

Kauzālo efektu novērtēšana ICCS sekundārajā analīzē

Reinis Alksnis

Latvijas Universitāte
2024. gada 3. oktobris

Prezentācijas saturs

1. Problēmas apraksts, kauzalitāte jeb cēloņsakarība
2. Kādas ir klasiskās pieejas to novērtēšanai
3. Beijesa aditīvo regresijas koku (BART) algoritms
4. Piemēri no IEA ICCS 2022 sekundārās analīzes

Kauzālais efekts

- **Kauzalitēte** jeb **cēloņsakarība** ir saistība starp kaut ko, kas notiek, un šī notikuma iemeslu; princips, ka nekas nevar notikt bez cēloņa. (*tulkots no Oksfordas vārdnīcas*)
- Kauzālītāti definējam caur hipotētisko scenāriju: **kādi būtu rezultāti, ja būtu citi apstākļi.**
- **Y** - atbildes mainīgais (*mūs interesē efekts uz šo lielumu*)
- **z** - *treatment* mainīgais (*interesē šī lieluma efekts*)
- **X** - kovariāti (*tiešā veidā neinteresē, bet arī ietekmē Y*)
- Kauzālo efektu matemātiski definējam kā

$$\mathbb{E}(Y(1) - Y(0) \mid \mathbf{z} = 1, \mathbf{X} = \mathbf{x})$$

Zelta standarts

Lai novērtētu kādas intervences kauzālo efektu randomizējam subjektus.

- Ar pietiekami lielu izlases apjomu varam izslēgt citu faktoru ietekmi.
- Rezultātā novērotās atšķirības varam attiecināt uz attiecīgo intervenci.
- Ja iespējams, atkārtoto mērījumu dizains.



Analoģija ar eksperimentālu pētījumu

Pētījuma mērķis: Izstrādāta jauna diēta un pētnieki vēlas noteikt, vai tā palīdz zaudēt svaru.

Eksperimenta dizains:

- z - izmanto diētu (1), neizmanto diētu (0)
- Y - svara zudums
- X - vecums un sākotnējais svars

Diētas efekts: $\mathbb{E}(Y(1) - \hat{Y}(0) | z = 1, X = x)$

- $Y(1)$ - faktiskā vērtība (*reāli novērota*)
- $\hat{Y}(0)$ - hipotētiskā vērtība (*nav reāli novērota*)



Hipotētiskās vērtības novērtēšana

Ideja:

1. Novērtējam modeli, kas skaidro Y izmantojot z un X
2. Samainam z vērtības no 0 uz 1 un otrādi
3. Izmantojot modeli veicam Y prognozi \hat{Y} pie $X = x$ un jaunā z

Iepriekšējās analogijas gadījumā:

1. Novērtējam modeli, kas skaidro svara zudumu (Y) izmantojot vecumu un sākotnējo svaru (X), kā arī to vai ievērota diēta vai nav (z).
2. Izmantojot modeli prognozējam hipotētisko svara zudumu \hat{Y} (pie X un apgriezta z).

Klasiskās pieejas

- **Lineārā regresija**

$$Y = \alpha + \beta Z + \gamma X + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

- pieņemam, ka \mathbf{z} nav korelēts ar ϵ
- pieņemam, ka \mathbf{X} satur visu būtisko informāciju, lai izskaidrotu \mathbf{Y} un \mathbf{z}
- Tad šeit kauzālo efektu iegūstam kā β novērtējumu $\hat{\beta}$

