

Seose tugevuse iseloomustamine I

2x2 sagedustabelid

Märt Möls
martm@ut.ee

1

Kas oliiviõli?

Uuringusse värvati terveid ameerika meditsiinitöötajaid (õdesid/arste). Küsiti, kas nad kasutavad või ei kasuta (alla 1 korra kuus) toidu tegemisel oliiviõli. Seejärel vaadati, kas neil ilmneb aasta jooksul mõni südamehaigus...

	süda haige	süda terve	Kokku
Muud õlid	1 971	397 715	399 686
Oliiviõli	3 516	892 123	892 123
Kokku	5 487	1 289 838	1 295 325

Kas seos eksisteerib?

> fisher.test(tabel)
p-value = 9.499e-16

Miks kasutasin Fisheri täpset testi,
mitte Barnardi täpset testi või hii-
ruut testi või tõepärasuhte testi?

Vastus: Barnardi täpse testi tegemine
võtab sedavõrd suure valimi korral
lihtsalt liiga palju aega....

2

Kas oliiviõli?

Uuringusse värvati terveid ameerika meditsiinitöötajaid (õdesid/arste). Küsiti, kas nad kasutavad või ei kasuta (alla 1 korra kuus) toidu tegemisel oliiviõli. Seejärel vaadati, kas neil ilmneb aasta jooksul mõni südamehaigus...

	süda haige	süda terve	Kokku
Muud õlid	1 971	397 715	399 686
Oliiviõli	3 516	892 123	892 123
Kokku	5 487	1 289 838	1 295 325

Seos on oluline – oliiviõli kasutatavad ameeriklased
jäävad harvem südamehaigustesse võrreldes teistega

Kas seos on ka meditsiini jaoks oluline? Kas haigusrisiki
muutus on piisavalt suur selleks, et seda tasuks edasi
uurida (sellest tasuks rääkida?)

NB! Antud uuringu põhjal ei saa veel väita, et
oliiviõli kasutamine (teiste õlide asemel) viib
väiksema südamehaiguste riskini – sest oliiviõli
tarbijad võivad olla näiteks terviseteadlikumad
inimesed ja käituda mingil muul moel
tervislikumalt. Põhjusliku mõju näitamiseks
peaksime tegema randomiseeritud katse...

3

Kas oliiviõli?

Uuringusse värvati terveid ameerika meditsiinitöötajaid (õdesid/arste). Küsiti, kas nad kasutavad või ei kasuta (alla 1 korra kuus) toidu tegemisel oliiviõli. Seejärel vaadati, kas neil ilmneb aasta jooksul mõni südamehaigus...

	süda haige	süda terve	Kokku
Muud õlid	1 971	397 715	399 686
Oliiviõli	3 516	892 123	895 639
Kokku	5 487	1 289 838	1 295 325

$$p_C = P(H + | C) \quad \hat{p}_C = \frac{1\,971}{399\,686} = 0,0049 \dots$$

$$p_T = P(H + | T) \quad \hat{p}_T = \frac{3\,516}{895\,639} = 0,0039 \dots$$

4

Ligikaudsed usalduspiirid tõenäosusele

$$\hat{p}_T = \frac{3516}{895639} = 0,0039 \dots$$

Kui täpne on hinnang?

$$Y = \begin{cases} 1 & \text{tõenäosusega } p \\ 0 & \text{tõenäosusega } 1-p \end{cases}$$

$$E(Y) = 1 \cdot p + 0 \cdot (1-p) = p$$

$$\begin{aligned} D(Y) &= (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) \\ &= (1-p)p \cdot ((1-p) + p) = p(1-p) \end{aligned}$$

$$D(\bar{Y}) = \frac{D(Y)}{n}$$

$$\hat{p} = \bar{Y} \quad D(\hat{p}) = \frac{p(1-p)}{n}$$

Tsentraalne piirteoreem
 $\hat{p} \sim N(p; p(1-p)/n)$

5

Ligikaudsed usalduspiirid tõenäosusele

$$\hat{p}_T = \frac{3516}{895639} = 0,0039 \dots$$

Kui täpne on hinnang?

