

Het
opleidingsniveaubestand



Belang opleidingsniveau in statistiek en onderzoek

- Arbeidsmarktparticipatie
- Inkomen
- Stemgedrag
- Crimineel gedrag
- Gezondheid
- Uitkeringsafhankelijkheid

CBS: twee bronnen voor opleidingsniveau

- Enquête Beroepsbevolking: 100.000 waarnemingen
- Opleidingsniveaubestand: 13.000.000 waarnemingen

Doel opleidingsniveaubestand: maximale dekkinggraad

Homerus: “een onsterfelijke creatie, niet menselijk, van voren een leeuw, van achteren een slang, en een geit in het midden, vuurspuwend en al”.

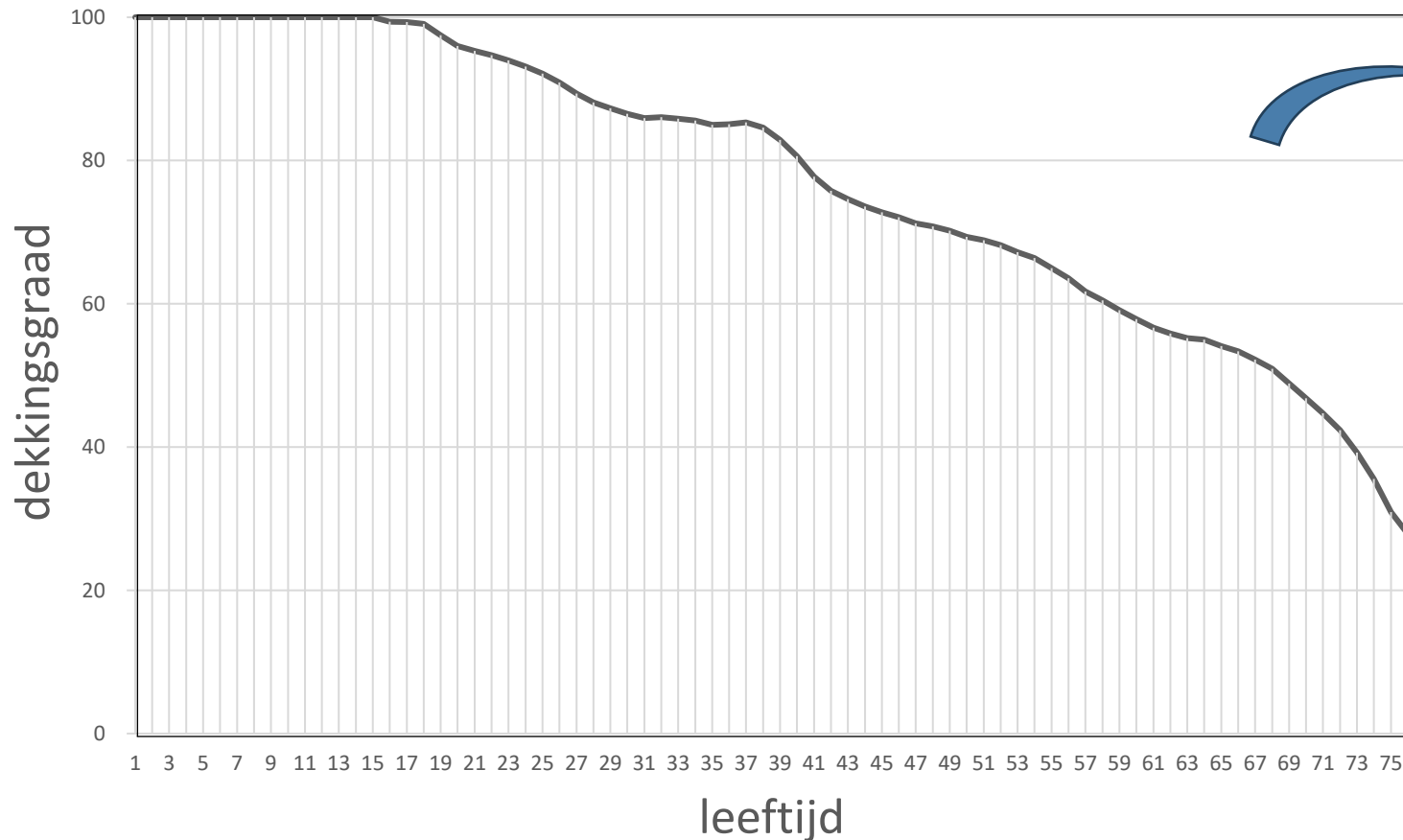


Bronnen en dekingsgraad

Bron	Aantal
OCW (inclusief WSF)	9.000.000
UWV	2.500.000
BRP	500.000
EBB (2004-2024)	1.000.000

Hoe combineren we registerdata en steekproefdata over opleidingsniveau op zo'n manier dat de gegevens representatief zijn voor de hele populatie?