Klasiskās pieejas

- **Novērojumu sapārošana**

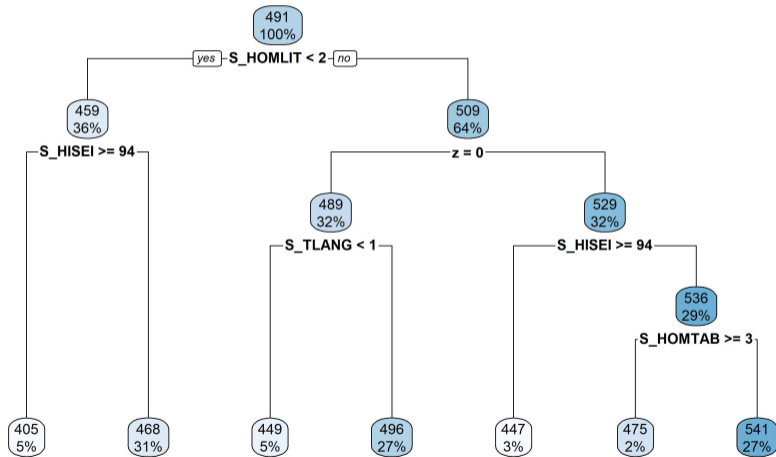
Ideja: Katram cilvēkam, kurš ievērojis diētu, atrodam maksimāli līdzīgu cilvēku (izmantojot \mathbf{X}), kurš nav ievērojis diētu.

- Šādi "sapārojot" salīdzinām vidējos svāra zudumus.
- Nepieciešams novērtēt $z = 1$ varbūtību (*propensity score*)

BART (Bayesian Additive Regression Trees)

- Chipman, H. A., George, E. I. and McCulloch, R. E. (2010). BART: Bayesian additive regression trees.
- Cabras, S. and Tena Horrillo, J. D. (2016). A Bayesian non-parametric modeling to estimate student response to ICT investment.
- Ferraro, S. (2018). Is information and communication technology satisfying educational needs at school?

Regresijas koks $Y = g(x; T, M) + \epsilon$



BART algoritms

- Modelējam Y kā šādu koku summu

$$Y = \sum_{j=1}^m g(x; T_j, M_j) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

- Beijesa pieeja
 - Regularitātes priori - vienkārši koki
 - Ticamības funkcija
 - MCMC *backfitting* algoritms
 - Aposteriorais sadalījums

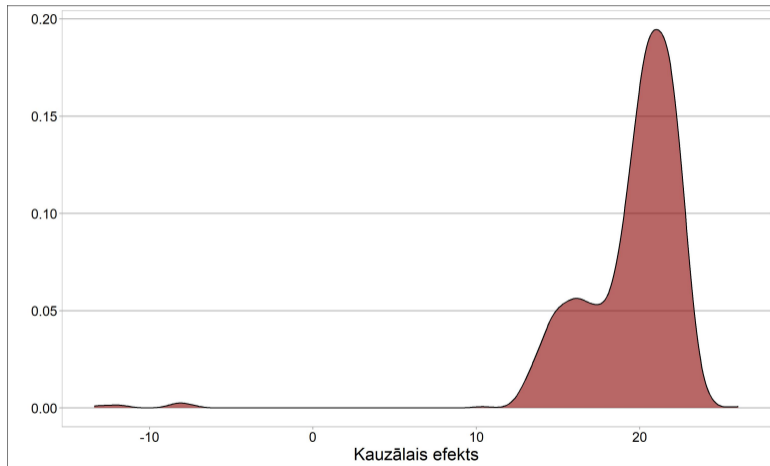
Piemērs (IEA ICCS 2022)

- Atbildes mainīgais (**Y**) - skolēnu pilsonisko zināšanu līmenis - ”*skolēnu izpratne par pilsoniskiem jēdzieniem, iesaistīšanās pilsoniskās aktivitātēs un attieksme pret pilsoniskumu*”
- *Treatment* (**Z**)- augstākais abu vecāku izglītības līmenis
 - 0 - nav bakalaura grāds
 - 1 - ir vismaz bakalaura grāds
- Kovariāti (**X**) - skolēnu un skolas līmeņa mainīgie lielumi

Izpētes jautājumi:

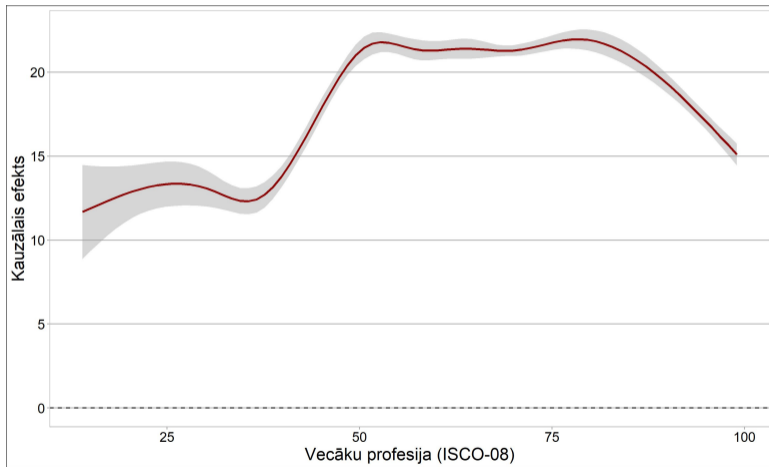
1. Kads ir vecāku izglītības efekts uz skolēnu sasniegumiem pilsonisko zināšanu testā?
2. Nelineārie efekti atkarībā no dažādiem kovariātu līmeņiem.

Kauzālā efekta sadalījums



- Pozitīva efekta varbūtība: **0.99**
- Efekta lielums: **19.83**

Efekts atkarībā no vecāku profesijas

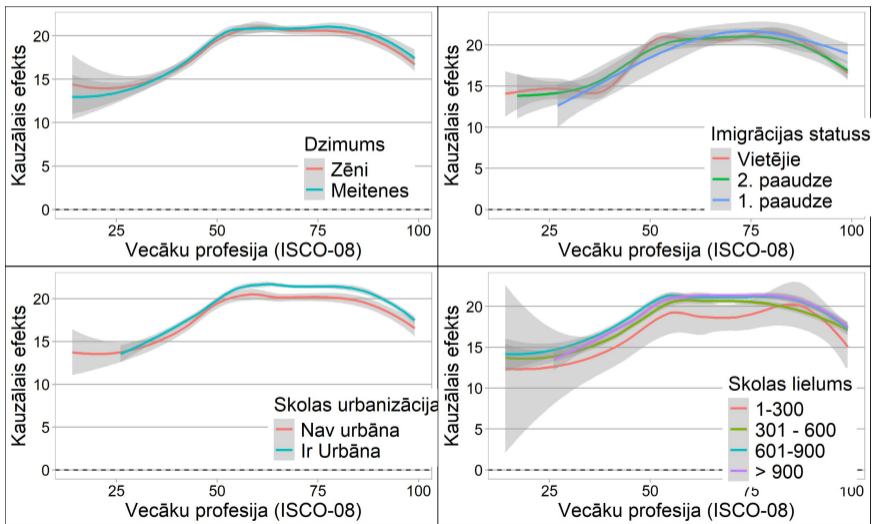


Efekts **lielāks**,
ja vecākiem
ir **augstākas**
klasifikācijas
profesija.