Tsentraalne piirteoreem

$$\hat{p} \sim N(p; p(1-p)/n)$$

$$P\left(p - 1,96\sqrt{\frac{p(1-p)}{n}} \leq \hat{p} \leq p + 1,96\sqrt{\frac{p(1-p)}{n}}\right) = 0,95 \quad | -p - \hat{p}$$

$$P\left(-\hat{p} - 1,96\sqrt{\frac{p(1-p)}{n}} \leq -p \leq -\hat{p} + 1,96\sqrt{\frac{p(1-p)}{n}}\right) = 0,95 \quad | \cdot (-1)$$

$$P\left(\hat{p} + 1,96\sqrt{\frac{p(1-p)}{n}} \geq p \geq \hat{p} - 1,96\sqrt{\frac{p(1-p)}{n}}\right) = 0,95$$

Ligikaudne 95% usaldusintervall tõenäosusele: $\hat{p} \pm 1,96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

6

Ligikaudsed usalduspiirid tõenäosusele

$$\hat{p}_T = \frac{3516}{895639} = 0,0039 \dots$$

Kui täpne on hinnang?

```
binom.test(3516, 895639)
95 percent confidence interval:
 0.003797232 0.004057370
```

Täpsed usalduspiirid
tõenäosusele

Ligikaudne 95%-usaldusintervall p_T -le
0,0038 ... 0,0041

Ligikaudne 95% usaldusintervall tõenäosusele: $\hat{p} \pm 1,96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

7

Kui kasulik võiks olla oliiviõli?

$$\hat{p}_T = 0,0039 \dots$$

$$\hat{p}_C = 0,0049 \dots$$

Riskide vahe

*Risk Difference (RD);
Attributable Risk Reduction (ARR);
Attributable Risk Increase (ARI); ...*

$$RD = p_C - p_T$$

$$\widehat{RD} = \hat{p}_C - \hat{p}_T = 0,0010 \dots$$

Suhteline Risk

Risk Ratio (RR)

$$RR = p_T / p_C$$

$$\widehat{RR} = \frac{0,0039 \dots}{0,0049 \dots} = 0,796 \dots$$

Šansside suhe

Odds Ratio (OR)

$$OR = \frac{p_T / (1 - p_T)}{p_C / (1 - p_C)} = \frac{p_T(1 - p_C)}{p_C(1 - p_T)}$$

$$\widehat{OR} = 0,803 \dots$$

Väikeste tõenäosuste korral
on OR ja RR sarnased

Muud

*Number Needed to Treat (NNT)
Relative Risk Reduction (RRR)*

$$NNT = 1 / (p_C - p_T)$$

$$RRR = (p_C - p_T) / p_C$$

$$\widehat{NNT} \approx 1000$$

$$\widehat{RRR} = 0,204 \dots$$

...

9

Usaldusintervallid

$$ARR = p_C - p_T$$
$$\widehat{ARR} = \hat{p}_C - \hat{p}_T$$

Suure valimi korral:

$$\hat{p} \sim N(p; p(1-p)/n)$$

$$\hat{p}_C \sim N(p_C; p_C(1-p_C)/n_C) \quad \hat{p}_T \sim N(p_T; p_T(1-p_T)/n_T)$$

$$\widehat{ARR} = \hat{p}_C - \hat{p}_T \sim N(p_C - p_T; p_C(1-p_C)/n_C + p_T(1-p_T)/n_T)$$
$$\sim N(ARR; D(\widehat{ARR}))$$

$$P\left(ARR - 1,96\sqrt{D(\widehat{ARR})} \leq \widehat{ARR} \leq ARR + 1,96\sqrt{D(\widehat{ARR})}\right) = 0,95$$

$$P\left(-1,96\sqrt{D(\widehat{ARR})} \leq \widehat{ARR} - ARR \leq +1,96\sqrt{D(\widehat{ARR})}\right) = 0,95$$

10

Usaldusintervallid

$$ARR = p_C - p_T$$
$$\widehat{ARR} = \hat{p}_C - \hat{p}_T$$

$$P\left(ARR - 1,96\sqrt{D(\widehat{ARR})} \leq \widehat{ARR} \leq ARR + 1,96\sqrt{D(\widehat{ARR})}\right) = 0,95$$

$$P\left(-1,96\sqrt{D(\widehat{ARR})} \leq \widehat{ARR} - ARR \leq +1,96\sqrt{D(\widehat{ARR})}\right) = 0,95$$

$$P\left(-\widehat{ARR} - 1,96\sqrt{D(\widehat{ARR})} \leq -ARR \leq -\widehat{ARR} + 1,96\sqrt{D(\widehat{ARR})}\right) = 0,95$$

$$P\left(\widehat{ARR} + 1,96\sqrt{D(\widehat{ARR})} \geq ARR \geq \widehat{ARR} - 1,96\sqrt{D(\widehat{ARR})}\right) = 0,95$$

