

Kvalitatiivsete andmete analüüs
3. praktikum
Seosekordajad 2x2 tabelitest

Ülesanne 1. Alkohol ja alaealised

Alkohoolsete jookide testostlemise käigus oli üheks uuritavaks teguriks alaealistelt alkoholiostjatelt dokumendi küsimine. Tabelis on saadud tulemused (osakaalud) soo kaupa.

	ei küsitud	küstiti
poiss	0.553	0.447
tüdruk	0.619	0.381

Millised seosekordajad saab antud tabeli põhjal hinnata? Mis tüüpi uuringuga on tegemist?

Tee arvutused. Kuidas tulemusi tõlgendada?

Kas antud tabeli põhjal saaks leida ka usaldusintervalle või kontrollida hüpoteese?

Uuringu läbiviijatelt saame teada, et testostlemises osales 47 poissi ja 21 tüdrukut. Kas seda informatsiooni kasutades on võimalik usaldusintervalle seosekordajatele (näiteks riskide vahele või absoluutse riski suurenemisele) leida? Riskide vahe korral 95%-usaldusintervallid olid leitavad nii:

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


 $p_C(1 - p_C)/n_C + p_T(1 - p_T)/n_T$

Vaata ka R-i käsku prop.test. Leia selle käsu abil 95%-usalduspiirid riskide vahele. Kas jõuad sama tulemuseni kui ise arvutades?

NB! Kui kohe sama tulemuseni ei jõudnud, siis uuri veel – prop.test käsu abil on võimalik arvutada usalduspiire ka täpselt sellesama loengus tutvustatud arvutuseeskirja järgi. Lihtsalt tuleb õige arvutusviis üles leida!

Ülesanne 2. Suitsetamine ja infarkt

Põhja-Itaalia haiglates viidi läbi uuring, kus infarktiga haiglasse sattunutelt uuriti nende suitsetamisharjumuste kohta (kas on praegu või minevikus olnud suitsetaja). Igale infarktipatsiendile valiti võrdluseks samadest haiglatest kaks tausta poolest sarnast kontrollpatsienti (nemad olid haiglas muude haiguste tõttu) ja küsiti neilt samad küsimused suitsetamise kohta. Tulemused on järgmises tabelis:

	infarkt	kontroll	
suitsetaja (minevikus/praegu)	172	173	345
pole kunagi suitsetanud	90	346	436
	262	519	781

Kas selle tabeli korral saab hinnata infarkti saamise tõenäosust suitsetajatel?

Millis(t)e näitaja(te) abil saab siin iseloomustada seost suitsetamise ja infarkti saamise vahel? Leia sobivad näitajad ja interpreteeri neid!

Leia ka hinnang šansside suhtele ja loengus tutvustatud valemit kasutades leia ka ligikaudne 95%-usaldusintervall šansside suhtele!

R-is saab usaldusintervalli šansside suhtele leida päris mitmel moel. Siin on ära toodud mõned võimalikud variandid (kõigil juhtudel on käsu väljundist võimalik leida usaldusintervalli šansside suhtele, OR-le):

```
sagedustabel=rbind(c(172,173),c(90,346))

fisher.test(sagedustabel)

# install.packages("epitools")
library(epitools)

oddsratio(tabel, method="midp")
oddsratio(tabel, method="wald")
```

Milline neist käskudest arvutab usaldusintervalle loengus tutvustatud viisil?

Märkus: kui üks marginaaljaotustest on juhuslik, siis on soovitatav kasutada „midp“-valikut (keskmiselt kitsamad usaldusintervallid – vähemalt nendest variantidest, mille puhul õige OR väärtus jääb usaldusvahemikku garanteeritult vähemalt tõenäosusega 95%)

Tudenginäide

Päriselus tahame seosekordajaid enamasti arvutada andmemaatriksite pealt (kus on üks rida ühe uuritava kohta). Sellisel juhul koostame sagedustabelid nn töö käigus. Uurime näiteks, kas meestudengid sattuvad sagedamini haiglasse kui naised:

```
# Loeme sisse Tartu Ülikooli tudengite
# küsitlemisel saadud andmestiku:
load(url("http://www-1.ms.ut.ee/mart/linmud2020/kysitlus.RData"))
tudengid[1:3,]
attach(tudengid)
# Leiame mõned huvipakkuvad seosekordajad, näiteks šansside suhte ja
riskide vahe:
epitab(table(sugu, haiglaravi))
```

Interpretatsioon: Meestel on hinnanguliselt 1,028 korda suuremad šansid haiglasse sattuda, kuid kuna usaldusintervall sisaldab väärtust 1 (tegelik šansside suhe võib olla ka alla ühe), siis on ka võimalik, et hoopis naised vajavad sagedamini haiglaravi.

Šansside suhte asemel võime muidugi vaadata ka riskide vahet. Riskide vahele leitud usaldusintervall sisaldab väärtust 0 – seega ei saa me kindlalt öelda, kas naistel või meestel on suurem haiglaravivajadus:

```
prop.test(table(sugu, haiglaravi))
```