Kestusandmed, longituudsus: sündmusanalüüs

Ristlõikeandmed, kordusmõõtmised, kestusandmed

- Ristlõikeandmed – konkreetsel ajahetkel mõõdetud andmed
- Staatilised, ei võimalda üldjuhul ajakomponenti (sündmuste järgnevust) analüüsida
- Selle võimaldamiseks kordusmõõtmised, kaks varianti:
  - Mõõdetavad tunnused samad, indiviidid erinevad
  - Indiviidid samad, mõõdetavad tunnused põhiosas samad
- Teise puhul ongi tegu kestusandmetega
  - ehk longituudandmetega
  - ehk pikilõikeliste andmetega
- Kas Euroopa Sotsiaaluuring on longituuduuring?

(Tooding 2015)

Longituudandmed, kestusandmed, pikilõikelised andmed

- Longituuduuringud rikkaliku andmestikuga, aga
  - metodoloogiliselt väljakutseterohked
    - nii andmekogumise kui andmeanalüüsi metoodika poolest
  - korralduslikult keerukamad
  - kulukad

Sündmusanalüüs, elukestusanalüüs

Survival analysis, time-to-event analysis, failure analysis

- Kogum analüüsimeetodeid, mis võimaldavad analüüsida
  - mingi sündmuseni kuluva aja kestust (perioodi/episoodi pikkust)
  - ja seda kestust mõjutavaid tegureid
- Olulised mõisted:
  - Sündmus
  - Episood, sündmusele eelnev ajaperiood

Sündmusanalüüsi näiteid

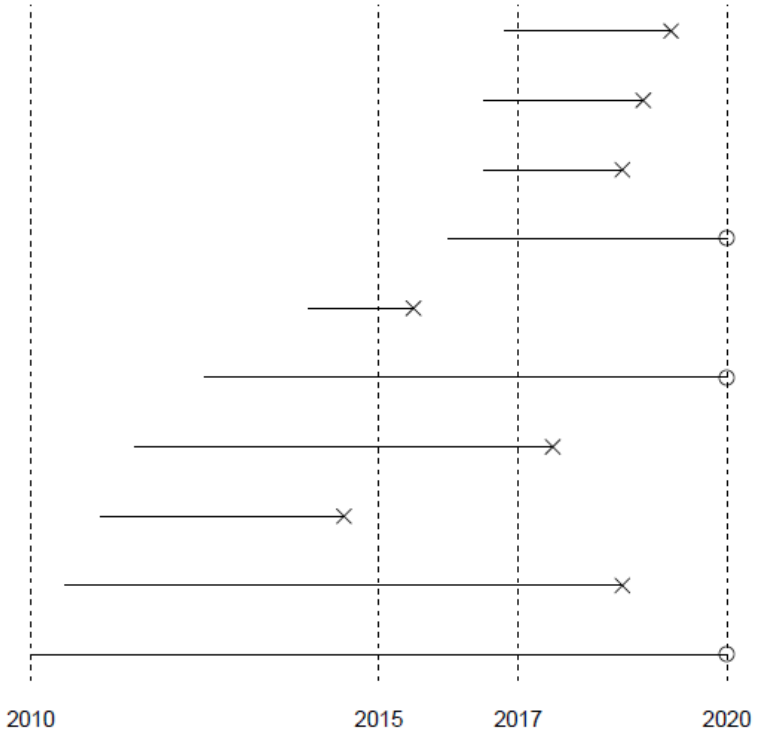
Uurimisvaldkond	Sündmus	Episood

- Võimalik uurida kestuse erinevust gruppides, tegureid, mis kestust ja sündmuseni jõudmist võivad mõjutada

