

Praktikum**Tõepärasuhte test ja põhjuslikud mõjud**

Ülesanne 1

Vaatlused $X_i \sim N(\mu; \sigma)$, $i=1 \dots n$ on teineteisest sõltumatud. Soovitakse testida hüpoteesi $H_0: \mu=0$. Milline näeb välja tõepärasuhte testi teststatistik?

Ülesanne 2

Aafrika savannis paikneb väike külake. Suvekuivuse all kannatavas savannis puhkeb tulekahju, mis põletab maha kõik, mis ta teele jääb. Nüüd võtab tulemüür suuna savannikülale ja miski, mida külaelanikud võiks ette võtta, ei suudaks küla tulemõllu eest kaitsta. Seda mõistes läheb Jack lähima maja juurde ja süütab maja. Maja põleb maani maha. Kas see, et Jack süütas maja, on maja mahapõlemise põhjuseks (lähtu kontrafaktuaalsest põhjuslikkusest). Põhjenda oma otsust.

Ülesanne 3

Haiglas vaevlevad haige südamega patsiendid. Kuna nad on hädised (arsti silma jaoks kõik ühtmoodi kehvast seisusest), siis võib südameoperatsioon neid tappa. Aga ka opereerimata jättes võivad nad surra. Kas opereerimine on kahjulik või kasulik? Mida on mõistlikum teha?

Oli patsiente, keda tegelikult opereeriti, oli ka neid patsiente, keda tegelikult ei opereeritud. Seda kas kedagi opereeriti ja mis patsiendiga edasi juhtus pandi kirja. Nii kogutud andmed kui ka see, mis patsientidega oleks juhtunud ühe või teise ravi korral on toodud ära järgmises tabelis (binaarne tunnus Y on tabelis esitatud kasutades kodeerinut 1 – patsient suri; 0 – patsient jäi ellu):

Nimi	Seda osa uurija ei näe		nähtud andmed	
	Y kui opereerime	Y kui ei opereeri	X (operatsioon)	Y (suri?)
Rheia	0	1	-	1
Kronos	1	0	-	0
Demeter	0	0	opereerime	0
Hades	0	0	opereerime	0
Hestia	0	0	opereerime	0
Poseidon	1	0	-	0
Hera	0	0	-
Zeus	0	0	opereerime
Artemis	1	1	-
Apollo	1	0	-

Milline tuleb nn keskmine põhjuslik mõju (eeldame, et siin lihtsuse mõttes, et antud 10 patsienti moodustavadki uuritava populatsiooni):

$$E(Y \text{ operatsiooni korral}) - E(Y \text{ ilma operatsioonita}) = ?$$

Milline on opereeritud patsientide ja opereerimata patsientide keskmiste erinevus?

$$E(Y|X=\text{opereerime}) - E(Y|X=-) = ?$$

Miks tuli tulemus erinev? Milles seisneb erinevus? Mida peaks muutma, et ka tegelikult kogutavate andmete põhjal võiksime jõuda sarnaste tulemusteni kui eelmisel leheküljel?

Gruppide võrreldavaks muutmine

Randomiseeritud katse või randomiseeritud uuring on parim viis põhjuslike mõjude tuvastamiseks ja nende mõju tugevuse kirjeldamiseks. Paraku on randomiseeritud katsete korraldamine sageli raske või võimatu – randomiseeritud uuringu korraldamine võib sageli olla ebaeetiline (näiteks kui soovime mõõta mingi teadaolevalt tervistkahjustava tegevuse põhjusliku mõju suurust – pole eetiline käskida juhuslikult valitud inimestel oma tervist kahjustama hakata) või kallis ja aeganõudev (peame leidma näiteks uued patsiendid, keda ravida ja tuleb oodata aastaid enne kui selgub, kuidas neil läks uut ravi saades või ravi mitte saades/tavapärast ravi saades). Sestap üritatakse ikkagi lugeda välja põhjuslikke mõjusid ja nende mõjude suuruseid olemasolevatest vaatlusandmetest.

Vaatlusandmete korral aga võivad uuritavate inimeste grupid olla juba algselt erinevad (näiteks: uut ravi saanud patsiendid võivad olla rikkamad, neil on vähem elumuresid ja nad saavad endale lubada paremat ravi ja hoolitsust võrreldes nendega kes jäid ravita). Sellisel juhul üritatakse kunstlikult teha inimesi või uurimisobjekte võrreldavateks. Näiteks uurides seda, kas õlle joomine teeb tudengi pikemaks või mitte (kas õlle joomine põhjustab tudengi pikemaks muutumist) peaksime mitte võrdlema õlut joovaid tudengeid nendega kes õlut ei joo (sest õlut joovad tudengid on enamasti mehed ja mehed kipuvadki olema pikemad kui naised) vaid peaksime võrdlema õlut joovaid naistudengeid õlut mittejoovate naistudengitega ja õlut mittejoovaid meestudengeid õlut mittejoovate meestudengitega (sellisel juhul avastaksime, et õlut joovad ja õlut mittejoovad samast soost tudengid on keskmiselt sama pikad – ehk kui muudame grupid algselt võrreldavaks, siis kaob ka uuritava tunnuse mõju).

