

Landelijke databestand HBO-Monitor

Handleiding bij het gebruik van het landelijke databestand dat aan hogescholen beschikbaar wordt gesteld in het kader van de HBO-Monitor.

1. Opbouw bestand
2. Soorten variabelen
3. Enkele kernvariabelen per onderwerp
4. Weging
5. Aandachtspunten

Dit document is eveneens van toepassing op het meerjaren bestand dat aan hogescholen beschikbaar wordt gesteld. In het meerjarenbestand van een hogeschool zitten alleen de respondenten van de hogeschool zelf. Variabelen als de weegfactor (*wf*) en instelling versus rest (*InstellingRest*) waarnaar in dit document verwezen wordt zitten niet in dit bestand.

1 Opbouw bestand

Het bestand volgt wat opbouw (volgorde van de variabelen) betreft de HBO-Monitor vragenlijst. Aan het begin staan enkele algemene en administratieve variabelen, zoals bijvoorbeeld de instellingsvariabele *InstellingRest*. Aan het einde van het bestand de instelling specifieke vragen uit eventuele inlegvellen. Het bestand kent dan ook de volgende opbouw:

- algemene variabelen
- gevolgde opleiding
- vooropleiding
- na de HBO opleiding
- huidige functie
- oordelen van afgestudeerden
- slotvragen / eigen vragen instelling

2 Soorten variabelen

Bij het werken met het bestand en het interpreteren van eventuele output is het van belang te weten dat er verschillende soorten variabelen zijn:

Tekst- en numerieke variabelen: de meeste variabelen zijn numerieke variabelen (bijvoorbeeld aanduiding F2.0 in SPSS) en hebben betrekking op getallen (bijv. leeftijd) of vragen met gesloten antwoordcategorieën. Tekstvariabelen (bijvoorbeeld aanduiding A100 in SPSS) zijn tekstvelden die door respondenten ingevoerd zijn, zoals bijvoorbeeld de naam van het bedrijf waar men werkt. Deze variabelen

lenen zich minder voor beschrijvende statistieken zoals een frequentieverdeling, maar zijn vaak van toelichtende waarde voor andere (numerieke) variabelen.

Administratieve variabelen: deze zijn niet direct uit de vragenlijst afkomstig. De variabelen zijn nodig om rapportages in te richten en het veldwerk gestructureerd en controleerbaar te laten verlopen zoals bijvoorbeeld de variabele *koppel_id* (instelling specifiek koppelnummer t.b.v. eigen administraties van hogescholen) en *vinput* (methode van enquêteren).

Vragenlijstvariabelen: variabelen die rechtstreeks uit de vragenlijst afkomstig zijn.

Samengestelde variabelen: de variabelen zijn herkenbaar aan de '*n_varnaam*' notatie. De samengestelde variabelen zijn geen originele vragenlijstvariabelen maar worden gebaseerd op één of meerdere vragen uit de vragenlijst. Behalve de regel dat originele variabelen uit de vragenlijst nooit worden gehercodeerd zijn er verschillende redenen waarom deze samengestelde variabelen worden aangemaakt:

- **Rapportagedoeleinden:** het is vaak wenselijk om in rapportages niet de volledige inhoud van een variabele te presenteren maar een samenvoeging van categorieën. Bijvoorbeeld de leeftijd in klassen *n_lftk* i.p.v. alle voorkomende leeftijden (de variabele *lft*).
- **Definitievariabelen:** sommige variabelen zijn gebaseerd op meerdere vragen uit de vragenlijst zoals de variabele beroepsbevolking (*n_berbv2*). Definitievariabelen kunnen alleen aangemaakt worden indien er een valide waarde is bij alle vragen die nodig zijn om de variabele te berekenen. Definitievariabelen hebben daarom in de regel iets meer 'missing values' dan originele variabelen.
- **Verschillen in vraagstelling:** het kan voorkomen dat tussen meetjaren of tussen vragenlijsten verschillen in vraagstelling zijn ontstaan doordat bijvoorbeeld een antwoordcategorie is weggevallen. Een samengestelde variabele kan dan een handige 'samenvatting' van de verschillende varianten geven zoals de variabele over het dienstverband *n_hdvb2*.

3 Enkele kernvariabelen per onderwerp

In deze paragraaf worden enkele variabelen besproken of genoemd die binnen de HBO-Monitor vaak gebruikt worden of bij het werken met het databestand een centrale rol vervullen. Voor een opsomming van alle variabelen die in het bestand zitten wordt verwezen naar het bij het bestand behorende codeboek.

Algemeen en gevolgde opleiding

InstellingRest: de variabele geeft aan of het een respondent van de eigen instelling of een andere instelling betreft. De resultaten die verkregen worden met deze variabele voor de andere instellingen zijn niet gelijk aan het landelijk totaal in instellingsrapportages. In de instellingsrapportages is het landelijke cijfer inclusief de eigen hogeschool. De variabele *InstellingRest* deelt het bestand op in eigen versus overige respondenten en zijn de cijfers voor de 'overigen', dus de overige hogescholen exclusief de eigen hogeschool.

Voplr: gevolgde opleiding op het niveau van 'opleiding' in de online instellingsrapportage

N_gorpli: gevolgde opleiding op het niveau van 'cluster' in de online instellingsrapportage. In veel gevallen is deze gelijk aan de *voplr*-variabele maar met name in de sectoren onderwijs en taal en cultuur zijn er vaak meerdere opleidingen (*voplr*) in één cluster (*n_gorpli*) ondergebracht.

N_govtdt: gevolgde opleidingstype d.w.z. voltijd, deeltijd of duaal.