Tsenseerimine

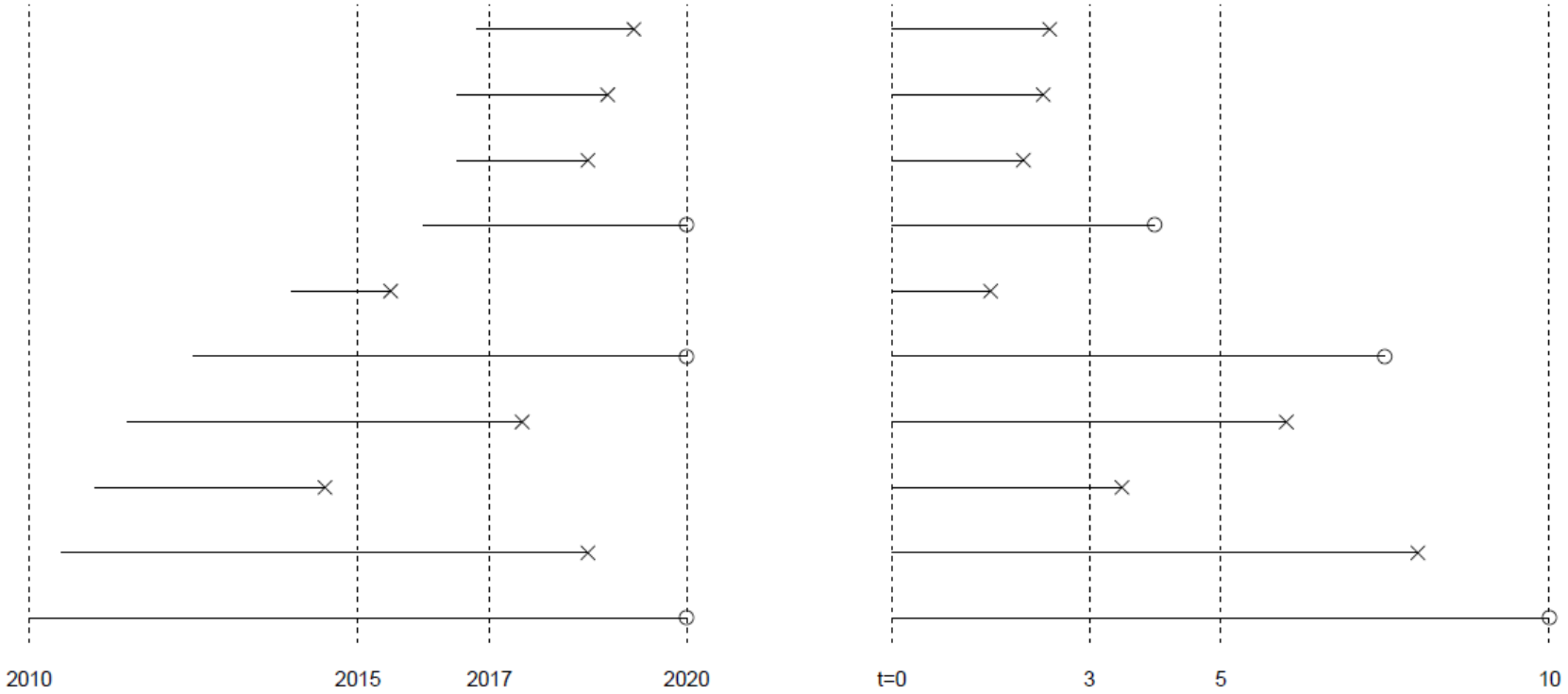
- Uuring võib lõppeda enne sündmuse toimumist
- Uuritav võib n-ö ära kaduda
- Sündmust ei pruugi toimuda, sest leiab aset muu sündmus (nt inimene ei leia tööd, sest kaotab õnnetuse tõttu töövõime)
- Nimetatakse tsenseerimiseks (konkreetse indiviidi andmed on tsenseeritud)
- Probleem, sest me ei tea, millal sündmus oleks toimunud

Tsenseerimine



(Stare 2020)

Tsenseerimine



(Stare 2020)

