

Biostatistika

Märt Möls
martm@ut.ee

Haiguste esinemissageduse mõõtmine ja standardimine

Biostatistika 1. loeng

Mis on biostatistika?

Statistiliste meetodite kasutamine meditsiini või bioloogia jaoks aktuaalsete probleemide lahendamiseks.

Miks on biostatistika eraldi õppeaine (miks lihtsalt statistika loengutest või teadmistest ei piisa)?

- erinevused terminoloogias (esitatud küsimus tuleb statistika-keelde tõlkida);
- andmete/uurimistöö tausta mõistmine vajalik et valida sobivaid eelduseid ja meetodeid;

Levimus (*prevalence*)

P(juhuslikult valitud inimene populatsioonist on haige) =

$$\frac{\text{haigete arv populatsioonis}}{\text{populatsiooni suurus}}$$

Levimuse hindamisel, usalduspiiride leidmisel või hüpoteeside testimisel (kas levimus on kahes riigis või elanikkonna grupis samasuur?) tuleb meele pidada, et tegemist on tõenäosusega (ehk keskväärtusega). Kuna kasutatav valim (või valimid) on enamasti suured, saab kasutada ka asümptotilisi meetodeid.

Usaldusintervalli leidmine

Valimi suurus: n
haigete osakaal valimis: PR

(valim enamasti suur)

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Täpsed statistilised meetodid – usaldusintervall

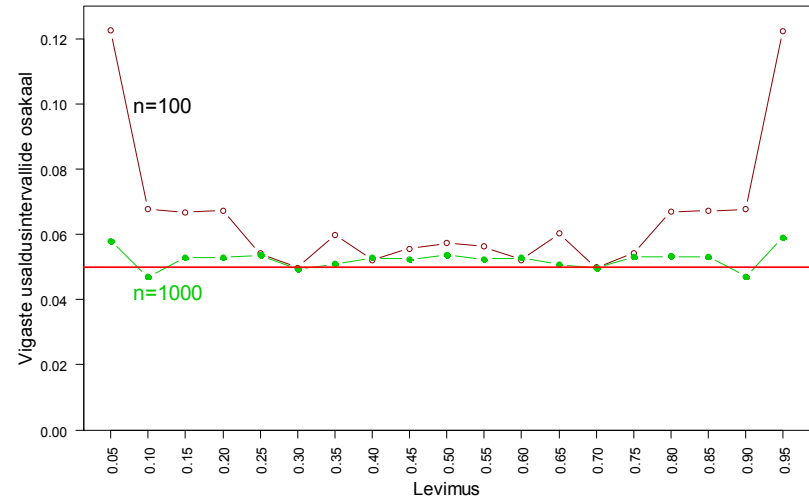
Ligikaudne (asümptootiline)

$$PR \pm z_{\alpha/2} \sqrt{\frac{PR(1-PR)}{n}}$$

Täpne usaldusintervall
tõenäosusuele (R)

`binom.test(25, 100)`
 ↑ ↑
 haigete arv uuritavaid (n)

Ligikaudne 95%-usaldusintervall



Kuna kasutada usaldusintervalle?

Hüpoteetiline näide:

Esti elanike arv 1. jaanuaril 2020: 1 328 360
 haigeid Eestis (kõik haiged): 100

Kas haiguse levimus on 7.52808 juhtu 100 000 inimese kohta või 7,5 (6.1... 9.2) juhtu 100 000 inimese kohta?

Esti elanike arv 2. jaanuaril 2020: 1 328 361
 haigeid Eestis (kõik haiged): 100

Haiguse levimus
 7.528074 juhtu 100 000 inimese kohta:
**MEDITSIINISÜSTEEM/KESKKOND ON
 EESTIS TÕESTATAVALT PAREMAKS
 MUUTUNUD!**

Haiguse levimus
 7.5 (95%-UI: 6,1...9,2) juhtu 100 000
 inimese kohta:
 Tõestatavaid muutuseid pole!

