



Technische documentatie HO instellingsprognose

1. Data selectie

De prognose wordt in eerste instantie gebaseerd om de referentieraming. De referentieraming geeft het verwachte aantal studenten in het bekostigd onderwijs in Nederland weer voor de komende jaren. Deze is uitgesplitst naar:

- Opleidingsvorm
- Soort HO
- Type HO
- Internationale herkomst (Nederlands, EER of niet-EER)

De raming kent ook uitsplitsingen naar leeftijd en richting (vergelijkbaar met CROHO onderdeel). Op deze uitsplitsingsniveaus wordt niet gestuurd in de ramingsproces. Ze worden daarom niet gebruikt voor de HO instellingsprognose.

Als tweede data bron is het 1-Cijferbestand van de HO inschrijvingen gebruikt. Deze is uitgesplitst naar:

- Gemeente van de student
- herkomst (Nederlands, EER of niet-EER)
- Leeftijd
- Instelling
- Opleidingsvorm
- CROHO onderdeel
- Hoogste vooropleiding
- Soort HO
- Type HO

Hier geeft gemeente de woongemeente van laatst bekende vooropleiding aan. Als deze niet beschikbaar is wordt de woongemeente ten tijde van HO inschrijving of de gemeente van de ho vestiging gebruikt.



Als laatste is de historische bevolkingstelling van CBS en de bevolkingsprognose, PRIMOS, van ABF research. Deze zijn uitgesplitst naar:

- Gemeente
- Leeftijd
- herkomst (Nederlands, EER of niet-EER)

De PRIMOS is opgesplitst naar postcode4, wat door middel van een tabel omgezet wordt naar gemeente.

Data bewerking

De data uit het 1-Cijferbestand is voorbereid om aan te sluiten bij de definities gebruikt door de Referentieraming. Zo zijn instromers gedefinieerd als studenten die in het vorige jaar in een andere combinatie van soort HO, type HO, richting en opleidingsvorm hun hoofdschrijving hadden, of geen HO hoofdschrijving hadden. Tweedejaars studenten staan voor precies twee achtereenvolgende jaren ingeschreven in een combinatie van soort HO, type HO, richting en opleidingsvorm, hogerejaars voor drie of meer achtereenvolgende jaren.

De leeftijden van het 1-Cijferbestand en de bevolkingsgegevens wordt gegroepeerd naar een drietal categorieën:

- 17 en onder
- 18 t/m 25
- Boven de 25

Voor de opleidingsvorm gebruiken we de opties voltijd of deeltijd. Duaal wordt omgecodeerd naar voltijd.

Verder wordt de vooropleiding ook in een aantal categorieën opgesplitst:

- Havo
- Vwo
- Mbo
- Hbo
- Wo
- Overig



Drie losse sets aan gegevens worden gemaakt, één voor instroom, één voor tweedejaars, en één voor hogerejaars. Hiermee kan hetzelfde prognose model gebruikt worden om instroom, tweedejaars en hogerejaars te voorspellen.

Als laatste worden voor de historische gegevens de gemeentes gefuseerd waar nodig op basis van de CBS gegevens over fusies van gemeente.

2. Prognose model

Aan de hand van de hiervoor gedefinieerde data kunnen we de prognose maken. Deze prognose heeft een aantal stappen. Als eerste splitsen we de studenten uit de referentieraming op naar de verschillende gemeentes. Met deze opsplitsing van studenten per gemeente kan het marktaandeel van de instellingen per gemeente gebruikt worden om deze studenten te verdelen over de verschillende instellingen.

