

PIAAC-aineisto ja sen analysointi

Kari Nissinen

PIAAC-tutkijatapaaminen, Helsinki

17.5.2024

Aineistojen saaminen käyttöön

OECD:n PIAAC-sivusto

<https://www.oecd.org/skills/piaac>

sisältää käytännössä kaiken tarvittavan

- Kattava kuvaus PIAAC-tutkimuksesta (sekä 1. sykli että 2. sykli)
- Tutkimusasetelmien ja mittareiden kuvaus
- Public Use -datat, oheisdokumentit
 - 1. syklin kyselylomakkeet, muuttujakuvaus ("codebooks"), jakaumataulukoita ("compendia")
- Linkit OECD:n PIAAC-julkaisuihin

Aineistojen saaminen käyttöön

Yhdistäminen muihin aineistoihin?

- Lähinnä kyseeseen tulisivat rekisteriaineistot
 - Tilastokeskus
 - Nordic-Baltic PIAAC Network => rakenteilla protokolla valikoitujen tietojen hankkimiseen eri maiden rekistereistä
 - aineiston käyttö lupa haettava
- Public Use -datat anonymisoitu ja tietosuojan kannalta kriittiseksi katsottuja muuttujia poistettu tai modifioitu
 - Onko Tilastokeskuksesta tarvittaessa saatavissa alkuperäinen laajempi data?

OECD:n PIAAC-sivusto

OECD:n PIAAC-sivuston **Data-välilehti**

<https://www.oecd.org/skills/piaac/data>

- **PIAAC Data** Vapaasti ladattavat anonymisoidut Public Use (PUF) -aineistot
 - tuetut formaatit SPSS, SAS, CSV + *import CSV -> Stata*
 - *myös kokeiden lokitiedot (PIAAC Log Files) ladattavissa*
- **Documentation** Laaja tietopaketti tutkimuksen toteutuksesta (aineiston keruu, käytetyt mittarit ym.)
 - linkit eri maiden taustakyselylomakkeisiin
 - Technical Report
 - Reader's Companion
 - codebooks, compendia

OECD:n PIAAC-sivusto

OECD:n PIAAC-sivuston Data-välilehti

<https://www.oecd.org/skills/piaac/data>

- **Tools for Analysis** Vapaasti ladattavia työkaluja datan analysointiin
 - SAS-makro, Stata-makro
 - Linkki IEA:n kehittämään International Database (IDB) Analyzer -ohjelmistoon
 - kansainvälisten arviointiaineistojen analysointiin räätälöity makropaketti (myös PISA, PIRLS, TIMSS,...)
 - Käytettävissä ainakin SPSS:n, SAS:n ja R:n kautta
- **Explore the Data** Interaktiivinen International Data Explorer (IDE) -ympäristö aineiston kuvaileviin tarkasteluihin

OECD:n PIAAC-sivusto

OECD Skills Surveys

[HOME](#)

[ABOUT](#) ▾

[PIAAC DESIGN](#)

[EVENTS](#)

[DATA](#) ▾

[PUBLICATIONS](#) ▾

[ONLINE ASSESSMENT](#)

[Survey of Adult Skills \(PIAAC\)](#) > [Data](#)

PIAAC Data and Tools

1st Cycle: Twenty-four countries participated in Round 1 of the Survey of Adult Skills (PIAAC), with data collection taking place from 1 August 2011 to 31 March 2012 in most countries. Nine countries took part in Round 2 of the assessment, with data collection taking place from April 2014 to end-March 2015. Six countries participated in Round 3, with data collection taking place from July to December 2017. Please consult the [list of countries](#) that took part in each of the Rounds.

2nd Cycle: Thirty-one countries are participating in Round 1. Data will be available on 10 December 2024.

For any queries, please contact edu.piaac@oecd.org

[Expand all](#)

+ [PIAAC Data](#)

+ [Documentation](#)

+ [Tools for Analysis](#)

+ [Explore the Data](#)

PIAAC-datan käytössä huomioon otettavaa

- 1) Aineisto on kerätty mutkikkaalla otanta-asetelmalla
 - 2) Osaamista arvioidaan ns. **plausible value** -lukuarvojen avulla (ei yksinkertaista mittaria!)
- Aineiston oikeaoppinen tilastollinen analyysi edellyttää kummankin seikan huomioon ottamista!