Dekkingsgraad naar leeftijd (2024)



Personen die:

- hun opleidingen hebben gevolgd voordat de registraties tot stand kwamen
- hun opleidingen in het buitenland hebben gevolgd

Weging: het principe



Weegmodel

≥ 35 jaar:

- leeftijd x geslacht +
- Inkomen +
- sociaal-economische categorie +
- type huishouden +
- gemeente +
- migratiemotief +
- al dan niet in het buitenland geboren x herkomstregio

< 35 jaar:

- leeftijd +
- geslacht x migratiemotief

Institutionele bevolking:

- leeftijd x geslacht

Resultaat: 'aparte' verdeling gewichten

persoonsidentificatie	gewicht
1	1.00000
2	1.00000
3	1.00000
4	6.28373
5	1.00000
8	1.00000
9	1.00000
10	12.9848
11	1.00000
12	1.00000
14	1.00000

$$\hat{\theta}(y^{(c)}) = \underbrace{\sum_{i \in U_R} I\{y_i = y^{(c)}\}}_{\text{telling}} + \underbrace{\sum_{i \in S_{NR}} w_i I\{y_i = y^{(c)}\}}_{\text{schatting}}$$

niet waargenomen onderwijs

Onderwijs in het buitenland of aan een particuliere onderwijsinstelling

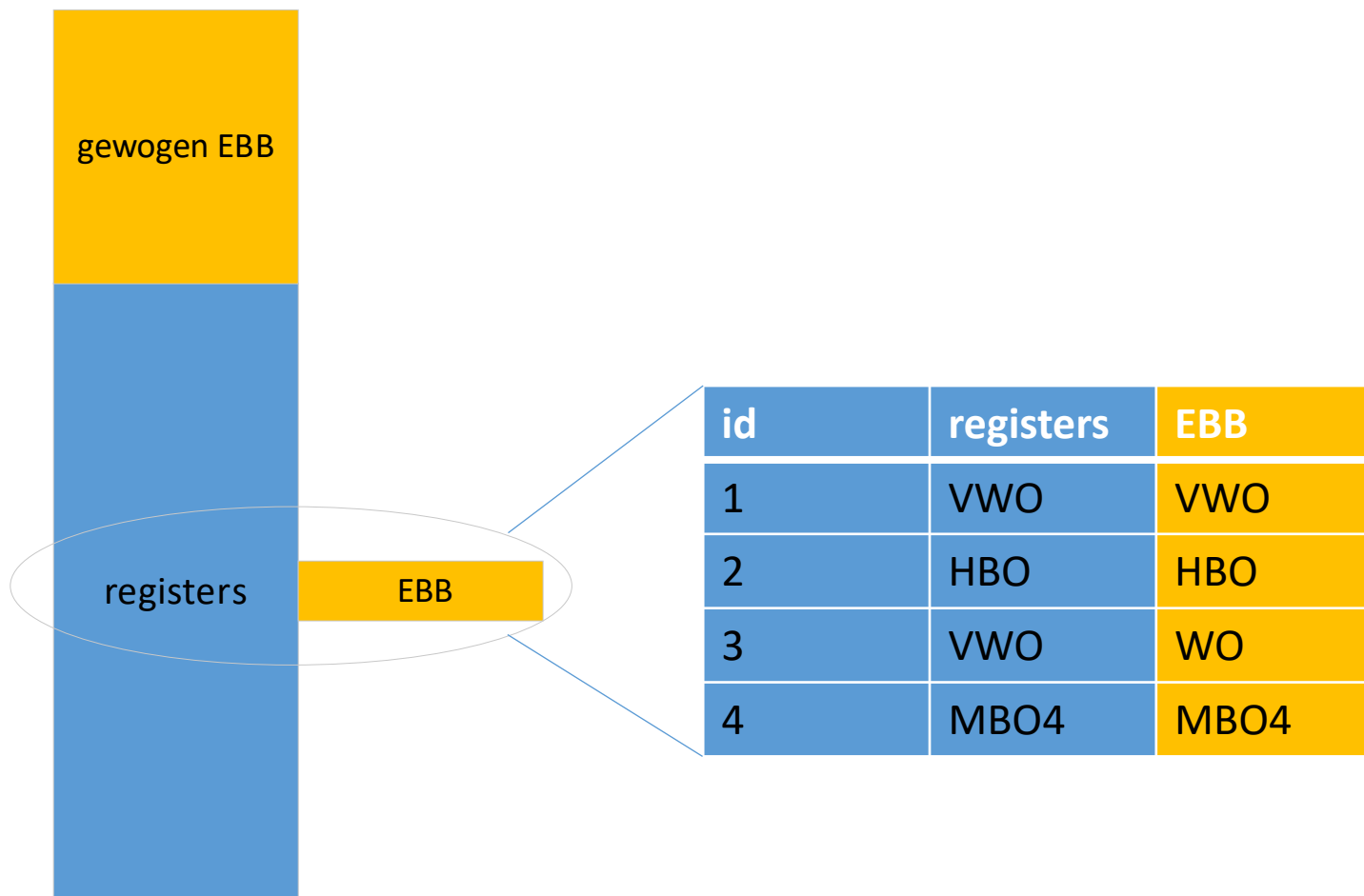
persoonsidentificatie	opleiding
9	BO
9	HAVO
9	VWO
9	WO

