STK4080 SURVIVAL AND EVENT HISTORY ANALYSIS

Slides 3: Censoring

Bo Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim

https://www.ntnu.no/ansatte/bo.lindqvist
bo.lindqvist@ntnu.no

University of Oslo, Autumn 2019

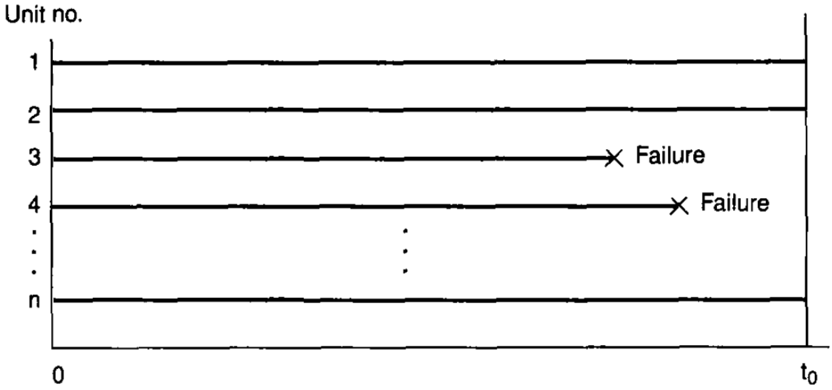
Lifetime data typically include *censored* data, meaning that:

- some lifetimes are known to have occurred only within certain intervals.
- The remaining lifetimes are known exactly.

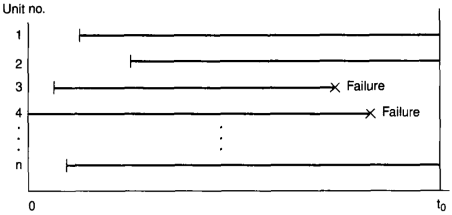
Categories of censoring:

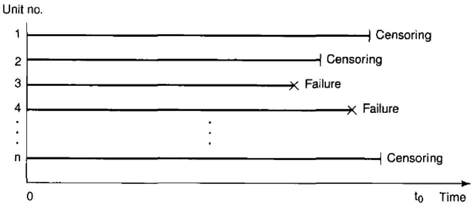
- right censoring
- left censoring
- interval censoring

n units put on test at time $t = 0$. Experiment stopped at time $t = t_0$.



Individuals enter the study at different times, and the terminal point of the study is predetermined.





n units are put on test at time $t = 0$.

The study continues until r individuals have failed, where r is some predetermined integer ($r < n$).

Advantage: It could take a very long time for all items to fail. Also, the statistical treatment of Type II censored data is simpler because the joint distribution of the order statistics is available.

This is a mix of Type I and Type II censoring. Choose both an end time t_0 as for Type I censoring and an $r < n$ as for Type II censoring. Stop the experiment at time t_0 or at the rth failure, whatever comes first.

- For each unit we define
  - T_i to be the potential lifetime
  - C_i to be the potential censoring time

  where
- T_i, C_i are **independent random variables**.
- Then we *observe* the pair $(\tilde{T}_i, D_i)$, where

$$\begin{aligned} \tilde{T}_i &= \min(T_i, C_i) \\ D_i &= \begin{cases} 1 & \text{if} \quad T_i \leq C_i \\ 0 & \text{if} \quad T_i > C_i \end{cases} \end{aligned}$$

Example of use: Cancer treatment, with T_i being the time of death due to this cancer; while C_i is the time of death of another cause, or an accident, or migration, etc.

(*Right censoring is the most common way of censoring.*)

Right censoring of Type I, II, III, IV can all be represented as follows:

n units are observed, with potential i.i.d. lifetimes $T_1, T_2, \cdots, T_n$. For each i, we observe a time $\tilde{T}_i$ which is either the true lifetime T_i, or a censoring time $C_i < T_i$, in which case the true lifetime is "to the right" of the observed time C_i.

The observation from a unit is the pair $(\tilde{T}_i, D_i)$ where the *censoring indicator* D_i is defined by

$$D_i = \begin{cases} 1 & \text{if} \quad \tilde{T}_i = T_i, \text{ in which case we observe the true lifetime } T_i \\ 0 & \text{if} \quad \tilde{T}_i = C_i, \text{ in which case it is only known that } T_i > Y_i \end{cases}$$

INDEPENDENT CENSORING

Consider a situation where n individuals are followed from time $t = 0$. The ith individual is followed until $\tilde{T}_i = \min(T_i, C_i)$, i.e. until either failure (death) or censoring at time C_i.

The ith individual is said to be at risk at time t if $t < \tilde{T}_i$, i.e. if the individual has not yet been censored and have not failed.

A sensoring scheme is said to satisfy the property of **independent censoring** if, at any time t, the individuals that are *at risk* are representative for the distribution of T in the sense that their probaility of failing in a small time interval $(t, t + h)$ is (in the limit as h tends to 0) is $\alpha(t)h$.

The censoring types we have considered so far all satisfy this independent censoring property.