Mutkikkaat otanta-asetelmat

- **Ositus**: esim. maantieteelliset alueet, ikäluokat, koulutustaso (ositus vaihtelee maittain)
- Yksilöiden poimintatodennäköisyys ei ole vakio
=> **Otantapainojen käyttö** (myös vastauskadon korjaus!)
(PIAAC-datassa muuttuja SPFWT0)
- Joissakin maissa **ryväсотanta**: poimitaan esim. naapurustoja tai asuntokuntia
=> Tarvitaan klusteroitumisen huomioon ottavia menetelmiä

Mutkikkaat otanta-asetelmat

Asetelma huomioitava data-analyysissä - **keskivirheiden estimointi kriittistä!**

- A. Asetelmaperusteinen lähestymistapa ("survey statistics")
- B. Malliperusteinen lähestymistapa ("mainstream statistics")

Arviointiaineistojen (PISA, PIAAC, ym.) analysointiin räätälöidyt ohjelmistopakettit ja makrot noudattavat **asetelmaperusteista lähestymistapaa**

- otantapainot
- varianssien/keskivirheiden estimointi (jackknife ym) siten, että ositus ja ryvästyminen huomioidaan laskennassa eksplisiittisesti (ei mallintamalla)

Osaamisen arviointi plausible values - lähestymistavalla

- Vallitseva lähestymistava muissakin kansainvälisissä arviointitutkimuksissa (PISA, TIMSS, PIRLS, ICILS,...)
- Taustalla ajatus, että yksilöiden osaamisen sijasta kartoitetaan *osaamisen kansallista jakaumaa*
 - jotta kartoitus olisi kattava, tehtäviä tarvitaan paljon ja monipuolisesti
 - aihealueiden ja vaikeustasojen variointi

Osaamisen arviointi plausible values - lähestymistavalla

- Tehtävärotaatio / "matrix sampling" = laaja kokoelma osaamista mittaavia tehtäviä, joista **yksilö tekee vain osan** (koska testausaika rajallinen)
 - eri yksilöillä erilaiset tehtäväpaketit
 - tehtävien vaikeustasot vaihtelevat
- Tehtäväädaptaatio: kokeen alussa osoitettu osaaminen vaikuttaa siihen, minkä tasoisia tehtäviä testattava enimmäkseen saa

Osaamisen arviointi plausible values - lähestymistavalla

- Yksilöiden tulokset yhteismitallistetaan IRT-mallinnuksen avulla (Item-Response Theory)
 - modernia osioanalyysiä
 - tehtävien vaikeustasoilla keskeinen rooli
 - vaikeissa tehtävissä onnistuminen antaa enemmän "pisteitä" kuin helpoissa tehtävissä onnistuminen
- Yhteismitallisista tuloksista johdetaan osaamisen kansallinen jakauma

Osaamisen arviointi plausible values - lähestymistavalla

- Plausible values -lähestymistavassa yksilöiden arvioitua osaamista kuvataan **posteriorijakauman** muodossa
 - yhden kokonaispistemäärän sijasta määritetään todennäköisyysjakauma

=> latentin osaamisen mittaamiseen liittyvä epävarmuus näkyväksi
- Jakauma kuvaa, mikä on yksilön todennäköinen osaamisen taso, kun hän
 1. onnistui saamissaan koetehtävissä tietyllä tavalla
 2. on taustaominaisuuksiltaan tietynlainen

Osaamisen arviointi plausible values - lähestymistavalla

- Posteriorijakaumaa ei voida ilmaista suljetussa muodossa, mutta sitä voidaan simuloida
 - => generoidaan 10 numeerista realisaatiota
 - = 10 plausible values (PV1-PV10)

(PV:t generoitu datoihin valmiiksi =>

10 muuttujaa PV1LIT-PV10LIT lukutaidolle, 10 muuttujaa PV1NUM-PV10NUM numerotaidolle, jne.)