Ligikaudne 95%-usaldusintervall *ARR*-le:

$$\widehat{ARR} \pm 1,96\sqrt{\hat{D}(\widehat{ARR})}$$

12

Usaldusintervallid (ligikaudsed)

ARR (riskide vahe)

$$\hat{p}_C - \hat{p}_T \pm 1,96\sqrt{\hat{p}_C(1-\hat{p}_C)/n_C + \hat{p}_T(1-\hat{p}_T)/n_T}$$

$\frac{a}{a+b}$ $\frac{c}{c+d}$ $a+b$ $c+d$

	haige	terve
ravi	<i>a</i>	<i>b</i>
kontroll	<i>c</i>	<i>d</i>

Ligikaudne 95%-usaldusintervall *ARR*-le:

$$\widehat{ARR} \pm 1,96\sqrt{\hat{D}(\widehat{ARR})}$$

$\hat{p}_T(1-\hat{p}_T)/n_T + \hat{p}_T(1-\hat{p}_T)/n_T$

14

Usaldusintervallid (ligikaudsed)

ARR (riskide vahe)

$$\hat{p}_C - \hat{p}_T \pm 1,96\sqrt{\hat{p}_C(1-\hat{p}_C)/n_C + \hat{p}_T(1-\hat{p}_T)/n_T}$$

	haige	terve
ravi	<i>a</i>	<i>b</i>
kontroll	<i>c</i>	<i>d</i>

RR (suhteline risk)

$$\widehat{RR} \cdot \exp\left(\pm 1,96\sqrt{\frac{b}{a(a+b)} + \frac{d}{c(c+d)}}\right)$$

OR (šansside suhe)

$$\widehat{OR} \cdot \exp\left(\pm 1,96\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\right)$$

Mida on vaja, et
usaldusintervall šansside
suhte oleks kitsas?

15

Tähelepanekuid

Uuring 1

10 90
190 810

Uuring 2

10 90
190 810

Miks tulevad
hinnangud
erinevad?

$$\widehat{RD} = 0,09$$

$$\widehat{RR} = 0,526 \dots$$

$$\widehat{NNT} = 11,111 \dots$$

$$\widehat{OR} = 0,473684 \dots$$

$$\widehat{RD} = 0,05$$

$$\widehat{RR} = 0,5.$$

$$\widehat{NNT} = 20.$$

$$\widehat{OR} = 0,473684 \dots$$

16

Rohkem kui kaks töötlust (2xk sagedustabelid)

	süda haige	süda terve	Kokku
Muud õlid	1 971	397 715	399 686
Tarbitud oliiviõli	2 658	638 583	641 241
kogus	367	106 313	106 680
päevas	491	150 743	151 234
Kokku	5 487	1 289 838	1 295 325

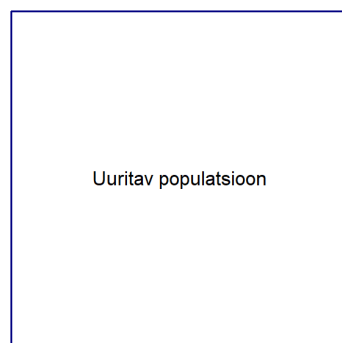
$OR = ???$

	OR	95%-usaldusintervall
ei tarbi	1	
0...1	0,84	0,79 ... 0,89
1...1,5	0,70	0,62 ... 0,78
1,5....	0,66	0,60 ... 0,73

20

Mida mõõdavad erinevad katseplaanid?