3. Populatie aandeel

Om het aantal HO studenten per gemeente te vinden, berekenen we eerst het aandeel van de bevolking dat HO student is (A^{HO}):

$$A^{HO}_{g,l,c,v,t,s,o,h} = \frac{L^{HO}_{g,l,c,v,t,s,o,h}}{P^{CBS}_{g,l,h}}$$

Met uitsplits variabelen g de gemeente, l de leeftijd categorie, c het CROHO onderdeel, v de hoogste vooropleiding, t & s de type en soort HO opleiding, o de opleidingsvorm en h de herkomst. Het aantal HO studenten L^{HO} ('Leerlingen Hoger Onderwijs') is berekend door het nemen van een gemiddelde over de afgelopen drie jaren:

$$L^{HO}_{g,l,c,v,t,s,o,h} = \sum_j \frac{1}{n_j} \sum_i L^{HO}_{j,i,g,l,c,v,t,s,o,h}$$

Met j het jaartal en $n_j = 3$ het aantal jaren dat terug gekeken wordt en i de instelling. Vergelijkbaar wordt de bevolkingstelling P^{CBS} ('Prognoses CBS') ook door een gemiddelde over de afgelopen drie jaren berekend:

$$P^{CBS}_{g,l,h} = \sum_j \frac{1}{n_j} P^{CBS}_{j,g,l,h}$$



Aan de hand van het aandeel HO studenten in de bevolking kunnen we een toekomst voorspelling maken over hoe de studenten zich verdelen over het land ($L^{HO,Prog}$):

$$L_{j,g,l,c,v,t,s,o,h}^{HO,Prog} = A_{g,l,c,v,t,s,o,h}^{HO} \times P_{j,g,l,h}^{PRIMOS}$$

Dit is berekend aan de hand van de bevolkingsprognose van ABF research P^{PRIMOS} . Nu hebben we een regionale prognose van het aantal HO studenten.

Deze aantal studenten komt echter niet overeen met de voorspelling van de referentieraming. Daarom is het belangrijk om deze te ijken op basis van de referentieraming:

$$L_{j,g,l,c,v,t,s,o,h}^{HO,Prog,RR} = \frac{L_{j,g,l,c,v,t,s,o,h}^{HO,Prog}}{L_{j,t,s,o,h}^{HO,Prog}} RR_{j,t,s,o,h}^{HO,Prog}$$

Zo komen we uit op een regionale prognose ($L^{HO,Prog,RR}$) op basis van de referentieraming ($RR^{HO,Prog}$).

4. Marktaandeel

Deze studenten zijn nu in alle vormen opgesplitst, maar nog niet verdeeld over de verschillende instellingen. Door gebruik van het marktaandeel (M) van iedere instelling (i) kunnen we deze verdeling maken:

$$M_{i,g,l,c,v,t,s,o,h} = \frac{L_{i,g,l,c,v,t,s,o,h}^{HO}}{L_{g,l,c,v,t,s,o,h}^{HO}}$$

De studenten aantallen die hier voor gebruikt worden zijn gemiddeld over de afgelopen twee jaren:

$$L_{i,g,l,c,v,t,s,o,h}^{HO} = \sum_j \frac{1}{n_j} L_{j,i,g,l,c,v,t,s,o,h}^{HO}$$

In dit geval is $n_j = 2$.

Deze marktaandelen kunnen nu gebruikt worden om de prognose verder over de instellingen op te splitsen:

$$L_{j,i,g,l,c,v,t,s,o,h}^{HO,Prog,RR} = L_{j,g,l,c,v,t,s,o,h}^{HO,Prog,RR} \times M_{i,g,l,c,v,t,s,o,h}$$

5. Ijken aan referentieraming



Nu hebben we een instellingsprognose op basis van de referentieraming, maar deze telt niet noodzakelijk op tot de referentieraming. De reden hiervoor is dat er groepen in de opsplitsing van de marktaandeelen die nul zijn in de regionale prognose. Dan tellen de marktaandeelen per instelling die meegenomen worden niet meer op tot één. Daarom moet er een herijking plaatsvinden op basis van de referentieraming:

$$Prog_{j,i,g,l,c,v,t,HO,SHO,o,h} = \frac{RR_{j,t,s,o,h}^{HO,Prog}}{L_{j,t,s,o,h}^{HO,Prog,RR}} L_{j,i,g,l,c,v,t,s,o,h}^{HO,Prog,RR}$$

6. Herberekening van instroomdefinitie

De prognose is nu gebaseerd op de instroom definitie van de referentieraming. In de referentieraming wordt niet gekeken naar de eerdere geschiedenis van een student, maar simpelweg naar of ze in het vorige jaar ook in dezelfde onderwijscategorie hun hoofdschrijving hadden of niet. Onderwijscategorie in het ho context komt overeen met de combinatie soort en type ho, CROHO onderdeel en opleidingsvorm, waarbij CROHO onderdeel *onderwijs* onderverdeeld wordt in lerarenopleidingen voor het basisonderwijs en lerarenopleidingen voor het voortgezet onderwijs.