- Osaamisen data-analyysi suoritetaan näiden 10 PV-muuttujan avulla
 - **moni-imputoinnin** metodologia sovellettuna arviointitutkimukseen
 - PV = "puuttuvalle osaamispistemäärälle imputoitu arvo"
 - 10 eri PV-arvoa => "imputointiepävarmuus" näkyväksi

Osaamisen estimointi IRT-mallinnuksella

- Halutaan estimoida yksilön **kyky**, joka on **latentti**
- Estimoinnin perustana kaksi mallia, joita käytetään yhdessä:
 - 1) **Osioanalyysin malli**: yksilön todennäköisyys onnistua tehtävässä riippuu **yksilön kyvystä ja tehtävän vaikeudesta**
 - 2) **Latentti regressiomalli**: määrittää latentin **kyvyn populaatiojakauman ja sen yhteydet taustamuuttujiin**

Osaamisen estimointi IRT-mallinnuksella

Osioanalyysin malli ja latentti regressiomalli tuottavat *posteriorijakauman*, joka on yksilön **kyvyn todennäköisyysjakauma**, kun tiedetään

- Miten hän onnistui kokeessa
 - Kuinka vaikeita tehtäviä hän teki
 - Taustamuuttujien arvot
 - Mikä on kyvyn jakauma populaatiossa
 - Miten kyky riippuu taustamuuttujista
- Osio-analyysi
- Latentti regressio

Plausible values

Plausible value (PV) = yksilön posteriorijakaumasta arvottu realisaatio
= eräs mahdollinen yksilön latentin kyvyn ”piste-estimaatti”

- Ei ole ”puhdas” koetulos, koska ehdollistettu taustamuuttujilla latentin regression kautta.
- Oikeampi tulkinta: PV on suoritettuun kokeeseen pohjaava **ennuste** sellaisen yksilön osaamiselle, jolla on tietyt taustamuuttujien arvot.

Plausible values

- Esim. mieshenkilö, jolla on tietyt taustamuuttujien arvot (ikä, asenteet, koulutustaso, ...) ja joka onnistui saamissaan tehtävissä tietyllä tavalla: mikä olisi näiden tietojen valossa uskottava (plausible) arvio hänen osaamistasostaan?
- Pelkkä koetulos ei ratkaise, vaan myös taustamuuttujilla on ennustetta laskettaessa väliä!
 - Koetehtäviä rajallinen määrä => ennustetta täytyy "paikata" taustainformaatiolla
 - Samalla taustamuuttujien jakautuminen populaatiossa antaa välillistä informaatiota osaamisen jakautumisesta populaatiossa

Plausible values

Analogia: korkeushyppy

- Kaksi henkilöä hyppää muutamasta eri korkeudesta (vrt. rajallinen määrä PIAAC-tehtäviä) ja kumpikin ylittää/pudottaa riman samoissa korkeuksissa (vrt. onnistuu samoissa tehtävissä)
- Toinen hyppääjä on pitkä ja sporttinen, toinen lyhyt ja ylipainoinen.

Onko uskottavaa, että molemmat hyppääjät hyppäisivät yhtä hyvin jatkossakin, jos hyppykertoja lisättäisiin eri korkeuksissa (ts. onko uskottavaa, että näiden kahden hyppääjän ”osaaminen” on kaikin puolin samantasoista)?

- Tiedämme empiriasta (datasta), että pitkät ja sporttiset hyppäävät keskimäärin paremmin kuin lyhyet ja ylipainoiset.

On perusteltua ennustaa, että tässäkin tapauksessa pitkän ja sporttisen henkilön ”hyppyosaaminen” on parempaa kuin lyhyen ja ylipainoisen, vaikka testissä heidän hyppytuloksensa olivat identtiset!

Plausible values

- Jokaiselle otoksen yksilölle generoidaan 10 PV-arvoa tarkasteltavasta arviointialueesta (lukutaito, numerotaito,...)
 - ⇒ 10 PV-muuttujaa ⇒ 10 "PV-dataa"
- Kukin PV on mahdollinen estimaatti tietynlaisen yksilön latentille osaamiselle
- PV:t **eivät ole** yksilön osaamisen harhattomia estimaatteja (koska ehdollistettu taustamuuttujilla)
 - Mutta: voidaan osoittaa, että PV:t tuottavat käytännössä **harhattomat estimaatit populaatiotunnusluville** (keskiarvot, varianssit,...)
 - Validi päättely kansalliselle tasolle ja "riittävän suurille" osapopulaatioille (alueet, sukupuolet,...)

