

Andmebaasi struktuur,  
genereeritud tunnused, kaalud.

Eesti standardtabelid.

Liili Abuladze

27.11. 2014

SHARE andmebaasi koolitus



TALLINNA ÜLIKOOL  
Eesti Demograafia Instituut



# Andmebaasi struktuur

- ▶ Iga moodul on eraldiseisev ja tähistatud mooduli tähekombinatsiooniga (AC; AS; CV; DN jne)
- ▶ Kõiki mooduleid üheks failiks ei ole võimalik luua (liiga mahukas, ükski programm ei toeta)
- ▶ Seega tuleb kõigepealt tutvuda ankeediga (iga laine eestindatud/venendatud küsimustikud Küsitlusinstrumentaariumi kataloogis <http://www.share-estonia.ee/index.php?id=47>)
- ▶ Genereeritud moodulite failid on sõnalise pealkirjaga, sh Eesti moodul (w4: EE\_eesti, 5.laines veel pole)



# Moodulite lühendid ja nimetused

ac	Tegevused	hh	Leibkond
as	Varad	ho	Elamuga seonduv
br	Käitumisriskid	housing	Elamuga seonduv - genereeritud
cf	Kognitiivne funktsioon	imputations	Imputeeritud tunnused
ch	Lapsed	isced	Haridus
co	Tarbimine	it	IT oskused (w5)
cv	Taustaandmed	iv	Intervjueerija märkused
depriva- tion	Deprivatsioon (w5) – gen.	mc	Lapsepõlvkodu (w5)
dn	Demograafilised andmed	mh	Vaimne tervis
eesti	Eesti moodul	networks	Võrgustikud – genereeritud
ep	Töötamine	pf	Kopsufunktsioon
ex	Ootused	ph	Füüsiline tervis
ft	Rahalised ülekanded	sn	Võrgustikud
gs	Käetugevus	sp	Sotsiaalne tugi
hc	Tervishoid	weights	Kaalud - genereeritud
health	Tervis – genereeritud	xt	Elulõpu andmed

# Iga mooduli fail sisaldab

- ▶ mergeid – indiviidi identifitseerimise tunnus, failide ühendamisel kasutatav põhitunnus
- ▶ hhid4, hhid5 – leibkonna identifitseerimine 4. ja 5.laines
- ▶ cvid, cvidp – valimi põhiisik, põhiisiku partner
- ▶ Ülejäänud vastava mooduli tunnused



TALLINNA ÜLIKOOL  
Eesti Demograafia Instituut



Euroopa Liit  
Euroopa  
Regionaalarengu Fond



Eesti tuleviku heaks

# Genereeritud moodulid

- ▶ gv\_health – genereeritud tervisenäitajad – isiku peale
- ▶ Genereeritud vastused (nt GALI või sphus, sphus2) on moodustatud rahvusvahelist analüüsitava järgides (nt w4\_health):
  - ▶ gali – oma igapäevategevustes piiratud (väärtusega 1) või mitte (0)
  - ▶ sphus- USA terviseküsimus (suurepärase, väga hea, hea); sphus2 – kahene koondatud sphus (suurepärase-väga hea ning vähem)
  - ▶ eurod – depressiooni skaala, koondatud erinevatest tunnustest
- ▶ gv\_housing: genereeritud elamutunnused (sh NUTS3), tüüp, korruselisus – leibkonna peale
- ▶ gv\_isced: genereeritud haridustunnus (alates w5 ka ISCED11) isiku peale
- ▶ gv\_weights: kaalud 50+ kogurahvastiku peale (indiviidid ciw4)
- ▶ imputations: arvestades puuduvaid vastuseid imputeeritud tunnused (Eestis suhteliselt vähe puuduvaid väärtusi w4-s, ciw4\_f=0 ja balanced\_ind ja balanced\_HH=0 → võib ignoreerida). Saab kasutada seal olevaid imputeeritud tunnuseid keskmiste arvutamiseks.



# Genereeritud moodulid: cv\_r

- ▶ *cv\_r* – *coverscreen*
- ▶ *cvid* – valimi põhiisik, *cvidp* – valimiisiku partner
- ▶ *partnerinhh* – partner on leibkonnas (1) või mitte (3)
- ▶ *dumfinr* - finantsküsimustele, *dumfamr* – peret puudutavatele, *dumhhr* – leibkonda puudutavatele küsimustele vastanud isik
- ▶ *relrpers* – suhe respondenti
- ▶ *hhsiz*e – leibkonna suurus



# Genereeritud moodulid: gv\_isced

- ▶ Haridustase: respondendi (isced\_r), partneri (isced\_sp), lapse (isced\_c), vanemate (isced\_m, isced\_f → w5) kohta
- ▶ 5.laines info puudu, sest 4.laines küsitud → tuleb ühendada 4.laine andmetega
- ▶ 5.laines on nii ISCED97 kui ISCED11 klassifikaatorite tunnused
- ▶ Vaata koodide tähendusi ja vastavusi nt siit: <http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx>