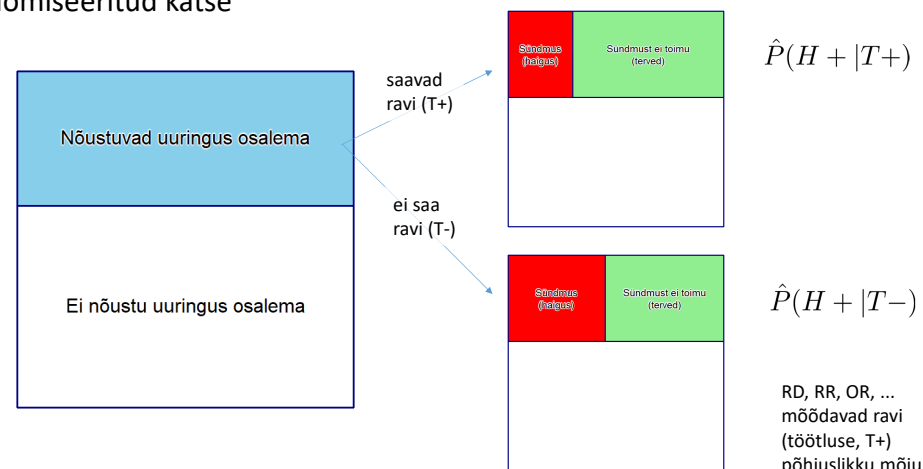
Randomiseeritud katse



21

Mida mõõdavad erinevad katseplaanid?

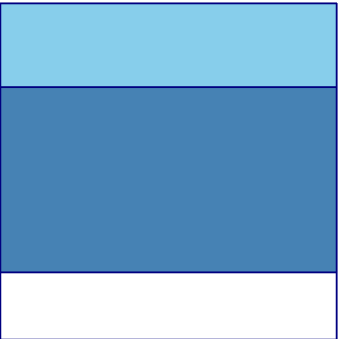
Randomiseeritud katse



23

Mida mõõdavad erinevad katseplaanid?

Mitterandomiseeritud kohortuuring



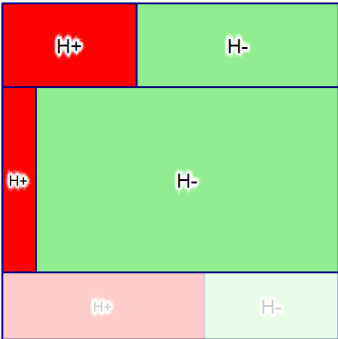
Saavad ravi
(sest on kõrge haigusriskiga või rikkad või hea arsti patsiendid...)

Ei saa ravi
(sest näivad terved või on vaesed või on uusi ravimeetodeid mittetundva arsti patsiendid...)

Inimesed kes ei nõustu vastama
(Keelduvad osalemast mistahes uuringutes või ei taha, et keegi saaks teada nende probleemidest...)

Mida mõõdavad erinevad katseplaanid?

Mitterandomiseeritud kohortuuring



$P(H + |T+)$ Tõenäosus, et riskigrupi patsient (kellele anti ennetavat ravi) haigestub...

Saavad ravi
(sest on kõrge haigusriskiga või rikkad või hea arsti patsiendid...)

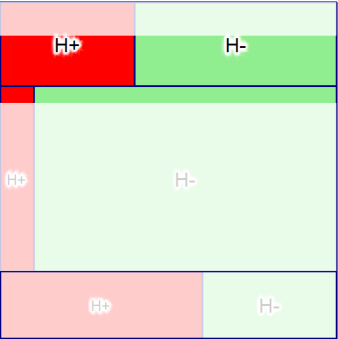
$P(H + |T-)$ Tõenäosus, et haigestub inimene kellel ohumärke polnud (ja seetõttu ennetavat ravi ei antud)

Ei saa ravi
(sest näivad terved või on vaesed või on uusi ravimeetodeid mittetundva arsti patsiendid...)

Inimesed kes ei nõustu vastama
(Keelduvad osalemast mistahes uuringutes või ei taha, et keegi saaks teada nende probleemidest...)

Mida mõõdavad erinevad katseplaanid?

Mitterandomiseeritud kohortuuring



Saavad ravi $\hat{P}(H + |T+)$

valim $\hat{P}(H + |T-)$

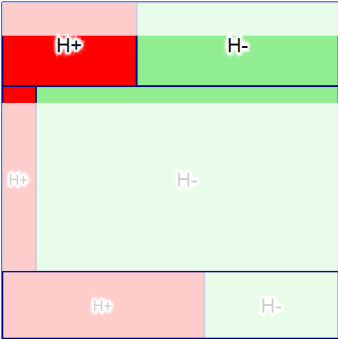
Ei saa ravi

Keerulisem on aga hinnata tõenäosuseid $P(H+)$ $P(T + |H+)$ $P(T - |H+)$

Inimesed kes ei nõustu vastama

Mida mõõdavad erinevad katseplaanid?