Plausible values

- Käyttämällä posteriorijakaumaa yksittäisen koetuloksen sijaan saadaan aidommin huomioitua latentin kyvyn estimointiin kuuluva epävarmuus
 - Yksilön kyvyn arviointi perustuu rajalliseen tehtävävalikoimaan, saatu tehtäväpaketti voi vaikuttaa
 - Varaudutaan mittausvirheeseen: yksilö voi yli- tai alisuoriutua kokeessa
- 10 PV-arvon käyttö tekee arviointiepävarmuuden mitattavaksi
 - Datasta tehtävissä analyyseissä (keskivirheiden laskennassa) otetaan huomioon 10 PV-arvon keskinäinen vaihtelu (ylimääräinen ”varianssikomponentti”) => **vältetään populaatiovarianssien aliestimointi**

PV-arvojen data-analyysi

- Sovelletaan **moni-imputoinnin** metodologiaa (Rubin ym.)
- Suoritetaan jokaiselle PV:lle (esim. lukutaito: muuttujat PVLIT1-PVLIT10) sama analyysi
 - ⇒ 10 (hieman) erilaista tulosta (piste-estimaatit, keskivirheet,...)
- ”Virallinen” piste-estimaatti on 10 piste-estimaatin keskiarvo
 - Esim. regressiokerroin = 10 kertoimen keskiarvo
- Keskivirhe on 10 keskivirheen keskiarvo, johon lisätään ”imputointiepävarmuus”
 - = komponentti, joka mittaa 10 piste-estimaatin keskinäistä vaihtelua

Entä jos PV-arvot ohitetaan?

Houkutusia (jos moni-imputointiin sopivaa kalustoa ei saatavilla):

- Käytetään vasteena PV-arvojen keskiarvoa
 - Paras piste-estimaatti yksilön posteriorikeskiarvolle...
 - Yksilöiden välinen varianssi voi aliestimoitua pahoin
 - Jos populaatiovarianssi ei kiinnosta (esim. korrelaatiotarkastelut – standardointi!), keskiarvoa voi käyttää
- Käytetään yhtä PV-arvoa kymmenen sijasta
 - Yksilöiden välinen varianssi voi edelleen aliestimoitua (imputointiepävarmuus ei mukana), mutta vähemmän
 - Suuressa datassa päätelmät usein lähellä kaikilla 10 PV:llä saatavia

Entä jos PV-arvot ohitetaan?

Eksploratiivisiin tarkoituksiin voi suorittaa alustavia analyysejä esimerkiksi vain yhtä PV:tä käyttäen

- Esim. relevanttien selittäjien kartoittaminen monimuuttujaiseen regressioon, rakenneyhtälömallin ”hahmottelu”
- Erilaisten mallivaihtoehtojen kokeilu
- **“Standardiohjelmistojen” (SPSS ym) käyttö mahdollista**
- **Mieluiten survey-analyysiin (“complex samples” tms) tarkoitetuilla proseduureilla => otantapainot käsitellään oikein**
 - *Muulloin otantapainot skaalattava summautumaan otoskokoon (“house weights”)*
- Keskivirheiden estimointi malliperusteisesti tai Taylor-sarjamenetelmällä, joissakin ohjelmistoissa myös replikaattimenetelmät mahdollisia
- **Alustavien analyysien jälkeen voi tehdä varsinaiset analyysit “virallisilla” (10 PV jne) työkaluilla**

Keskivirheiden estimointi ja tilastollinen päättely

A) Malliperusteinen (<-> "mainstream"-tilastotiede)

- Havaintojen oletetaan "generoituvan" tietyistä todennäköisyysjakaumasta (esim. normaalijakauma), "ääretön superpopulaatio"
- "Todellinen" populaatio on superpopulaation realisaatio
- Varianssit ja keskivirheet johdetaan tällä oletuksella

B) Asetelmaperusteinen (<-> "survey-tilastotiede")

- Havainnot tulevat äärellisestä kiinteästä populaatiosta, ei jakaumaoletuksia, ei superpopulaatiota
- Satunnaisvaihtelu on yksinomaan peräisin "otantamekanismista", jolla havaintoaineisto on saatu
- Varianssit ja keskivirheet johdetaan niistä otosten välisistä eroista, jotka havaittaisiin jos erilaisia otoksia poimittaisiin valitulla otanta-asetelmalla äärellisestä populaatiosta

Keskivirheiden estimointi ja tilastollinen päättely

A) Malliperusteinen

- Lähes koko tilastomenetelmien "paletti" käytettävissä, kaikki standardiohjelmistot käytettävissä
- Jakaumaoletus => maximum likelihood ym.
- Jos aineistossa ei hierarkiaa / klustereita => "standardianalyysit" (+ "house weights")
- Klusteroituminen => monitasomallit (+ "house weights")
- Mutkikkaiden otanta-asetelmien erityispiirteiden huomiointi saattaa olla työläästi toteutettavissa

Kansainvälisissä arviointitutkimuksissa käytössä mutkikkaat aineistonkeruuasetelmat => on omaksuttu *asetelmaperusteinen* lähestymistapa ja sitä tuetaan!