Täpsed statistilised meetodid – kahe populatsiooni võrdlus

Ligikaudne (asümptootiline)

```
haige1 = c(1,0,0,1,0,...
haige2 = c(0,1,0,1,1,...
t.test(haige1, haige2)
```

Fisher'i täpne test

```
fisher.test(
  cbind( c(25, 100),
         c(40, 100) ) )
```

↑ ↑
 haigete arv tervete arv,
 1. valim. 1. valim

↑ ↑
 haigete arv tervete arv,
 2. valim. 2. valim

Samuti kasutatakse:
 logistiline regressioon
 (võimaldab keerukamaid mudeleid)

Täpsed statistilised meetodid – kahe populatsiooni võrdlus

Logistiline regressioon
(ligikaudne; asümptootiline)

```
haige = c(1,0,0,1,0,...
popul = c(1,1,2,2,2,...
model=glm(haige~factor(popul),
           family=binomial())
drop1(model, test="Chisq")
```

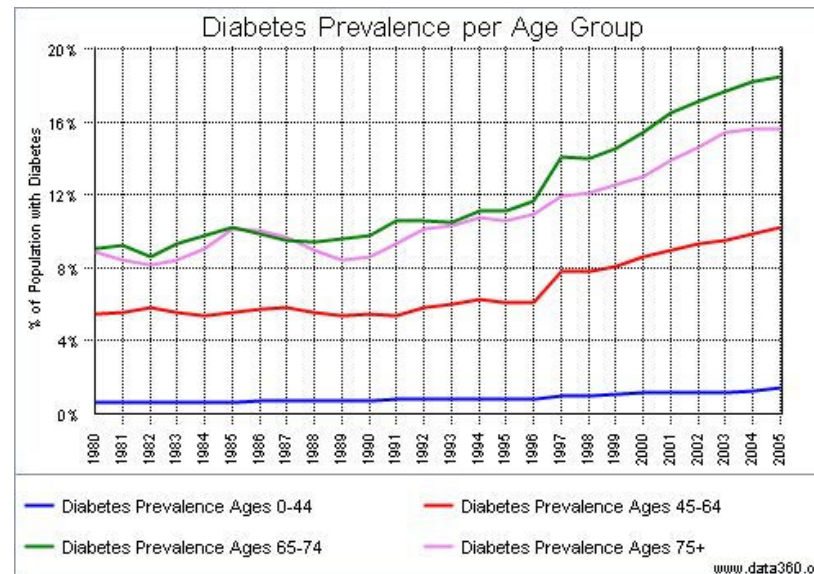
Fisheri täpne test

```
fisher.test(
  cbind( c(25, 100),
         c(40, 100)  ))
```

haigete arv
1. valim. tervete arv,
1. valim

haigete arv
2. valim. tervete arv,
2. valim

Samuti kasutatakse:
logistiline regressioon
(võimaldab keerukamaid
mudeleid)



Levimuse mõõtmise erivormid:

- hetkelevimus (point prevalence)
 - kui suur osa elanikkonnast on haige mingil ajahetkel
- perioodilevimus (period prevalence)
 - tõenäosus põdeda mingit haigust mingi ajaperioodi jooksul (kui suur osa elanikkonnast kannatab külmetushaiguste all aasta jooksul)
- Lifetime prevalence
 - milline on tõenäosus inimesel oma elu jooksul antud haigusega kokku puutuda

Esmahaigestumus (Incidence)

Vahel ka:
avaldumus

Variant 1 - Incidence Cases; Incidence
uute haigusjuhtude ehk haiguse esmasjuhtude arv mingil ajavahemikul

Näiteks Eestis (2016) – uued haigusjuhud:

Infarkte:	2 826
Suhkrutõbi	7 385
Parkinson	825
Kopsupõletik	12 830

Raske (mõttetu) võrrelda:

USA uusi vähijuhte 1975: ~ 650 000
USA uusi vähijuhte 2009: ~ 852 000

Aga elanike arv 1975: 216 mil.
2009: 307 mil.