werkelijkheid

persoonsidentificatie	opleiding
9	BO
9	HAVO
9	VWO

waargenomen

Confrontatie met EBB-gegevens



Modelschatting en imputatie

id	x	y	z	p	q	r	registers	EBB
1	4	3	3	1	5	5	VWO	VWO
2	3	6	3	2	1	3	HBO	HBO
3	2	5	1	1	3	5	VWO	WO
4	6	3	2	1	4	2	MBO4	MBO4



$$P(O_j) = f(O_i, x, y, z, p, q, r); O_j > O_i$$



imputatie hoger niveau in registerpopulatie

Nadelen imputatie

- Niveau in 18 categorieën
- Geen echte waarnemingen
- Risico op vertekening
- Longitudinale inconsistenties

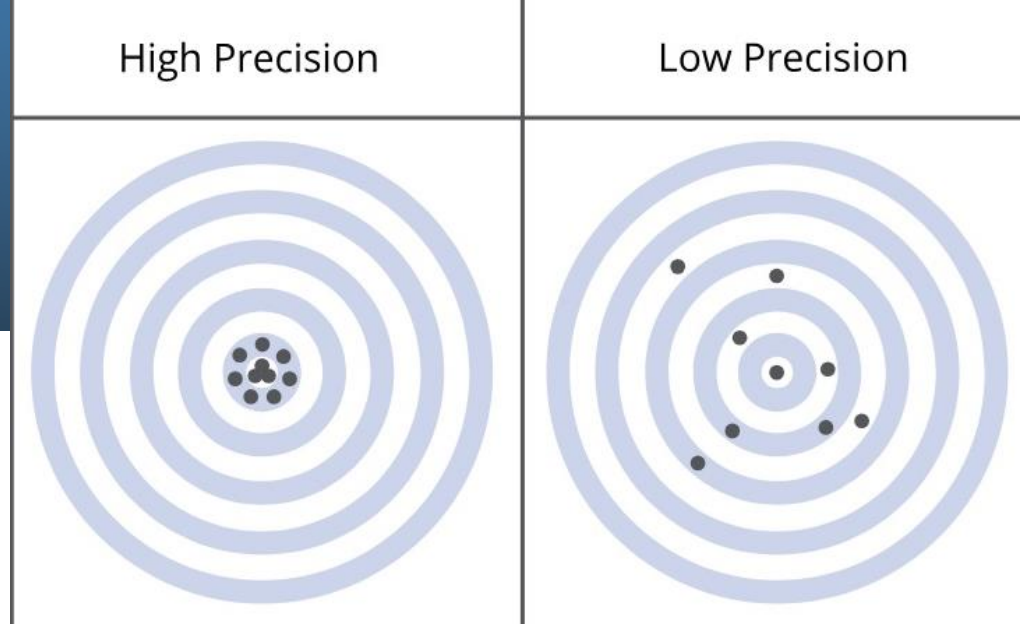


Precisie → bootstraptool

Deels steekproefwaarneming → steekproeffout

Precisie van puntschattingen kan worden bepaald met behulp van bootstraptool

- 500 steekproeven trekken
- De weging wordt 500 keer uitgevoerd → 500 extra kolommen met gewichten
- De tool schat een tabel 500 keer, één keer met elk gewicht → 95% BI
- De tool herschaalt de gewichten bovendien naar de achtergrondkenmerken



500 extra kolommen met gewichten

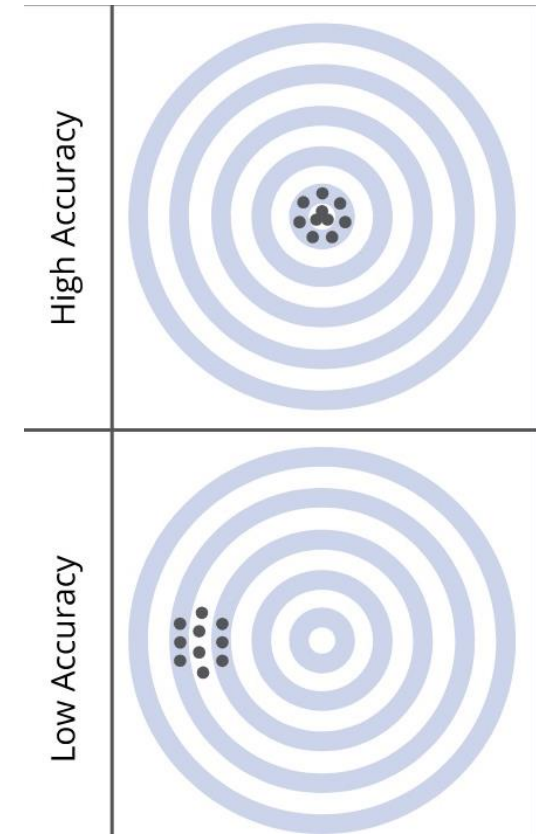
persoonsidentificatie	gewicht
1	1.00000
2	1.00000
3	1.00000
4	6.28373
5	1.00000
8	1.00000
9	1.00000
10	12.9848