TALLINNA ÜLIKOOL  
Eesti Demograafia Instituut



# Genereeritud moodulid: gv\_networks (w4)

- ▶ Moodustatud SN, CH, SP, FT, DN moodulite põhjal
- ▶ sizeofsocialnetwork – väärtusega 0 kuni 7
- ▶ spousesnet – partner ka võrgustikus?
- ▶ famnet – pereliikmed võrgustikus: childnet, gchildnet, siblingnet, parentnet (arv, dummy, %)
- ▶ kontaktihedus (1 – kõige tihedam, 7 – mitte kunagi), geograafiline kaugus, emotsionaalne lähedus, abi andmine ja saamine
- ▶ sn\_satisfaction – rahulolu võrgustikuga (1-10), k.a 0-liikmelist võrgustikku omavatel





# Genereeritud moodulid: gv\_db\_deprivation (w5)

- ▶ Indeks: BR, CO, HO, HH moodulite põhjal – 15 küsimust 17-st küsitud leibkonna tasemel
- ▶ Loetakse deprivatsiooniks kui vähemalt üks leibkonna liige on maininud ilmajäetust
- ▶ deprivation\_equal – kõigile indeksi elementidele on antud võrdne kaal
- ▶ deprivation\_hedonic – kaalutud Eurobaromeetris leiduva olulisusastmega iga elemendi suhtes (v.a. CH, IL)



# Genereeritud moodulid: gv\_db\_deprivation (w5)

Sotsiaalne tõrjutus (w5) (MN032_socex=1) - ilmajäetus			
BR033	Liha toidulaual	CO207	Vaesuse tõttu vanad riided
BR034	Puuvili toidulaual	CO208	Vaesuse tõttu vanad jalanõud
HH023	Vandalismioht naabruskonnas	CO209	Vaesuse tõttu külm Kokkuhoidmise tõttu lükkas
HC024	Naabruskond puhas	CO211	hambaarsti külastust edasi
HH027	Olemas pangautomaat	CO213	Loobunud vajalikest prillidest või nende ostust kokkuhoiu eesmärgil
HH029	Olemas perearst	HO017	Regulaarse laenu maksmine
HH030	Olemas apteek	HO022	Maksed võlgu 2+ kuud
C0202	Saab lubada 1-näd puhkust	HO010	Üürivõlg 2+ kuud
CO206	Saab lubada ettenägematuid kulutusi		



# Genereeritud moodulid: gv\_weights

- ▶ dw\_w4 – disainikaal, ei soovitata kasutada
- ▶ Kalibreeritud kaalud – üldistuste tegemiseks
  - ▶ chw\_w4 – kalibreeritud leibkonna kaal → analüüsiüksus: leibkond
  - ▶ ciw\_w4 – kalibreeritud indiviidi kaal → analüüsiüksus: individ
- ▶ \_f – flag – keda kaalumisel pole arvestatud (=1)
  - ▶ alla 50-aastased vastajad (respondentide partnerid), puuduva infoga (sünniaeg, sugu, NUTS1 kood), puuduva disainikaaluga
- ▶ Loe rohkem SHARE w4 rel guide 1.1.1, peatükk

## 14.3: Weights (Dropboxis)



TALLINNA ÜLIKOOL  
Eesti Demograafia Instituut



**SHARE**  
Survey of Health, Ageing  
and Retirement in Europe  
50+ in Europe



Euroopa Liit  
Euroopa  
Regionaalarengu Fond



Eesti tuleviku heaks

# Küsimuste vastuste loogika (1)

- ▶ Iga küsimust identifitseerib moodul, küsimuse number ja alakriips nt AC001\_
- ▶ Küsimused, mis on mitu korda küsitavad või mitmeste valikutega (mis tüüpi sissetulek, mis perioodil)
  - ▶ ph068\_3 viitab, et see on vastus antud küsimuse kolmanda juhu kohta,
  - ▶ ep112\_2\_3 viitab aastale, millal saadi 2.tüüpi pensioni esimest korda vastavas 3.perioodis,
  - ▶ ph010d3 näitab küs ph010\_ kolmandale vastusvariandi valikut
- ▶ Rahalist väärtust hindavate küsimuste puhul tuleb lisaks arvestada, et nõutse vastusele (nt ho005e) on lisaks ka vahemikväärtused
  - ▶ ho005ub – ütleb, milline vahemik on valitud ja kui mitme isiku poolt,
  - ▶ ho005v1 , ho005v2, ho005v3 on väärtused kvartiiliti,
- ▶ Kui on keeldunud või ei ole osatud /teatud vastata, siis on tunnuses need kajastatud – (miinus) märgiga. Võib kodeerida ümber, et saada aru, kuivõrd eriline rühm see vastajatest on. Juhul, kui küsimusest on ülehüpe, siis on vastused puudu (missing) ja tervikkogumi hindamiseks tuleb nad lisada ning samuti ümberkodeerida (tavaliselt 0).