Esmahaigestumus (Incidence)

Variant 2 – Haigestumusrisk

Incidence Proportion (IP); cumulative incidence; Incidence Risk
haigestumuskindaja, mis väljendab haigusjuhtude arvu rahvaarvu suhtes mingil ajavahemiku

$$IP = \frac{\text{uute haigusjuhtude arv ajavahemikus}}{\text{inimeste arv ajavahemiku alguses}}$$

Esitatakse sageli kui juhte 100 000 või 1000 või ...
inimese kohta aastas

Näiteks Eestis (2016) – uued haigusjuhud (100 000 elaniku kohta):

Infarkte:	208
Suhkrutõbi	561
Parkinson	63
Kopsupõletik	975

Esmahaigestumus (Incidence)

Variant 3 – Incidence Rate (IR)

haigestumuskindaja, mis väljendab haigusjuhtude arvu rahvaarvu suhtes mingil ajavahemiku

$$IR = \frac{\text{uute haigusjuhtude arv}}{\text{riskiaeg}}$$

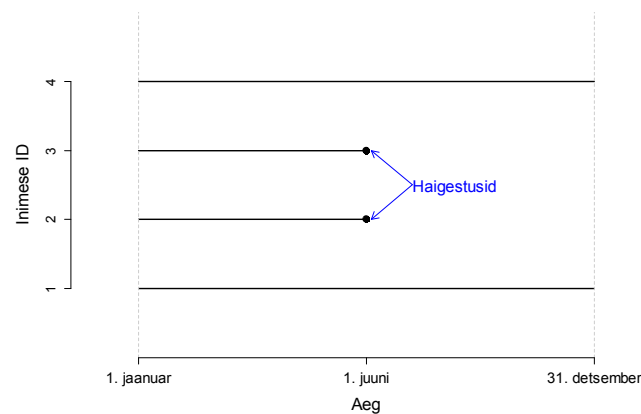
Riskiaeg (*risk time, time at risk*) – haigestuda võivate (nn *vastuvõtlike*) inimeste poolt kokku vaadeldud ajavahemikul elatud aeg. Kasutatakse ka mõisteid *person-years* või *person-time*.

Esmahaigestumus (Incidence)

Näide – IR ja IP võrdlus

$$IP = 2 / 4 = 0,5$$

$$IR = 2 / 3 = 0,666\dots$$



Esmahaigestumus (Incidence)

X - aeg haigestumiseni

$$X \sim \text{Exp}(\lambda)$$

Incidence rate

$$F_X(x) = 1 - \exp(-\lambda \cdot x)$$

Relation between CI and IR

$CI(t) = 1 - e^{-IR \times t}$, where e is the base to the natural logarithm (2.718) and t is the time unit of concern. When the CI < 0.10 (10%), the formula approximate to:

$$\exp(\epsilon) \approx 1 + \epsilon$$

$$CI(t) = IR \times t$$

Usaldusintervallide leidmisel alustame tähelepanekust, et uute haigusjuhtude arv (I) on enamasti (vähemalt ligikaudu) Poissoni jaotusega juhuslik suurus:

$$I \sim \text{Poi}(IR * \text{riskiaeg});$$

Poissoni jaotuse korral on aga juhusliku suuruse dispersioon võrdne keskvaartusega ehk Poissoni jaotuse parameetriga:

$$D(I) = IR * \text{riskiaeg}.$$

Seega

$$D(\widehat{IR}) = D(I/\text{riskiaeg}) = D(I)/\text{riskiaeg}^2 = IR/\text{riskiaeg} = IR^2/I$$

ja ligikaudne 95%-usaldusintervall tegelikule esmahaigestumusele on leitav valemiga

