Week 5

Wednesday, 4 August 2021 12:43

DATA, TIME & LOCATION

DOES EVENT x HAPPENS AFTER Δt?

BINARY — CONTINOUS

TIME-TO-EVENT DATA

INDIVIDUAL EVENT TIME UNKNOWN

EVENTS OCCURRING OVERTIME,
OFTEN WE DO NOT RECORD THE EVENTS,
INSTEAD WE GROUP THEM INTO TIME INTERVALS

EX. DATA ABOUT DEATH & DISEASE RATE
GROUPED BY LOCATION AND YEAR

UNKNOWN TIME PERIODS

INCIDENT RATE $= \hat{IR} = \dfrac{\#\text{ CASES PER YEAR}}{\#\text{ PERSON-TIME AT RISK}}$

→ 1 YEAR PER PERSON IN THE SAMPLE

ASSUMPTION: LIVED 1 YEAR IN PENSYLVANIA, OBSERVED FOR 1 YEAR

$\dfrac{\hat{p}}{\Delta t} = $ RATE

TECHNICALLY THIS IS NOT A PROPORTION BUT A RATE
LOOKING AT PROPORTIONS AFTER 1 YEAR OF FOLLOW-UP

INDIVIDUAL EVENT TIMES KNOWN

STUDIES FROM DEFINED STARTING POINT TO Δt

COMPLETE VS CENSORED OBSERVATION

↓

PARTIAL OBSERVATION
(WE JUST HAVE LOWER BOUND)

- TREATING DEATH BINARY (IGNORING TIME)
  % WHO DIED IN THE FOLLOW-UP YEAR

PROBLEM: NOT ALL SUBJECT HAVE THE SAME INFLUENCE
HERE WE ARE IGNORING TIME AT RISK

USELESS

- TREATING FOLLOW-UP TIME CONTINUOUS (ONLY TIME)

$$\bar{x} = \frac{x_1 + x_2 + \dots x_N}{N}$$

PROBLEM JUST AVG FOLLOW-UP TIME
NOT CAPTURING AVG TIME TO DEATH

- INCIDENT RATE

$$\boxed{\hat{IR} = \frac{\text{TOTAL \# OF DEATH}}{\text{TOTAL \# OF FOLLOW-UP TIME}}}$$

ASSUMING $\hat{IR}$ IS CONSTANT ACROSS FOLLOW-UP YEAR

COMPARING NUMERICALLY

HAZARD RATIO $\frac{\hat{IR}_X}{\hat{IR}_Y}$ $\hat{IRR}$

INCIDENT RATE RATIO (IMPORTANCE OF Δt)

$\hat{IR}_s$ ARE DIFFICULT TO INTERPRET

KAPLAN-MEIER PROCESS
ANALYZING SURVIVAL CURVES

SUMMARIZING TIME-TO-EVENT DATA

SURVIVAL CURVE $S(t)$

PROPORTION REMAINING EVENT-FREE (SURVIVING)
AT LEAST TO TIME t OR BEYOND

CURVE CAN ONLY DECREASE

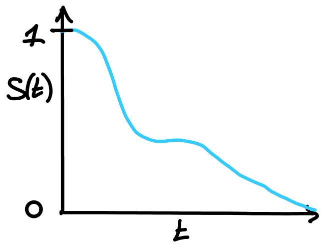
ESTIMATED BY DATA

$\hat{S}(t)$ ESTIMATED CURVE (KAPLAN MEIER)

KM - CURVE

SHOWS HOW DEATH IS COMMON
DURING FOLLOW-UP PERIOD

$\Uparrow$ IR AVERAGED ACROSS
FOLLOW-UP PERIOD



WE CAN USE BOTH
COMPLETE & INCOMPLETE DATA

EXAMPLE SMOKING CESSATION WORKSHOP

2 3+ 6 8 9+ 10
15+ 16 18 24+ 27 30

DATA
N = 12

"+" MEANS HE/SHE
DID NOT QUIT SMOKING

$$\boxed{S(t) = \frac{N(t) - E(t)}{N(t)} \cdot S(\text{previous event time})}$$

NOT HAD EVENT

$N(t)$ # STILL AT RISK OF HAVING EVENT AT TIME t

$E(t)$ # OF PEOPLE WHO HAD THE EVENT AT TIME t

$S(t)$ ESTIMATE PROPORTION OF PEOPLE WHO
DID NOT HAVE THE EVENT

$\hat{S}(0) = 1$ (100%)
CURVE WILL REMAIN 1 UNTILL FIRST EVENT

$$\hat{S}(2) = \frac{N(2) - E(2)}{N(2)} = \frac{12-1}{12} \cdot 1 = 0.92$$

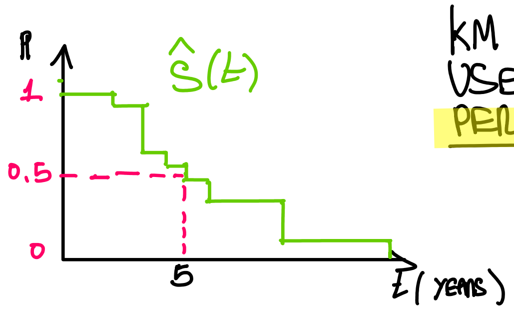
AT $t = 3$ ONE PERSON QUITS $N(3) = 11$

$$\hat{S}(6) = \frac{N(6) - E(6)}{N(6)} S(2) = \frac{10-1}{10} \cdot 0.92 = 0.83$$

$$\hat{S}(8) = \frac{N(8) - E(8)}{N(8)} S(6) = \frac{9-1}{9} \cdot 0.83 = 0.74$$

IF LITTLE DATA, KM IS LESS SMOOTH & FLUID

KM-CURVE DOES NOT ASSUME ANY STRUCTURE BETWEEN DATA POINTS, <u>NO INTERPOLATION</u>



KM CURVE IS USEFUL FOR **PERCENTILES**

$1 - \hat{S}(t)$ CURVE

PROPORTION OF SAMPLE WHO HAD THE EVENT CAN <u>ONLY INCREASE</u>

COMPLIMENT CURVE

MORE EASY TO WORK WITH PERCENTILES

ADDING RICHNESS TO TIME-TO-EVENT DATA

GRAPHICALLY COMPARING GROUPS ON TIME-TO-EVENT OUTCOMES

- COMPARING $\hat{IR}_1$, $\hat{IR}_2$ $\quad \hat{IRR} = \dfrac{\hat{IR}_1}{\hat{IR}_2}$
- K-M GRAPH (PLOTTING KM_1, KM_2 TOGETHER)