Elukestusfunktsioon, elulemusfunktsioon

Survival function

- Sündmuse mittetoimumise tõenäosus ajahetkeni t (t kaasa arvatud)
- Nt tõenäosus, et
  - vähipatsient jääb ajahetkeni t ellu
  - töötu ei ole t kuud pärast arvelevõtmist tööd leidnud
- Indiviidide osakaal, kellel ajahetkeni t EI OLE sündmus aset leidnud

$$S(t) = P(T > t)$$

- Kui tsenseeritud indiviide ei esineks, saaks elulemusfunktsiooni arvutada lihtsalt

$$\hat{S}(t) = \frac{Indiviidide\ arv, kellel\ T > t}{Kõigi\ indiviidide\ arv}$$

- Mida teha tsenseeritud indiviididega?
  - Kui viskame välja, võime üle hinnata sündmuste osakaalu (elulemus väiksem)
  - Kui võrdsustame tsenseerimise sündmuse mittetoimumisega, võime alahinnata sündmuste osakaalu (elulemus tegelikust suurem)

Riskifunktsioon

Hazard function

- Sündmuse toimumise tõenäosus ajahetkel *t* indiviidil, kellel selle ajahetkeni sündmust toimunud ei ole
  - Ehk sündmuse toimumise tõenäosus ajahetkel *t* nende hulgas, kes on selle hetkeni elus (kellel pole veel sündmust toimunud)
  - Ehk sündmuse toimumise tinglik tõenäosus vaadeldaval ajahetkel eeldusel, et sündmus ei ole toimunud enne ajahetke

$$\lambda_i = P(T = a_i | T \geq a_i)$$

$$\lambda(t) = \lim_{\Delta t \to 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

Kaplan-Meieri hinnangufunktsioon

Kaplan-Meier estimator, product-limit estimator

- Mitteparameetriline meetod elukestusfunktsiooni hindamiseks
  - st hindamiseks, kui suur on tõenäosus, et ajahetkeni t pole sündmust toimunud

$$S(t) = \prod_{i:\, t_i \le t} (1 - \frac{d_i}{n_i})$$

  - t_i = 1, 2, …, k – sündmuse toimumise momendid vaatlusperioodi vältel
  - d_i – ajavahemikus (t_{i-1}, t_i] toimuvate sündmuste arv
  - n_i – indiviidide arv, kellel enne ajahetke t_i ei ole sündmust toimunud ($T \geq t_i$)

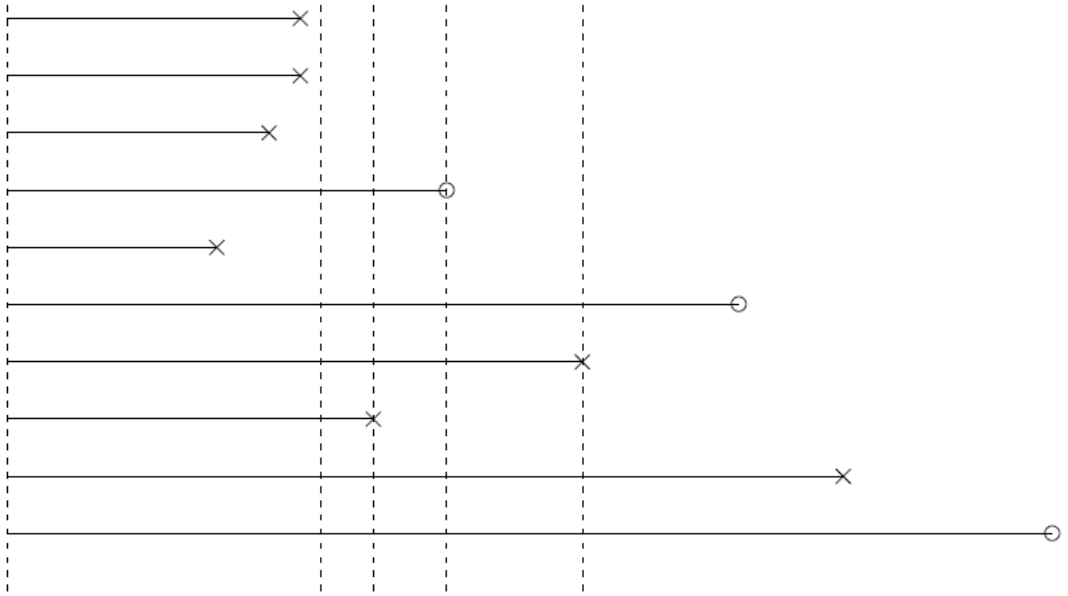
- Elukestusfunktsiooni hindamine konkreetse ajahetke kohta