Keskivirheiden estimointi ja tilastollinen päättely

B) Asetelmaperusteinen

- **Tilastollisten menetelmien valikoima suppeampi** (perustunnusluvut, ristiintaulukot, regressiot,...)
 - Menetelmävalikoima alkanut viime vuosina laajentua (esim. rakenneyhtälömallit)
- Tarvitaan tilastollisia **erikoisproseduureja**
 - SPSS Complex Samples, SAS Survey Procedures, IDB Analyzer, Statan ja R:n tietyt proseduurit, tietyt Mplus-optiot...
- **Kriittistä on varianssien/keskivirheiden estimointi**
- Kaksi lähestymistapaa:
 - (1) Replikaatiomenetelmät (erityisesti Jackknife ja sen muunnelmat)
 - (2) Taylor-linearisointi / approksimaatio

Asetelmaperusteinen otantavarianssin estimointi

1) Taylor-linearisointi

- ”Suljetun muodon” approksimaatio, ei replikaattipainoja
- Löytyy valmiina monista ohjelmistoista, nopea soveltaa

2) Replikaatiomenetelmät

- Arviointitutkimusten ”oletuslähestymistapa”
- PIAAC Cycle 1: Jackknife-menetelmä (JK)
- PIAAC Cycle 2:ssa tullaan suosimaan BRR/Fay-menetelmää?
- Menetelmien yleinen idea:
 - *Entä jos* otoksemme olisikin ollut hieman erilainen? Kuinka paljon tuloksemme olisivat muuttuneet?
 - Kokeillaan erilaisilla otoksilla, jotka poimittu annetulla otanta-asetelmalla

=> Saadaan otantavaihtelun estimaatti => asetelmaperusteinen päättely

Jackknifen perusidea

- Vastataan ”*entä jos*”-kysymykseen generoimalla uusia otosaineistoja alkuperäisestä datasta
- Poistetaan vuorotellen, tiettyä systematiikkaa seuraten, alkuperäisestä datasta havaintoja => saadaan **joukko uusia, hieman toisistaan poikkeavia replikaattidatoja**, jotka koostuvat jo kerran (annetulla otanta-asetelmalla) kerätyistä havainnoista
- Kukin replikaattidata jäljittelee aineistoa, joka **olisi voitu saada** poimimalla perusjoukosta otos annetulla otanta-asetelmalla
- Jokaisesta replikaattidatasta lasketaan sama tunnusluku tai analyysi => **saatujen tulosten variointi tuottaa asetelmaa noudattavan estimaatin otantavaihtelulle** => keskivirheet => tilastolliset päätelmät
- Huomaa: ei jakaumaoletusta, hyödynnetään ainoastaan kerättyä dataa ja otanta-asetelman ominaisuuksia!
- Sovellettavissa kaikenlaisille analyyseille / tunnusluvuille

Jackknifen perusidea

Eri versioita:

- "Delete-one jackknife, Random jackknife" (JK1)
 - Yksinkertaiset otanta-asetelmat (ei ositusta, ei ryvästystä)
 - Poistetaan aineistosta kukin havainto (tai satunnainen havaintojoukko) vuoron perään => suoritetaan haluttu analyysi kaikilla näin muodostetuilla replikaattiaineistoilla => **analyysitulosten vaihtelu toimii otantavaihtelun estimaattina => keskivirheiden estimaatit**
- "Paired jackknife" (JK2)
 - Mutkikkaammat otanta-asetelmat
 - Poistetaan "havaintoryypäitä" ositteiden sisällä vuorotellen => replikaattiaineistot => jne