Het gebruik bij HO&S, de Raden en overige onderzoek op basis van het cijferbestand HO is instroom te bepalen aan de hand van de eerdere geschiedenis van een student. Een student is dan alleen instromer als ze nooit eerder in een bepaalde domein ingeschreven zijn geweest.

Voor de prognoses zijn de verhoudingen tussen de referentieraming definitie van instroom en (1) instroom voor het eerst in het gehele ho en (2) voor het eerst in de combinatie van type en soort ho uitgewerkt. Daarmee kan een vertaling gemaakt worden tussen de ene instroom definitie en de ander.

Deze vergelijking wordt uitgevoerd op samples die getrokken worden uit de studentenpopulatie (hoofdschrijvingen in domein ho) in de jaren 2016 t/m 2023, per combinatie van instelling (Instellingscode), soort ho, type ho en herkomst (EER, niet-EER, Nederlandse). De groottes van samples varieerden tussen de 5% en 50%



van de totale populatie van een combinatie. Per sample werd het aantal studenten dat als instromer, tweedejaars en hogerejaars aangemerkt zou worden volgens de verschillende definities berekend. Over de samples heen worden vervolgens mixed-effects regressiemodellen gefit. Deze modellen worden vervolgens gebruikt om het aantal instromers, tweedejaars en hogerejaars in domein ho en in domein type binnen soort te voorspellen aan de hand van de RR-instroom, tweedejaars en hogerejaars aantallen in de prognose zelf.

Voor de instroom definitie domein hoger onderwijs is het model gespecificeerd als volgt.

InstroomDomeinHO_{ij}

$$\begin{aligned} &= \beta_{0j} + \beta_{1j}InstroomRR_{ij} + \beta_{2j}t_{HO, SHO_{ij}} + \beta_{3j}h_{ij} \\ &+ \beta_{4j}InstroomRR_{ij} \times t_{HO, SHO_{ij}} + \beta_{5j}InstroomRR_{ij} \times h_{ij} + \beta_{6j}t_{HO, SHO_{ij}} \times h_{ij} \\ &+ \beta_{7j}InstroomRR_{ij} \times t_{HO, SHO_{ij}} \times h_{ij} + e_{ij}, \end{aligned}$$

waar

$$\beta_{0j} = \gamma_{00} + \gamma_{01}i_j + \gamma_{02}t_{HO, SHO_j} + \gamma_{03}h_j + \gamma_{04}i_j \times t_{HO, SHO_j} + \gamma_{05}i_j \times h_j + \gamma_{06}t_{HO, SHO_j} \times h_j + \gamma_{07}i_j \times t_{HO, SHO_j} \times h_j + u_{0j},$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}i_j + \gamma_{12}t_{HO, SHO_j} + \gamma_{13}h_j + \gamma_{14}i_j \times t_{HO, SHO_j} + \gamma_{15}i_j \times h_j + \gamma_{16}t_{HO, SHO_j} \times h_j + \gamma_{17}i_j \times t_{HO, SHO_j} \times h_j + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}i_j + \gamma_{22}t_{HO, SHO_j} + \gamma_{23}h_j + \gamma_{24}i_j \times t_{HO, SHO_j} + \gamma_{25}i_j \times h_j + \gamma_{26}t_{HO, SHO_j} \times h_j + \gamma_{27}i_j \times t_{HO, SHO_j} \times h_j + u_{2j},$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}i_j + \gamma_{32}t_{HO, SHO_j} + \gamma_{33}h_j + \gamma_{34}i_j \times t_{HO, SHO_j} + \gamma_{35}i_j \times h_j + \gamma_{36}t_{HO, SHO_j} \times h_j + \gamma_{37}i_j \times t_{HO, SHO_j} \times h_j + u_{3j},$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41}i_j + \gamma_{42}t_{HO, SHO_j} + \gamma_{43}h_j + \gamma_{44}i_j \times t_{HO, SHO_j} + \gamma_{45}i_j \times h_j + \gamma_{46}t_{HO, SHO_j} \times h_j + \gamma_{47}i_j \times t_{HO, SHO_j} \times h_j + u_{4j},$$