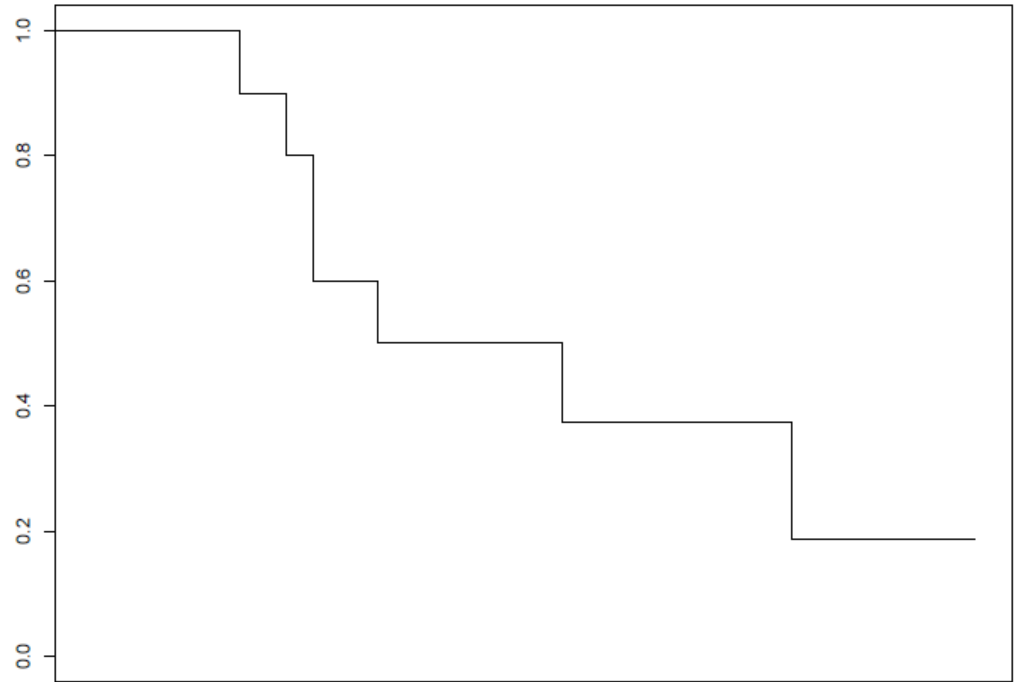
$$S(t_i) = S(t_{i-1}) \cdot (1 - \frac{d_i}{n_i})$$