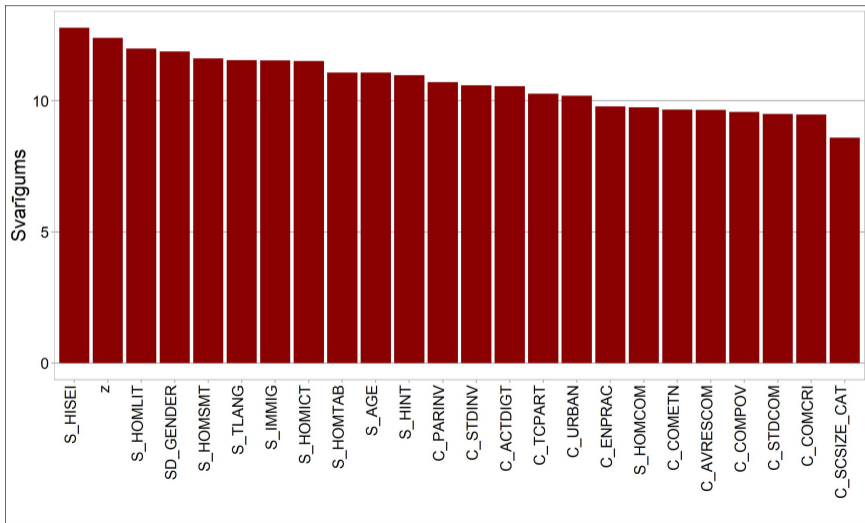
Nosacītie efekti

Grupa	Varbūtība	Efekta	Ticamais intervāls (95%)
Puiši	0.993	19.37	(7.79, 24.32)
Meitenes	0.995	20.28	(8.82, 24.42)
Vietējie	0.994	19.97	(8.70, 24.37)
2. paaudze	> 0.999	17.69	(7.23, 24.06)
1. paaudze	> 0.999	19.01	(9.75, 24.38)
Nav urbāna	0.992	19.23	(7.76, 22.89)
Ir urbāna	> 0.999	20.72	(8.72, 24.55)
1 - 300	0.970	18.75	(2.69, 23.59)
301 - 600	> 0.999	18.93	(7.32, 23.58)
601 - 900	> 0.999	20.64	(12.56, 24.08)
> 900	> 0.999	20.19	(8.81, 24.29)

Nosacītie efekti



Mainīgo svarīgums

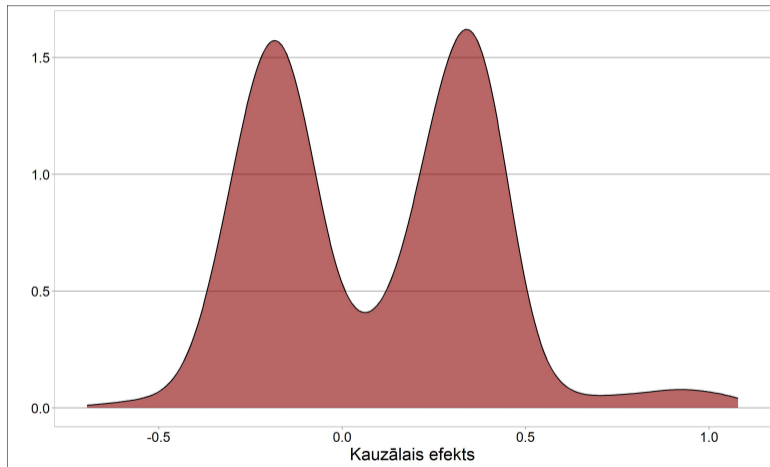


2. Piemērs (IEA ICCS 2022)

- Atbildes mainīgais (**Y**) - skolēnu uzticību pilsoniskajām institūcijām
- *Treatment (z)*- augstākais abu vecāku izglītības līmenis
 - 0 - nav bakalaura grāds
 - 1 - ir vismaz bakalaura grāds
- Kovariāti (**X**) - skolēnu un skolas līmeņa mainīgie lielumi

