



# Instellingsprognoses HO technische documentatie

## 1 Data selectie

De prognose wordt in eerste instantie gebaseerd op 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.

### 1.1 Data bewerking

De data uit het 1-Cijferbestand is voorberekt 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 zitten in hun tweede verblijfsjaar en zijn beide jaren ingeschreven in een combinatie van soort HO, type HO, richting en opleidingsvorm. De rest van de studenten zijn geclassificeerd als hogerejaars. Dit includeert dus en studenten die meer dan twee achtereenvolgende jaren in dezelfde combinatie zitten en studenten die tussendoor uit- en dan weer ingestroomd zijn. 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 ( $A^{\text{HO}}$ ):

$$A_{g,l,c,v,t,s,o,h}^{\text{HO}} = \frac{L_{g,l,c,v,t,s,o,h}^{\text{HO}}}{P_{g,l,h}^{\text{CBS}}}$$

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  $L^{\text{HO}}$  ('Leerlingen Hoger Onderwijs') is berekend door het nemen van een gemiddelde over de afgelopen drie jaren:

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

Met  $j$  het jaartal en  $n_j = 3$  het aantal jaren dat terug gekeken wordt en  $i$  de instelling.

Vergelijkbaar wordt de bevolkingstelling  $P^{\text{CBS}}$  ('Prognoses CBS') ook door een gemiddelde over de afgelopen drie jaren berekend:

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

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 ( $L^{\text{HO,Prog}}$ ):

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

Dit is berekend aan de hand van de bevolkingsprognose van ABF research  $P^{\text{PRIMOS}}$ . 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:



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

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

## 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:

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

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

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

In dit geval is  $n_j = 2$ .

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

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

## 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 marktaandelen die nul zijn in de regionale prognose. Dan tellen de marktaandelen per instelling die meegenomen worden niet meer op tot één. Daarom moet er een herijking plaatsvinden op basis van de referentieraming:

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

## 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 1cijferbestand HO is instroom te bepalen aan de hand van de eerdere geschiedenis van een student. Studenten zijn dan alleen instromers 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:

$$V_{i,s,t,o,h}^{k \rightarrow l} = \frac{L_{i,s,t,o,h}^{k,l}}{L_{i,s,t,o,h}^k}$$

$$L_{j,i,s,t,o,h}^{l, \text{Prog}} = \sum_k V_{i,s,t,o,h}^{k \rightarrow l} \times L_{j,i,s,t,o,h}^{k, \text{Prog}}$$

**Toelichting:** De vertaling wordt uitgevoerd door in totaal 18 verdeelsleutels te berekenen. Voor elke verstaalslag van ( $k$ ) het aantal studenten binnen een studiejaar volgens de referentieraming definitie (instroom RR, tweedejaars RR, hogerejaars RR) naar ( $l$ ) een combinatie van studiejaar en instroom definitie in het eencijfer bestand ho (instroom, tweedejaars en hogerejaars in het gehele ho; instroom, tweedejaars en hogerejaars in hetzelfde type binnen soort ho) wordt er zo'n verdeelsleutel berekend. Al deze berekeningen worden wederom uitgevoerd voor elke combinatie van instelling ( $i$ ), soort en type ho ( $s, t$ ), opleidingsvorm ( $o$ ) en herkomst ( $h$ ) en gemiddeld over de afgelopen drie jaar.

Elke verdeelsleutel wordt bepaald door het aandeel van studenten te berekenen dat als instromer, tweedejaars of hogerejaars telt en volgens de referentieraming definitie én volgens een van de twee andere definities ( $L_{i,s,t,o,h}^{k,l}$ ) onder het totale aantal studenten binnen een studiejaar volgens de referentieraming definitie ( $L_{i,s,t,o,h}^k$ ). Vervolgens kan per prognose jaar  $j$  het voorspelde aantal studenten binnen een studiejaar volgens de referentieraming definitie  $L_{j,i,s,t,o,h}^{k, \text{Prog}}$  vermenigvuldigd worden met de verdeelsleutel  $V_{i,s,t,o,h}^{k \rightarrow l}$ . Om het gewenste aantal instromers, tweedejaars of hogerejaars voor een bepaalde definitie  $L_{j,i,s,t,o,h}^{l, \text{Prog}}$  te bepalen wordt over de drie studiejaars volgens de referentieraming definitie  $k$  gesommeerd.

Bijvoorbeeld, om voor prognosejaar  $j$  het aantal tweedejaars studenten binnen het gehele ho  $L_{j,i,s,t,o,h}^{\text{tweedejaars domein ho, Prog}}$  te bepalen worden er drie verdeelsleutel berekend:

$$V_{i,s,t,o,h}^{\text{instroom RR} \rightarrow \text{tweedejaars domein ho}} = \frac{L_{i,s,t,o,h}^{\text{instroom RR, tweedejaars domein ho}}}{L_{i,s,t,o,h}^{\text{instroom RR}}}$$

$$V_{i,s,t,o,h}^{\text{tweedejaars RR} \rightarrow \text{tweedejaars domein ho}} = \frac{L_{i,s,t,o,h}^{\text{tweedejaars RR, tweedejaars domein ho}}}{L_{i,s,t,o,h}^{\text{tweedejaars RR}}}$$

$$V_{i,s,t,o,h}^{\text{hogerejaars RR} \rightarrow \text{tweedejaars domein ho}} = \frac{L_{i,s,t,o,h}^{\text{hogerejaars RR, tweedejaars domein ho}}}{L_{i,s,t,o,h}^{\text{hogerejaars RR}}}$$

Het gezochte aantal  $L_{j,i,s,t,o,h}^{\text{tweedejaars domein ho, Prog}}$  kan nu bepaald worden o.b.v. van het voorspelde aantal instroomers ( $L_{j,i,s,t,o,h}^{\text{instroom RR, Prog}}$ ), tweedejaars ( $L_{j,i,s,t,o,h}^{\text{tweedejaars RR, Prog}}$ ) en hogerejaars ( $L_{j,i,s,t,o,h}^{\text{hogerejaars RR, Prog}}$ ) volgens de referentieraming definitie, door de som van de producten van de voorspelde aantallen met de bijhorende verdeelsleutel te berekenen:

$$L_{j,i,s,t,o,h}^{\text{tweedejaars domein ho, Prog}} =$$

$$V_{i,s,t,o,h}^{\text{instroom RR} \rightarrow \text{tweedejaars domein ho}} \times L_{j,i,s,t,o,h}^{\text{instroom RR, Prog}} +$$

$$V_{i,s,t,o,h}^{\text{tweedejaars RR} \rightarrow \text{tweedejaars domein ho}} \times L_{j,i,s,t,o,h}^{\text{tweedejaars RR, Prog}} +$$

$$V_{i,s,t,o,h}^{\text{hogerejaars RR} \rightarrow \text{tweedejaars domein ho}} \times L_{j,i,s,t,o,h}^{\text{hogerejaars RR, Prog}}$$



## 7

# Verspreiding van resultaten

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

- Instelling
- Instelling, vestiging
- Instelling, herkomst
- Instelling, vestiging, herkomst
- Instelling, soort en type HO
- Instelling, vestiging, soort en type HO
- Instelling, soort en type HO, herkomst
- Instelling, vestiging, soort en type HO, herkomst
- Instelling, soort en type HO, opleidingsvorm
- Instelling, vestiging, soort en type HO, opleidingsvorm
- Instelling, soort en type HO, opleidingsvorm, herkomst
- Instelling, 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.

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