- Saame esitada elukestusfunktsiooni väärtused võrdlevalt rühmiti

Kaplan-Meieri hinnangufunktsioon

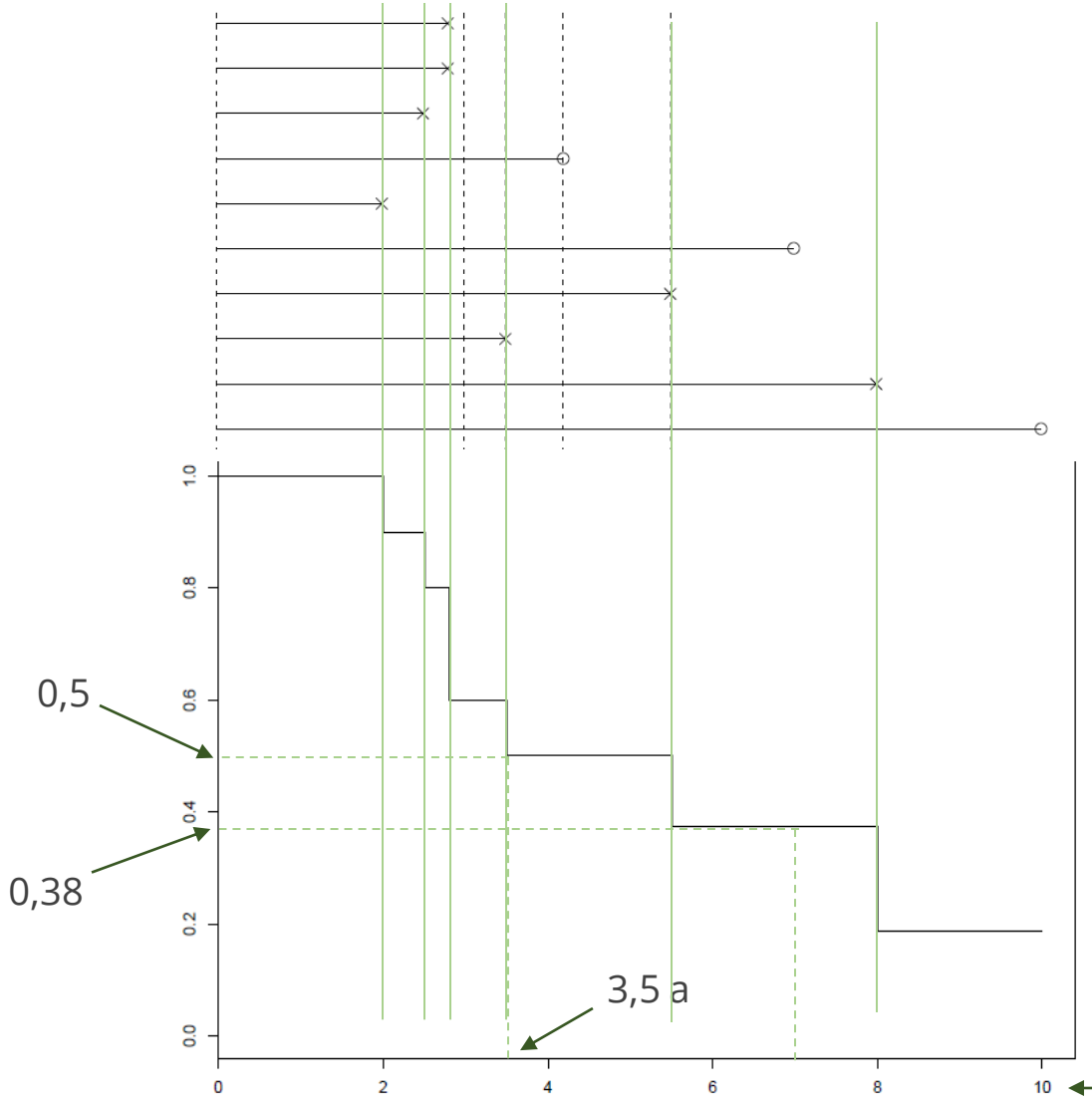
Kaplan-Meier estimator, product-limit estimator





(Stare 2020)

t_i	2,0	2,5	2,8	3,5	5,5	8,0
d_i	1	1	2	1	1	2
n_i	10	9	8	6	4	2
$1-d_i/n_i$	9/10	8/9	6/8	5/6	3/4	1/2
$1-d_i/n_i$	0,9	0,89	0,75	0,83	0,75	0,5
$S(t_i)$	0,9	0,8	0,6	0,5	0,38	0,19



- Tavaliselt kestuste jaotus asümmeetriline, seetõttu keskmise elukestuse / episoodi pikkuse asemel eelistatakse elukestuse mediaani (episoodide mediaan)

- Milline on tõenäosus, et seitsme aastaga ei ole sündmust toimunud?
- Nt töötute registris pikaajalise töötuse episoodid, sündmus on töö leidmine => milline on tõenäosus, et seitse aastat pärast arvelevõtmist ei ole pikaajaline töötu tööd leidnud?