11	1.00000
12	1.00000
14	1.00000

+

gewicht_1	gewicht_2	...	gewicht_499	gewicht_500
1.00000	1.00000		1.00000	1.00000
1.00000	1.00000		1.00000	1.00000
1.00000	1.00000		1.00000	1.00000
8.39572	3.50968		9.23498	5.23957
1.00000	1.00000		1.00000	1.00000
1.00000	1.00000		1.00000	1.00000
1.00000	1.00000		1.00000	1.00000
5.39574	9.21383		10.2398	13.1233
1.00000	1.00000		1.00000	1.00000
1.00000	1.00000		1.00000	1.00000
1.00000	1.00000		1.00000	1.00000

Nauwkeurigheid

- Herontwerp EBB onderwijsvraagstelling in 2013 → werkt asymptotisch door in opleidingsniveaubestand
- Selectiviteit van de EBB-respons
- Kwaliteit UWV-registratie



Waarneming migranten

leeftijdsklasse	geboren in NL (%)	niet geboren in NL (%)
15-24	99,8	74,2
25-34	99,6	47,3
35-44	87,1	52,8
45-54	69,3	61,7
55-64	55,5	62,6
65-74	41,0	51,4

bron	geboren in NL (%)	niet geboren in NL (%)
OCW	71,3	53,9
EBB	4,8	3,1
UWV	17,7	40,7
BRP	6,2	2,2

Kolommen in het opleidingsniveaubestand

- RINPERSOONS/RINPERSOON → koppelsleutel
- GEWICHTHOOGSTEOPL → altijd gebruiken
- OPLNRHB → zonder imputatie
- OPLNIVSOI2021AGG4HBmetNIRWO → met imputatie