Üldisemal kujul kirjapandult: kui soovime uurida X -tunnuse põhjuslikku mõju tunnusele Y aga erineva X -tunnuse väärtustega inimesed kipuvad juba algselt erinema ka S -tunnuse väärtuste poolest (S on segav faktor või segav tunnus, näiteks sugu õllejoomise mõju tuvastamise näites) siis X -tunnuse põhjusliku mõju kirjeldamiseks peaksime võrdlema tinglikke jaotuseid. Kui $f_{Y|X,S} = f_{Y|S}$ siis ütleme, et X -tunnusel pole põhjuslikku mõju Y -le (eeldades, et oleme tinglikustanud üle kõikide mõeldavate segavate faktorite/tunnuste). Samuti võime võrrelda näiteks tinglikke keskväärtuseid saamaks aru kui tugev on põhjuslik mõju: $E(Y|X=\text{“uus ravi“}, S) - E(Y|X=\text{“vana ravi“}, S)$.

Toodud ideed illustreerib järgmine ülesanne.

Ülesanne 4

Enamik valvearste (kuid mitte kõik) kipuvad raskemas seisundis (R) haiglasse toodud patsientidele määrama ravi A, heas seisundis patsientidele (H) aga määratakse enamasti ravi B.

Oletame, et ravi A saanud patsientide (grupp A) algset seisundit ja ravitulemuste jaotust kirjeldab järgmine tabel (ravitulemused – tervise seisund pallides peale ravi, suuremad numbrid näitavad paremat tervislikku seisundit):

Arsti hinnang seisundile:	R	R	H	H
Ravitulemus ravi A korral:	2	4	6	8
Antud grupi osakaal	0,3	0,3	0,2	0,2

Ravi B saanud patsientide (grupp B) algset seisundit ja potentsiaalset ravitulemust kui nad oleksid saanud ravi A kirjeldab aga järgmine jaotustabel:

Arsti hinnang seisundile:	R	R	H	H
Ravitulemus ravi A korral:	2	4	6	8
Antud grupi osakaal:	0,1	0,1	0,4	0,4

Seega tõenäosusega 0,3 on ravi A saanud patsient selline, kelle algne tervislik seisund oli raske ja kelle tervisliku seisundi hinnanguks peale ravikuuri ravimiga A anti 2 palli.

Märka, et ravimi A mõju on mõlemas grupis sama – pooled rasketest (R) patsientides saavutavad peale ravikuuri ravitulemuse 2 ja pooled saavutavad ravitulemuse 4. Samuti heas seisundis (H) patsientidest saavutavad pooled ravitulemuse 6 ja pooled ravitulemuse 8 – seda ühtemoodi nii grupis A kui ka grupis B.

Kui mõlemat patsientide gruppi (grupp A-sse sattunud patsiente ja grupp B-sse sattunud patsiente) oleks ravitud ühtemoodi (kõiki oleksid saanud ravimit A), siis milline oleks tulnud keskmiste ravitulemuste erinevus?

$$E(\text{ravitulemus} \mid \text{grupp A, ravi A}) - E(\text{ravitulemus} \mid \text{grupp B, ravi A}) = \dots$$

Järeldus: antud grupid pole juba algselt võrreldavad, kui me oleksime neid ravinud samal viisil siis oleksime näinud (keskmiselt) erinevat ravitulemust – seda siis tänu segava faktori (patsiendi algne seisund) mõjule.

Millised oleksid need keskmiste erinevused tulnud siis, kui oleksime kasutanud tinglikke keskväärteid – võrrelnud raskes seisundis patsiente (R) eraldi ja heas seisundis patsiente (H) eraldi?

$$\begin{aligned} E(\text{ravitulemus} \mid \text{grupp A, ravi A, R}) &= \dots \\ E(\text{ravitulemus} \mid \text{grupp B, ravi A, R}) &= \dots \\ E(\text{ravitulemus} \mid \text{grupp A, ravi A, R}) - E(\text{ravitulemus} \mid \text{grupp B, ravi A, R}) &= \dots \end{aligned}$$

$$\begin{aligned} E(\text{ravitulemus} \mid \text{grupp A, ravi A, H}) &= \dots \\ E(\text{ravitulemus} \mid \text{grupp B, ravi A, H}) &= \dots \\ E(\text{ravitulemus} \mid \text{grupp A, ravi A, H}) - E(\text{ravitulemus} \mid \text{grupp B, ravi A, H}) &= \dots \end{aligned}$$

Järeldus: vaadates raskes seisundis patsiente on kaks gruppi võrreldavad (grupis A olevad raskes seisundis patsiendid saaksid sama ravi korral sama ravitulemuse kui grupis B olevad raskes seisundis patsiendid). Samuti on omavahel võrreldavad heas seisundis patsiendid (grupis A olevad heas seisundis patsiendid saaksid sama ravi korral sama ravitulemuse kui grupis B olevad heas seisundis patsiendid).