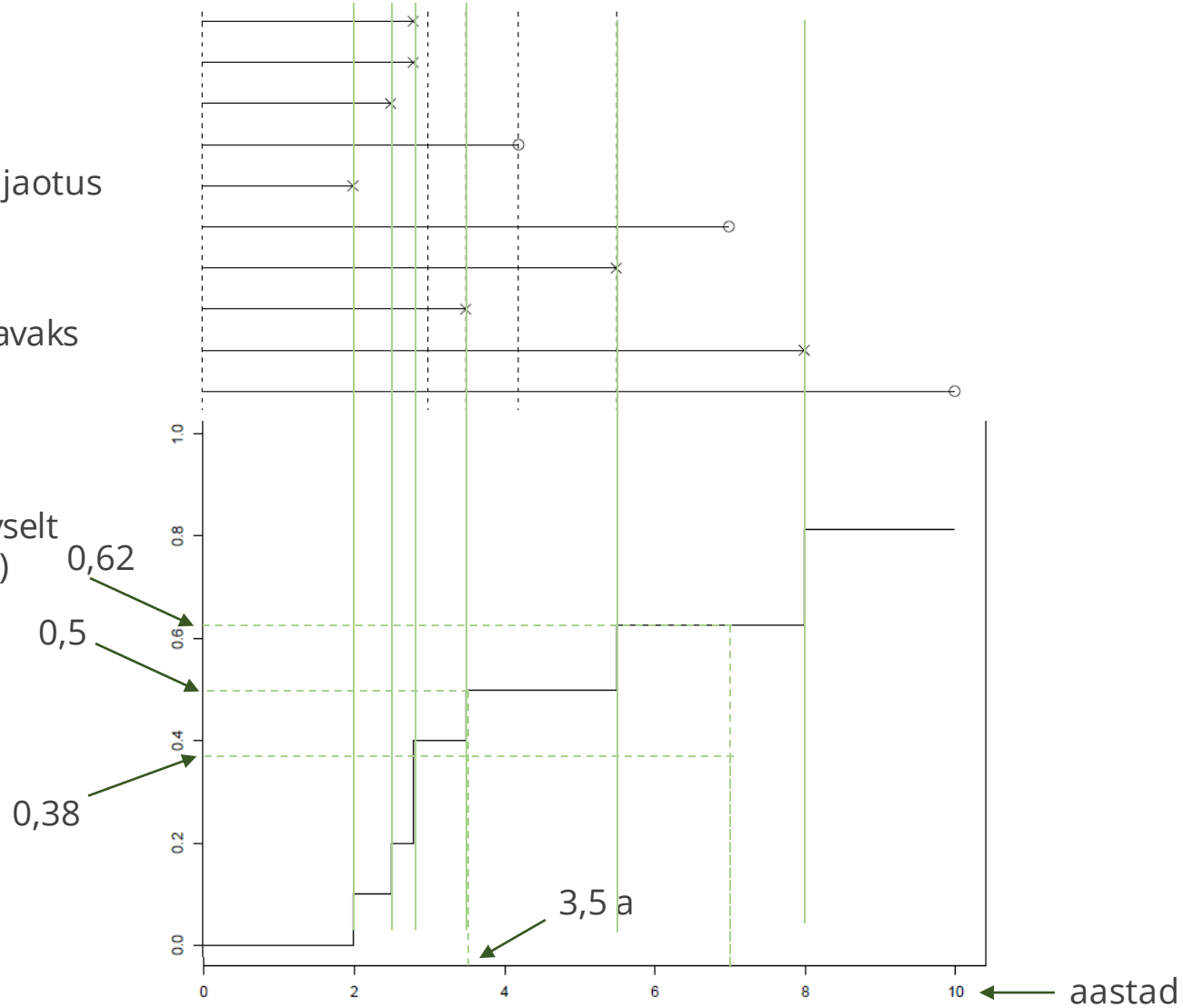
(Stare 2020)