$$\widehat{IR} \pm 1,96 \frac{\widehat{IR}}{\sqrt{I}}$$

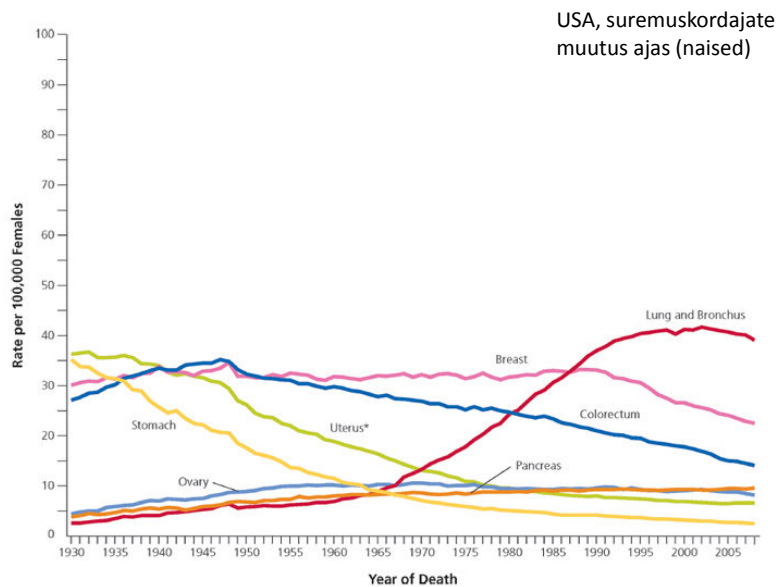
Analoogsed näitajaid on veel, näiteks **suremuskordaja** (*mortality rate / death rate*):

$$MR = \text{surmade arv mingil ajavahemikul} / \text{riskiaeg}$$

Ligikaudne seos (antud lühikese ja intensiivse haiguse) suremuskordaja ja haigestumuskordaja vahel:

$$MR \approx IR * P(\text{haigus lõpeb surmaga}).$$

NB! Eelmisel slaidil toodud valem usaldusintervalli leidmiseks esmahaigestumuskordajale IR-le ja suremuskordajale MR—le ei sobi kasutamiseks nakkushaiguste korral. Kas oskad põhjendada, miks?



Haigestumus(suremus)kordaja oma toorel kujul võib kergesti inimesi eksiteele viia. Näiteks oli suremuskordaja Inglismaal 1901. aastal 15,7 (1000 inimese kohta aastas), 1981. aastal aga 15,6 (1000 inimese kohta aastas). Kas see tähendab, et asjalood olid 1981.a. sama halvad kui 1901. aastal?

Vaatame suremust vanusegrupiti:

Vanusegrupp	suremuskordaja (MR)		vanusegrupi osakaal %	
	1901	1981	1901	1981
15-19	3,5	0,8	15,36	11,09
20-24	4,7	0,8	14,07	9,79
25-34	6,2	0,9	23,76	18,81
35-44	10,6	1,8	18,46	15,99
45-54	18,0	6,1	13,34	14,75
55-64	33,5	17,7	8,68	14,04
65-74	67,8	45,6	4,57	10,65
75-84	139,8	105,2	1,58	4,28
84+	276,5	226,2	0,17	0,64

Suremus(haigestumus)kordaja standardimine I

Leiame, milline oleks olnud 1901. aasta suremuskordaja, kui ühiskonna vanuseline struktuur oleks olnud samasugune kui 1981.a.

Vanusegrupp	vanuserühma suremuskordaja aastal 1901	vanuserühma osakaal 1981.a.	
15-19	3,5	0,1109	$3,5 * 0,1109 = 0,38815$
20-24	4,7	0,0979	0,46013
25-34	6,2	0,1881	1,16622
35-44	10,6	0,1599	1,69494
45-54	18,0	0,1475	2,65500
55-64	33,5	0,1404	4,70340
65-74	67,8	0,1065	7,22070
75-84	139,8	0,0428	5,98344
84+	276,5	0,0064	1,76960
MR			26,04

Ehk kui peaksime praeguse elanikkonnaga hakkama saama 1901.a meditsiini ja elutingimustega, sureks aastas 1000 inimese kohta 10 inimest rohkem kui praegu.

Suremus(haigestumus)kordaja standardimine II

Kahjuks tuleb sageli ette olukordi, kus pole selge, millise riigi või rahvastikugrupi suremusega millisel aastal lugejad uuringu tulemust võrdlema soovib hakata. Lahenduseks on nn. standardrahvastiku kasutamine: tulemuste võrreldavaks tegemiseks kasutatakse etteantud teoreetilist vanusestruktuuri.