Oletame, et ravi B puhul tuleb ravitulemus tegelikult alati 1 palli võrra kehvem, kui ravi A kasutades. Milline näeb välja uurijate poolt vaadeldud ravitulemuste jaotus?

Grupp A (said ravi A)					Grupp B (said ravi B)				
Ravitulemus	Ravitulemus
P(Ravitulemus)	P(Ravitulemus)

Milline tuleb keskvärtuste erinevus:

$$E(\text{Ravitulemus} \mid \text{ravi A}) - E(\text{Ravitulemus} \mid \text{ravi B}) = \dots$$

Kas saadud keskvärtuste erinevus kirjeldab ravimi B tegelikku põhjuslikku mõju (seda, et ravimit B kasutades saavutame alati 1 ühiku võrra kehvema ravitulemuse võrreldes ravimiga A)?

Vaatame nüüd keskvärtuste erinevust eraldi heas seisundis patsientide ja halvasti seisundis patsientide jaoks:

$$\begin{aligned} E(\text{ravitulemus} \mid \text{grupp A, ravi A, R}) - E(\text{ravitulemus} \mid \text{grupp B, ravi B, R}) &= \dots \\ E(\text{ravitulemus} \mid \text{grupp A, ravi A, H}) - E(\text{ravitulemus} \mid \text{grupp B, ravi B, H}) &= \dots \end{aligned}$$

Kas saadud keskvärtuste erinevus kirjeldab ravimi B tegelikku mõju?

Kõige järgi ei tohi tinglikustada!

Mitte iga andmestikus oleva või mõõdetud tunnuse järgi ei tohi tinglikustada – vale tunnuse järgi tinglikustades võib tulemuseks saada üsnagi eksitavaid tulemusi (kui eesmärgiks on põhjusliku mõju hindamine)! Nimelt ei tohi analüüsi tehes tinglikustada nende tunnuste järgi, mida võiks mõjutada Y-tunnus! Järgnev ülesanne näitlikustab seda, mis võib juhtuda kui muudame “grupe võrreldavaks” kasutades vale tunnust.

Ülesanne 5

Ühe haiguse patsiendid jagatakse kahte rühma. All on toodud kõigi uuritud patsientide kohta, mitu “tervisepalli” (suuremad väärtused näitavad paremat tervist) nemad saaksid peale nädalast ravi ravimiga A:

Rühm A:	1	7	13
Rühm B:	1	7	13.

Pane tähele, et antud näites on rühmad algselt võrreldavad – sama ravi saades oleksid tervisetulemused mõlemas rühmas samasugused.

Patsiendid rühmast A saavad ravi A ja rühma B kuuluvad patsiendid saavad ravi B. Lisaks panevad uurijad kirja, kas patsiendile on külastajad lilli toonud või mitte.

Oletame nüüd, et ravi B tagab ravist A 2 palli võrra parema tervisetulemuse. Oletame samuti, et meditsiiniõed lubavad lilli tuua vaid patsientidele, kellel terviseskoor on vähemalt 8. Millist andmestikku näeb uurija?

ravi	ravitulemus	lilled
.....
.....

Arst, kes uuringut läbi viib kahtlustab, et lillede olemasolu võib tõsta patsiendi meeleolu ja viia tema kiirema paranemiseni. Seetõttu soovib ta, et ravitulemuste võrdlemisel võetakse arvesse lillede olemasolu (adjusted for the presence of flowers), st. vaadeldakse neid patsiente, kellel oli palatis lilli (milline oli nende patsientide puhul ravimite erinevus), eraldi nendest, kellel lilli polnud. Milliseks osutub ravimite erinevus sellise võrdluse korral?

$$\begin{aligned} E^*(\text{ravitulemus} \mid \text{ravi A, lilled on}) - E^*(\text{ravitulemus} \mid \text{ravi B, lilled on}) &= \dots \\ E^*(\text{ravitulemus} \mid \text{ravi A, pole lilli}) - E^*(\text{ravitulemus} \mid \text{ravi B, pole lilli}) &= \dots \end{aligned}$$

Kus E^* tähistab praegu keskvärtuse hinnangut (keskmist).

Millise võrdluse abil oleksime praegu kätte saanud ravi tegeliku efekti?