# Küsimuste vastuste loogika (2)

- ▶ Lahtised vastused (st lisavariandid olemasolevatele vastusvariantidele) dot tunnuses
  - ▶ ph080dot (millises organis veel oli vähk, kui 22 eelnevas polnud)
- ▶ Vanus – 50+ rahvastiku peale arvestamiseks kasutatakse laine üht kindlat ajapunkti (w4 1.01.2010, w5 1.01.2013, võetud Eurostatist, (nüüd on uus rahvaarv, ei vasta enam täpselt))
- ▶ Vanus sündmuse hetkel ((sündmuse aasta\*12 + kuude arv) – (sünniaasta \* 12 + sünnikuu)) / 12
- ▶ Tunnused, mille järgi soovitakse tabuleerida, tasub luua esimestena ja kaasata igast moodulist vajalike tunnuste ühendamiseks. Sugu, vanus ja vanusrühm tuleb kaasa vedada erinevate moodulite tabeldamiseks (mitte igal pool pole sugu ja sünniandmeid ees)



# Küsimuse vastustes erisused - flag

- ▶ Kui küsimustes või vastustes on maa-spetsiifilised erisused, siis on see dokumenteeritud tunnusega samanimelises tunnuses, mille lõpus on f. Viimane tähistab flag'i e.lipp – tuleb erijuhtumiga arvestada.
- ▶ Imputation failis on imputeeritud tunnuste puhul samuti ära toodud iga maa kohta tunnuste lipud.



TALLINNA ÜLIKOOL  
Eesti Demograafia Instituut



# Eesti tabeldatud andmed

- ▶ Tervise Arengu Instituudi ja Statistikaameti kodulehtedel <http://share-estonia.ee/index.php?id=59&L=0>
- ▶ Tervise enesehinnang, Toimetuleku ja igapäevategevuste piirangud, Terviseseisund, Ravimite kasutamine, Vaimne tervis, Tervishoiu- ja sotsiaalteenuste kasutamine, Kehaline aktiivsus ja kehakaal, Toitumine, Alkoholi tarvitamine, Suitsetamine, Sotsiaalne võrgustik, Ootused ja rahulolu, Majanduslik aktiivsus ja materiaalne toimetulek, Sotsiaalmajanduslikud tausttunnused, Terviseseisund, Elukoht, Toimetulek, Hõive



TALLINNA ÜLIKOOL  
Eesti Demograafia Instituut



# Eesti tabeldatud andmed (2)

- ▶ TAI andmebaasis:

- ▶ Tervisekäitumine ja tervis

<http://pxweb.tai.ee/esf/pxweb2008/Database/Uuringud/databasetree.asp>

- ▶ Statistikaameti andmebaasis:

- ▶ Vanemaealiste vaesus, toimetulek ja sotsiaalsed suhted

[http://pub.stat.ee/px-web.2001/Database/Sotsiaalelu/12Sotsiaalne\\_terjutus\\_Laekeni\\_indikaatorid/02Vanemaealiste\\_vaesus\\_ja\\_toimetulek/02Vanemaealiste\\_vaesus\\_ja\\_toimetulek.asp](http://pub.stat.ee/px-web.2001/Database/Sotsiaalelu/12Sotsiaalne_terjutus_Laekeni_indikaatorid/02Vanemaealiste_vaesus_ja_toimetulek/02Vanemaealiste_vaesus_ja_toimetulek.asp)

- ▶ Terviseseisund

[http://pub.stat.ee/px-web.2001/Database/Sotsiaalelu/13Tervishoid/05Tervislik\\_seisund/05Tervislik\\_seisund.asp](http://pub.stat.ee/px-web.2001/Database/Sotsiaalelu/13Tervishoid/05Tervislik_seisund/05Tervislik_seisund.asp)

- ▶ SHARE andmed on ainult nendes tabelites, milles on andmed alates 2010.a.



TALLINNA ÜLIKOOL  
Eesti Demograafia Instituut



Euroopa Liit  
Euroopa  
Regionaalarengu Fond



Eesti tuleviku heaks



# Praktikum

- ▶ 1. Valime analüüsi objektid, ühendame failid
- ▶ 2. Vanuse tunnuse moodustamine (CMC)
- ▶ 3. Haridustaseme leidmine 4.lainest 5.laines osalenutele
- ▶ 4. Deprivatsiooni indeksi kasutamine
- ▶ 5. Kaalude kasutamine
- ▶ 6. Imputeeritud tunnuste kasutamine



TALLINNA ÜLIKOOL  
Eesti Demograafia Instituut