$$\beta_{5j} = \gamma_{50} + \gamma_{51}i_j + \gamma_{52}t_{HO, SHO_j} + \gamma_{53}h_j + \gamma_{54}i_j \times t_{HO, SHO_j} + \gamma_{55}i_j \times h_j + \gamma_{56}t_{HO, SHO_j} \times h_j + \gamma_{57}i_j \times t_{HO, SHO_j} \times h_j + u_{5j},$$

$$\beta_{6j} = \gamma_{60} + \gamma_{61}i_j + \gamma_{62}t_{HO, SHO_j} + \gamma_{63}h_j + \gamma_{64}i_j \times t_{HO, SHO_j} + \gamma_{65}i_j \times h_j + \gamma_{66}t_{HO, SHO_j} \times h_j + \gamma_{67}i_j \times t_{HO, SHO_j} \times h_j + u_{6j}, \text{ en}$$

$$\beta_{7j} = \gamma_{70} + \gamma_{71}i_j + \gamma_{72}t_{HO, SHO_j} + \gamma_{73}h_j + \gamma_{74}i_j \times t_{HO, SHO_j} + \gamma_{75}i_j \times h_j + \gamma_{76}t_{HO, SHO_j} \times h_j + \gamma_{77}i_j \times t_{HO, SHO_j} \times h_j + u_{7j}.$$



Hier representeert de bovenste vergelijking het *fixed effects* gedeelte, waarin per sample i de Instroom aantallen voor de instroomdefinitie Domein Hoger Onderwijs voorspelt wordt op basis van de drie in indicatoren, alle mogelijke 2-weg-interacties tussen indicatoren, de 3-weg interactie tussen alle drie indicatoren en een error-term. De onderste 8 vergelijkingen representeren het *random effects* gedeelte. Hier worden de intercept en alle regressie-coëfficiënten van de fixed effects-vergelijking voorspelt op basis van de overkoepelende waardes (door alle samples heen) van de drie in indicatoren, alle mogelijke 2-weg-interacties tussen indicatoren, de 3-weg interactie tussen alle drie indicatoren en een (aparte) error-term. Subscript j representeert voor elke variabele in kwestie de gemiddelde waarde over de hele populatie (d.w.z. de combinatie van alle samples) heen.

Omdat er een lineaire schatter is gebruikt (die een normale verdeling van de error-termen verondersteld) kunnen er studentenaantallen van kleiner dan nul voorspelt zijn. Daarom zijn alle waardes die als kleiner dan nul voorspelt zijn in een aparte post-processing stap op nul gezet.

7. Verspreiding van resultaten

De eindresultaten worden op twee manieren gepresenteerd. Als eerste wordt een tabel geleverd met verschillende opsplitsing niveaus:

- Instellingscode
- Instellingscode, vestiging
- Instellingscode, soort en type HO
- Instellingscode, vestiging, soort en type HO
- Instellingscode, soort en type HO, herkomst
- Instellingscode, vestiging, soort en type HO, herkomst
- Instellingscode, soort en type HO, opleidingsvorm
- Instellingscode, vestiging, soort en type HO, opleidingsvorm
- Instellingscode, soort en type HO, opleidingsvorm, herkomst
- Instellingscode, vestiging, soort en type HO, opleidingsvorm, herkomst

Verschillende tabellen zijn nodig omdat aantallen kleiner dan 5 in de realisatie niet geleverd mogen worden in het kader van AVG. Bij de kleine instellingen komen er snel kleine aantallen voor.



Herkomst is hier gedefinieerd als Nederlands, EER of niet-EER.

Als laatste zijn ook grafieken met ingeschrevenen per Instellingscode en instromers per Instellingscode, type en soort geleverd.