Mitterandomiseeritud kohortuuring



Saavad ravi $\hat{P}(H + |T+)$

valim $\hat{P}(H + |T-)$

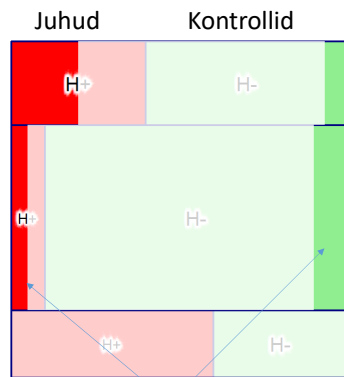
Ei saa ravi

Keerulisem on aga hinnata tõenäosuseid $P(H+)$ $\neq P(T+)$

$P(H + |valim) = P(H + |T+, valim)P(T + |valim) + P(H + |T-, valim)P(T - |valim) \neq P(T-)$

Mida mõõdavad erinevad katseplaanid?

Juht-kontrolluuring (tagasivaatav analüüs)



Saavad ravi

Saab hinnata

$$\hat{P}(T+|H+) \quad \hat{P}(T+|H-)$$

Ei saa ravi

Ei saa hinnata

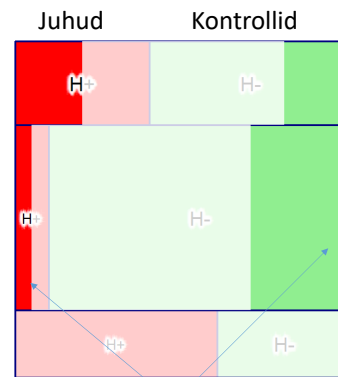
$$\hat{P}(H+|T+) \quad \hat{P}(H+|T-)$$

$$\begin{aligned} P(H+|T+, \text{valim}) &= \\ &= P(T+|H+, \text{valim})P(H+|\text{valim})/P(T+|\text{valim}) \\ &\approx P(T+|H+) \neq P(H+) \end{aligned}$$

34

Mida mõõdavad erinevad katseplaanid?

Juht-kontrolluuring (tagasivaatav analüüs)



Saavad ravi

Saab hinnata

$$\hat{P}(T+|H+) \quad \hat{P}(T+|H-)$$

Ei saa ravi

Ei saa hinnata

$$\hat{P}(H+|T+) \quad \hat{P}(H+|T-)$$

Ei saa hinnata ka nende
tõenäosuste põhjal
arvutatud suuruseid:

Riskide vahe, suhteline risk, ...

Ühe erandiga....

šansside suhet (OR)
saab jätkuvalt hinnata!!

Meenutuseks...

$$OR = \frac{P(H+|T+)/P(H-|T+)}{P(H+|T-)/P(H-|T-)}$$

$$= \frac{P(H+|T+)P(T+)/[P(H-|T+)P(T+)]}{P(H+|T-)P(T-)/[P(H-|T-)P(T-)]}$$

$$P(A|B)P(B) = P(B|A)P(A)$$

$$= \frac{P(T+|H+)P(H+)/[P(T+|H-)P(H-)]}{P(T-|H+)P(H+)/[P(T-|H-)P(H-)]}$$

$$= \frac{P(T+|H+)/P(T+|H-)}{P(T-|H+)/P(T-|H-)}$$

Neid tõenäosuseid oskame hinnata
ka juht-kontroll uuringu korral!

37

Meenutuseks...

Näide 2: kas COVID-vaktsiinist on kasu (2021 algus)?

	COVID	muu haigus
vaktsineeritud	1	18
vaktsineerimata	186	212
KOKKU	187	230

Küsitletakse 187 COVID-iga
arsti juurde pöördunud
inimest ja 230 muude
hingamisteede haigustega
arsti juurde sattunud
patsienti

Mida kasutaksid siin seose tugevuse iseloomustamiseks?

~~riskide vahe~~

~~suhteline risk~~

šansside suhe

$$\widehat{OR} = \frac{1 \cdot 212}{18 \cdot 186} \approx 0,0633$$

Tõenäosuseid $\hat{\pi}_{1|1}$ ja $\hat{\pi}_{1|2}$
pole võimalik hinnata!

41