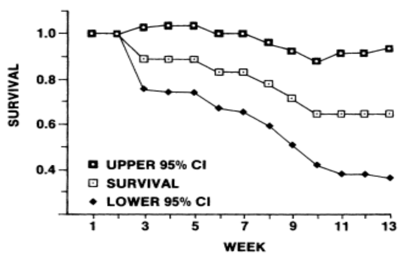


Fig. 1. The Kaplan-Meier survival function for northern bob-white quail radiotagged in North Carolina, spring 1985.

Example 1: Northern Bobwhites

- **One population estimation problem:**

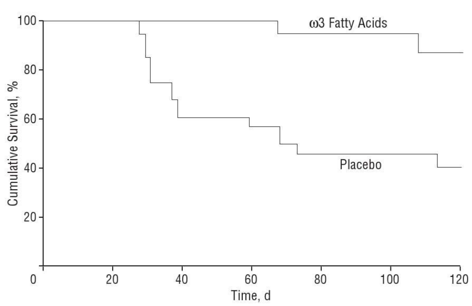
Let $X_1, \cdots, X_n$ be a random sample of survival times, some are possibly censored.

How to estimate the distribution of the survival time?

Example 2: Bipolar disorder

- Stoll et al. (1999). Arch. Gen. Psychiatry *56* 407-12.
- Goal: Test if Omega 3 fights bipolar disorder.
- 3 patients with bipolar disorder were recruited.
- Half given placebo, half given Omega3 fatty acid.
- Patients followed for 4 months.
- Variable of interest: Time to bipolar disorder relapse.

Example 2: Bipolar disorder



Example 2: Bipolar disorder

- **Two population comparison**
  - Let $X_1, \cdots, X_n$ be a random sample of censored survival data from population 1.
  - Let $Y_1, \cdots, Y_n$ be a random sample of censored survival data from population 2.
  - Are survival times from two populations different?

Example 3: Kidney transplant

- Goal: Investigate the effects of risk factors such as Gender, Race and Age on survival of Kidney transplant patients
- A total of 863 Kidney transplant patients were followed until death or censoring.
- All patients had their transplant performed at the Ohio State University Transplant Center during the period of 1982-1992.
- The maximum follow up time was 9.47 years.
- Patients were censored if they moved from Columbus or still alive at the end of the study.

Example 3: Kidney transplant

- Part of the data

Patient ID	Time(Days)	Status	Gender	Race	Age
1	3019	1	M	B	56
2	5	0	M	W	51
3	3281	0	F	W	47
4	198	1	F	W	42
5	17	1	M	W	47

Example 3: Kidney transplant

- Regression problem
- Assess the effect of risk factors to time-to-event.
- Suppose the data consists of $\{X_i, \boldsymbol{Z}_i, \mathrm{i} = 1, \cdots, n\}$ where $X_i{}'$s are censored survival data and $\boldsymbol{Z}_i{}'$s are covariates.
- What are the effects of covariates on the survival time?

Course Organization

- Introduction
- Kaplan-Meier Estimation
- Log-rank test
- Parametric regression
- Cox's proportional hazard model