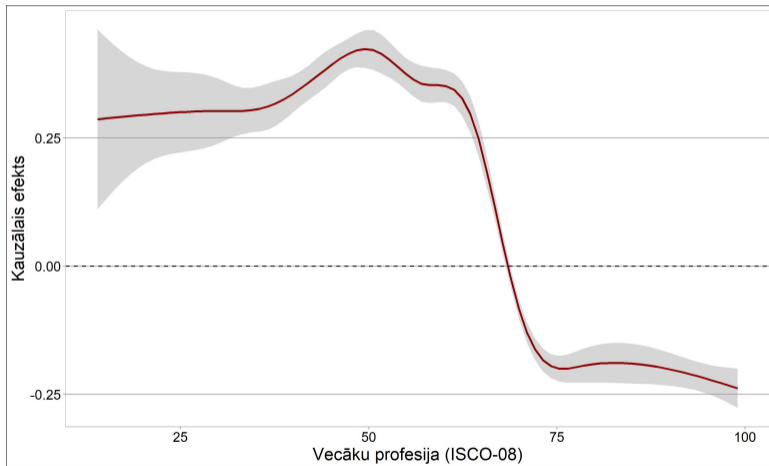


Kauzālā efekta sadalījums



- Pozitīva efekta varbūtība: **0.55**
- Efekta lielums: **0.10**

Efekts atkarībā no vecāku profesijas

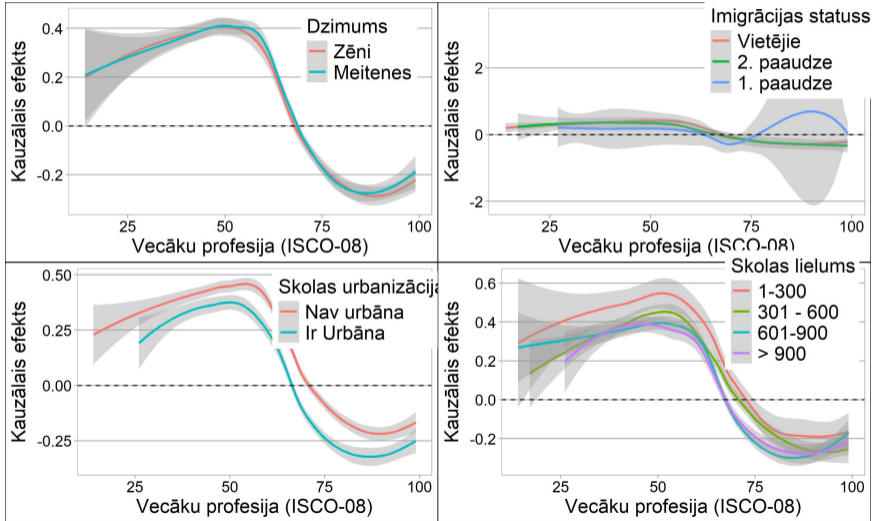


Efekts **negatīvs**,
ja vecākiem
ir **augstākas**
klasifikācijas
profesija.

Nosacītie efekti

Grupa	Varbūtība	Efekta	Ticamais intervāls (95%)
Puiši	0.53	0.09	(-0.36, 0.86)
Meitenes	0.56	0.11	(-0.35, 0.63)
Vietējie	0.55	0.10	(-0.34, 0.76)
2. paaudze	0.63	0.09	(-0.38, 0.96)
1. paaudze	0.52	-0.05	(-0.38, 0.33)
Nav urbāna	0.60	0.17	(-0.29, 0.80)
Ir urbāna	0.49	0.01	(-0.38, 0.80)
1 - 300	0.53	0.16	(-0.31, 1.00)
301 - 600	0.62	0.15	(-0.34, 0.62)
601 - 900	0.52	0.07	(-0.32, 0.46)
> 900	0.54	0.07	(-0.36, 0.81)

Nosacītie efekti



Kopsavilkums

- Kauzālo efektu novērtēšana var būt izaicinoša novērojumu tipa datiem.
- Izmantojot BART algoritmu ērti novērtēt nelineārus, kā arī jauktos efektus.
- Var iekļaut lielu skaitu skaidrojošo mainīgo \mathbf{X} .
- Metode nav plaši lietota liela mēroga izglītības pētījumos.
- Daudz vietas attīstībai.

Paldies par uzmanību!

Jautājumi?

- Reinis Alksnis
- reinis.alksnis@lu.lv
- +371 22317921