Vanuserühm	Maailm (Segi, 1960)	Euroopa	WHO World (2000-2025)
0-4	12000	8000	88 569
5-9	10000	7000	86 870
10-14	9000	7000	85 970
15-19	9000	7000	84 670
20-24	8000	7000	82 171
25-29	8000	7000	79 272
30-34	6000	7000	76 073
35-39	6000	7000	71 475
40-44	6000	7000	65 877
45-49	6000	7000	60 379
50-54	5000	7000	53 681
55-59	4000	6000	45 484
60-64	4000	5000	37 187
65-69	3000	4000	29 590
70-74	2000	3000	22 092
75-79	1000	2000	15 159
80-84	500	1000	9 097
85-	500	1000	6 348
Kokku	100 000	100 000	1 000 000

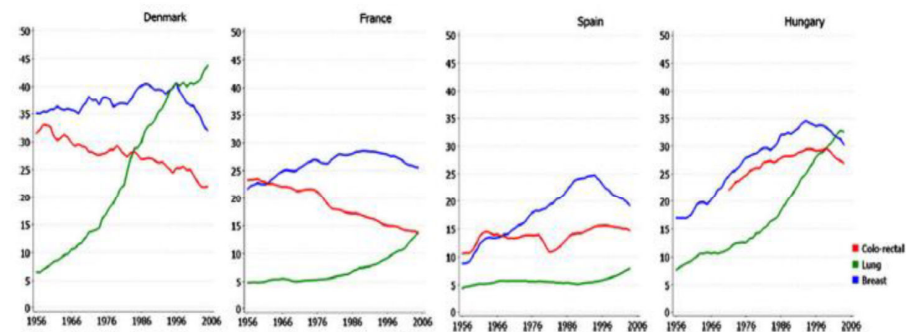


Fig. 4b - Trends in age-standardised (European standard) mortality rates from colorectal, lung and breast cancers in females in Denmark, France, Spain and Hungary (Source: WHO mortality database⁹).

Standardiseeritud suremuskordaja täpsus

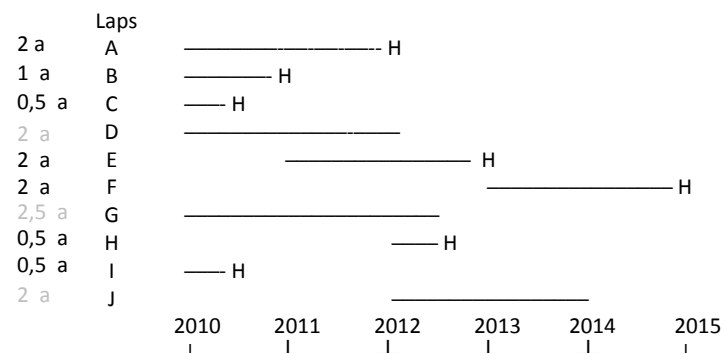
Eri vanuserühmadele leitud suremuskordaja hinnangud on sõltumatud. Seega:

$$D(0,25 MR_{\text{noored}} + 0,55 MR_{\text{keskealised}} + 0,2 MR_{\text{vanad}}) = 0,25^2 D(MR_{\text{noored}}) + 0,55^2 D(MR_{\text{keskealised}}) + 0,2^2 D(MR_{\text{vanad}}),$$

kus $D(MR_{\text{noored}})$ on noorte suremuskordaja dispersioon (arvutatud samamoodi kui esmahaigestumuskordaja dispersioon). Kui valimimahud on (enamikes vanuserühmades) piisavalt suured võib standardiseeritud haigestumuskordaja hinnangut vaadelda kui normaaljaotusega juhuslikku suurust.

Inimaastate (riskiaja) arvutamine II

Alustame lapse jälgimist sünnihetkest...



Kas haigestumus on vanuses 0..1 aastat suurem kui vanuses peale 1. sünnipäeva?

$$IR_{0..1} = 4/8,5 = 0,47 \text{ juhtu inimaastas}$$

$$IR_{>1} = 3/6,5 = 0,46 \text{ juhtu inimaastas}$$