Esimerkki: JK2 PIAAC-aineistossa

- Jaetaan otosaineisto 80 likimain yhtä suureen "varianssiositteeseen" (**VARSTRAT**) ja nämä jaetaan kahteen likimain yhtä suureen "varianssiyksikköön" (**VARUNIT**)
- Jako varianssiositteisiin tapahtuu otannassa käytettyjen ositteiden sisällä => otanta-asetelmaan kuuluva ositus voimassa, myös mahdolliset klusterit säilytetään varianssiyksikköjä muodostettaessa
- Yhdestä 80 varianssiositteesta poistetaan satunnaisesti jompikumpi varianssiyksikkö, ja jäljelle jääneen yksikön otantapainot tuplataan
- Muut varianssiositteet ennallaan
- Tämä suoritetaan vuoron perään jokaiselle varianssiositteelle => **80 replikaattidataa, jotka eroavat otantapainoiltaan!**
- **Jackknife-varianssi saadaan replikaattidatoista laskettujen tunnuslukujen vaihteluna alkuperäisillä otantapainoilla (**SPFWT0**) kokonaisaineistosta lasketun tunnusluvun ympärillä**
- **Tarvittavat 80 painovektorit valmiina datassa!** (muuttujat **SPFWT1-SPFWT80**)

Esimerkki: JK2 PIAAC-aineistossa

Osite	Varianssi- osite	Varianssi- yksikkö	Repl 1	Repl 2	Repl 3	Repl 4	Repl 5
1	1	1	0	1	1	1	1
1	1	2	2	1	1	1	1
1	2	1	1	2	1	1	1
1	2	2	1	0	1	1	1
1	3	1	1	1	0	1	1
1	3	2	1	1	2	1	1
2	4	1	1	1	1	0	1
2	4	2	1	1	1	2	1
2	5	1	1	1	1	1	0
2	5	2	1	1	1	1	2

Esimerkki: BRR/Fay PIAAC-aineistossa

Osite	Varianssi- osite	Varianssi- yksikkö	Repl 1	Repl 2	Repl 3	Repl 4	Repl 5
1	1	1	1,5	0,5	0,5	1,5	0,5
1	1	2	0,5	1,5	1,5	0,5	1,5
1	2	1	1,5	1,5	0,5	0,5	1,5
1	2	2	0,5	0,5	1,5	1,5	0,5
1	3	1	1,5	0,5	1,5	0,5	0,5
1	3	2	0,5	1,5	0,5	1,5	1,5
2	4	1	1,5	1,5	0,5	1,5	0,5
2	4	2	0,5	0,5	1,5	0,5	1,5
2	5	1	1,5	1,5	1,5	0,5	1,5
2	5	2	0,5	0,5	0,5	1,5	0,5

Työkaluja PIAAC-datan analysointiin

IDB Analyzer, käyttää SPSS-, SAS- ja R-ohjelmistoja

- Valikkopohjainen, valmiita ”helppokäyttötoimintoja”
- Soveltuu kaikkien tunnetuimpien kansainvälisten arviointiaineistojen analyysiin: PISA, TIMSS, PIRLS, PIAAC, ICCS,...
- Hakee annetun datan PV-muuttujat ja otanta-asetelmamuuttujat (otantapainot, replikaattipainot,...) automaattisesti, käyttäjän ei tarvitse määritellä niitä
- Ladattavissa osoitteista:
 - <https://www.oecd.org/pisa/data/>
 - <https://www.iea.nl/index.php/data-tools/tools>

Työkaluja PIAAC-datan analysointiin

Data & Tools > Tools

عربي

ES

FR

JOIN US

CONTACT

Search



STUDIES

RESEARCH SERVICES

PUBLICATIONS

DATA & TOOLS

NEWS & EVENTS

ABOUT US

Tools



IEA IDB Analyzer

User Guides

ILSA Gateway

NCES Data Explorer

IEA Studies and Data Protection

IEA IDB Analyzer

The IEA International Database Analyzer (IDB Analyzer) is a free software tool which can be used to combine and analyze data from all IEA's large-scale assessments. It can also analyze data from most major large-scale assessments, including those from the US National Assessment of Educational Progress (NAEP), and those conducted by the Organisation for Economic Co-operation and Development (OECD). The IDB Analyzer includes a built-in integrated help file which serves as a technical manual for users. Additionally, all IEA studies publication on the Data Repository includes respective user guides that will guide users when working with data.

The IEA IDB Analyzer will work on most computers using the most recent versions of the Microsoft Windows operating system. You will also need SPSS, SAS, or R (Version 4.2.0 or above) and Posit (formerly known as RStudio) installed on your machine to execute the code created by the IEA IDB Analyzer.