t_i	2,0	2,5	2,8	3,5	5,5	8,0
d_i	1	1	2	1	1	2
n_i	10	9	8	6	4	2
$1-d_i/n_i$	9/10	8/9	6/8	5/6	3/4	1/2
$1-d_i/n_i$	0,9	0,89	0,75	0,83	0,75	0,5
$S(t_i)$	0,9	0,8	0,6	0,5	0,38	0,19

Jaotusfunktsioon ehk sündmuste kumulatiivne jaotus

$$F(t) = 1 - S(t)$$

- Näitab tõenäosust, et sündmus on vaadeldavaks ajahetkeks toimunud
- Elukestusfunktsiooni peegelpilt
- Vahel kergem intuitiivselt mõista (nt praktikutel)



(Stare 2020)

Kaplan-Meieri hinnangufunktsioon

Kaplan-Meier estimator, product-limit estimator

- Elukestuskõveraid saab arvutada eri rühmadele ja võrrelda,
  - kuidas sündmuse toimumise tõenäosus muutub ajas rühmiti
  - kas rühmad erinevad sündmuse toimumise tõenäosuse poolest
- Võimalik arvutada kõveratele usaldusvahemikud ja võrrelda ka nende alusel
- Episoodide jaotus asümmeetriline
  - tsenseeritud juhtumitel pigem pikem episood =>
  - usaldusvahemikud elukestuse suuremate väärtuste puhul laiemad
- Elukestuse erinevuste hindamiseks rühmades võimalik
  - võrrelda usaldusvahemikega K-M kõveraid visuaalselt
  - arvutada erinevust/seost mõõtev teststatistik
  - kasutada keerulisemaid (parameetrilisi ja semiparameetrilisi) meetodeid

(Stare 2020)

Logaritmiline astaktest

Log-rank test, Mantel-Haenszel test

- Mitteparameetriline test elukestusfunktsioonide võrdlemiseks kahes või enamas rühmas
  - H0: sündmuse toimumise hetkel on sündmuse kumulatiivsed tõenäosused rühmades võrdsed
  - H1: sündmuse kumulatiivsed tõenäosused rühmades erinevad
- Testi loogika sarnane hii-ruut-testile
  - Iga ajahetke kohta, mil mõnel indiviidil toimub sündmus, arvutatakse teoreetiline sündmuste arv rühmades, mis peaks esinema eeldusel, et rühmade elukestustes erinevused puuduvad
  - Iga ajahetke kohta arvutatakse tegeliku ja teoreetilise sündmuste arvu vahe, võetakse ruutu, jagatakse teoreetilise sündmuste arvuga ja summeeritakse
  - Võrreldakse hii-ruut-jaotusega => saadakse teststatistiku olulisuse tõenäosus

$$\chi^2 = \sum_{i=1}^{g} \frac{(O_i - E_i)^2}{E_i}$$

  - g – rühmade arv
  - O_i – tegelik sündmuste arv hetkel i
  - E_i – teoreetiline sündmuste arv hetkel i
- K-M elukestuskõverad ei tohiks lõikuda
  - Kui lõikuvad, on antud testi suutlikkus tuvastada rühmadevahelisi erinevusi elukestuses väga madal
- Kasutatakse väga laialdaselt

(Clark et al 2003; Tiit ja Tooding 2019: 161)