The IEA IDB Analyzer installation requires the .NET Framework version NET 4.7.2 or above. If you intend to install the IDB Analyzer on a machine with Windows 8 or earlier, please make sure you have installed the .NET Framework on your system. The .NET Framework can be obtained from the following website: <https://dotnet.microsoft.com/en-us/download/dotnet-framework/net472>

To download the IEA IDB Analyzer, [click here](#).

Access [video tutorials](#) on how to use the IEA IDB Analyzer.

Download the latest manual detailing updates, and a step-by-step guide to using the IDB Analyzer.

If you have questions or feedback, please contact idb-analyzer@iea-hamburg.de

User Guides

User Guides for IEA studies contain information on the structure and variables within the data and how to conduct analyses using the IEA IDB Analyzer. Please see our publications database for a full overview of available user guides.

Työkaluja PIAAC-datan analysointiin

- **SAS-makro** 'PIAAC_Tool' Cycle 1 -datan analysointiin, ladattavissa OECD:n PIAAC-sivustolta
- **Stata**, ja erityisesti 'Repest'-moduli (ei henk.koht. kokemusta)
- **R** (ei henk.koht. kokemusta)
- **Mplus**
 - Runsaasti mahdollisuuksia asetelman huomioivaan analyysiin (painot, klusterointi,...)
 - Plausible value -analyysi mahdollista (TYPE=IMPUTATION)

Mitä työkaluilla voidaan tehdä?

- IDB Analyzerissa ja SAS-makrossa käytännössä sama analyysivalikoima
- Asetelmaperusteinen varianssien estimointi (ainakin Jackknife)
- Plausible value -analyysi

Tilastollisten menetelmien valikoima suppea, kattaa vain tavallisimmat perusmenetelmät

- Keskiarvot, persentiilit
- Frekvenssi- ja prosenttijakaumat, ristiintaulukot
- Lineaarinen regressio
- Logistinen regressio

Mitä työkalut tekevät?

- Hyödyntävät **SAS / SPSS / R** -ohjelmistojen valmiita proseduureja (asetelmaperusteiset menetelmät, ”complex survey statistics”)
- Käytännössä työkalut ovat valmisproseduurien ympärille ohjelmoituja silmukoita (loops), joiden sisällä ajetaan halutut proseduurit
- Lopuksi yhdistetään silmukoissa saadut laskentatulokset lopulliseksi tuloksiksi

Mitä työkalut tekevät?

1) Uloin silmukka plausible value -muuttujien PV1-PV10 yli (jos kyseessä PV-analyysi)

2) Kunkin PV:n alla ajetaan sisempi silmukka 80 replikaattipainomuuttujan yli (Jackknife)

(Ulompi + sisempi silmukka: kaikkiaan ajetaan $10 * 80$ proseduurikutsua => 800 estimointitulosta)

3) Sisemmässä silmukassa (siis yksi PV kerrallaan) lasketaan halutun tunnusluvun piste-estimaatti sekä alkuperäisillä otantapainoilla (SPFWT0) että 80 replikaattipainolla (painot SPFWT1-SPFWT80).

4) Tunnusluvulle lasketaan Jackknife-varianssi: 80 replikaattituloksen vaihtelu alkuperäispiste-estimaatin ympärillä.

Ulompi + sisempi silmukka => 10 PV-piste-estimaattia ja kullekin Jackknife-varianssi

Mitä työkalut tekevät?

Lopuksi kootaan yhteen 10 eri PV:lle edellä saadut tulokset (10 piste-estimaattia ja niiden Jackknife-variانسsit)

5) Lasketaan 10 PV-piste-estimaatin (PV1-PV10) keskiarvo

=> **lopullinen piste-estimaatti**

6) Lasketaan 10 Jackknife-variانسsin (PV1-PV10) keskiarvo ja lisätään siihen imputaatiovariانسsi (= 10 PV-piste-estimaatin välinen variانسsi)

=> **lopullinen variانسsiestimaatti**

=> **piste-estimaatin keskivirhe tämän neliöjuurena**

Jos kyseessä ei PV-analyysi, vaiheet 5 ja 6 jäävät pois, ts. toteutetaan vain yksi 80 replikaation silmukka.

Kiitos!

Kari Nissinen
Koulutuksen tutkimuslaitos
Jyväskylän yliopisto
kari.nissinen@jyu.fi