Vooropleiding

N_vohh: hoogst genoten vooropleiding

Na de HBO-opleiding

N_vvoge: de variabele geeft aan of de respondent aan een vervolgopleiding begonnen is. Deze variabele is een selectievariabele voor het nader analyseren van variabelen die betrekking hebben op het vervolgonderwijs.

N_vvod: geeft de status van de vervolgopleiding aan in termen van afgerond, nog bezig of ongediplomeerd gestopt.

N_vokop1, *n_vokop2*: het niveau en het niveau en de richting van de vervolgopleiding.

Vvoasn: de variabele geeft een indicatie in hoeverre de vervolgstudie aansluit bij de gevolgde HBO-opleiding.

Cbogmd: de variabele geeft aan of de respondent een cursus of bedrijfsopleiding gevolgd heeft.

N_aanb2: geeft aan of de afgestudeerde zich aanbiedt op de arbeidsmarkt. Respondenten die zichzelf als 'student' zien vallen in de categorie 'niet aanbieden'.

N_berbv2: werkzame en werkloze beroepsbevolking. Respondenten die zichzelf als 'student' zien bieden zich niet aan op de arbeidsmarkt (zie ook *n_aanb2*) en worden bij deze variabele buiten beschouwing gelaten. De variabele is een belangrijke selectievariabele voor het nader analyseren van variabelen die betrekking hebben op de baankenmerken.

Huidige functie

N_hbergr: beroepsgroep

N_hbragr: bedrijfsgroep

N_hdvb1: soort dienstverband

N_hasd: tijdelijke/vaste aanstelling

N_hnwg: voor functie vereist opleidingsniveau minimaal eigen opleidingsniveau

N_horwg: voor functie vereiste opleidingsrichting is eigen of verwante richting

N_hau33: full time werkzaam versus part time

N_hbi: bruto maandinkomen
Oordelen van afgestudeerden

N_haso: oordeel over de aansluiting tussen de gevolgde opleiding en functie

Ohfcar: oordeel met betrekking tot carrièremogelijkheden huidige functie

Ohftevr: tevredenheid over huidige functie

N_goazo1: achteraf bezien opnieuw voor zelfde - of andere opleiding kiezen

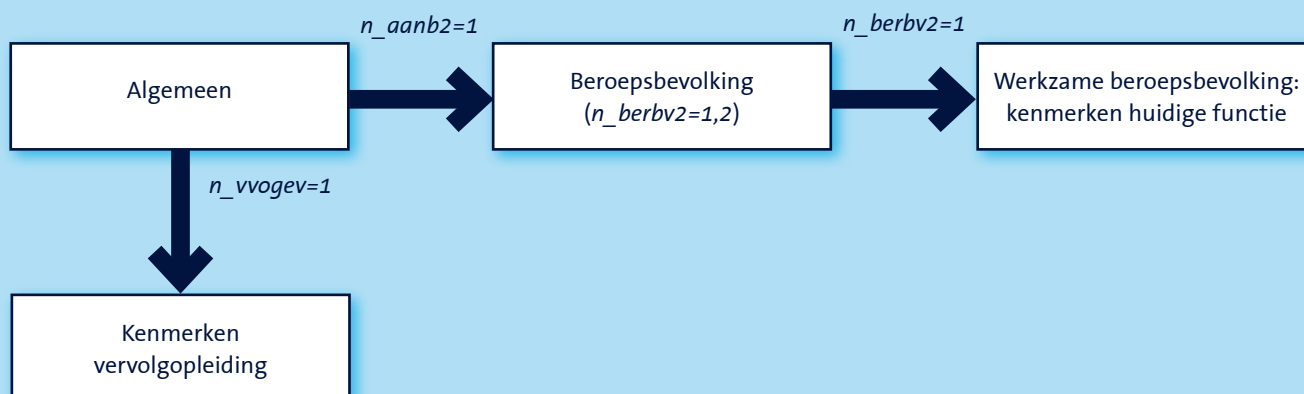
4 Weging

De weging (variabele *wf*) in het bestand vindt plaats op basis van een viertal factoren te weten de opleiding (*croho*), de regio, het type opleiding (voltijd/deeltijd/duaal) en het geslacht. De weging vindt plaats omdat de uiteindelijke verdeling van de respondenten in het bestand niet meer helemaal gelijk is aan die van de populatie. Alhoewel de dekking binnen de HBO-Monitor ongeveer 85% is betekent dit niet dat in alle regio's alle opleidingen even goed in de te benaderen populatie zitten. Daarnaast kan er door selectieve respons een klein verschil ontstaan tussen de responsgroep en de aanvankelijk benaderde populatie. Zo responderen vrouwen bijvoorbeeld in de regel iets beter dan mannen en zou er zelfs bij een dekking van 100% een scheefverdeling tussen mannen en vrouwen ontstaan. De respondenten worden gewogen naar de populatie-aantallen welke afkomstig zijn van het CBS.

5 Aandachtspunten

Selecties i.v.m. de juiste percenteerbasis

Bij het analyseren van vragen over de vervolgopleidingen en de huidige functie is het van belang dat er eerst geselecteerd wordt op de juiste deelgroep. Zo is het bij statistieken over de huidige functie van belang eerst te selecteren op de volgens de op de officiële definitie gebaseerde werkzame beroepsbevolking (*n_berbv2=1*). Dit omdat anders uitspraken over de arbeidsmarkt gedaan worden - en indirect over de gevolgde opleiding - die niet alleen gebaseerd zijn op de echte intreders op de arbeidsmarkt, maar ook op mensen die bijvoorbeeld minder dan 12 uur werken en studenten met een bijbaantje.



Kleine aantallen

Wanneer er gewerkt wordt met een gewogen databestand verdient het aanbeveling om bij twijfel altijd te controleren of het aantal cases ongewogen ook voldoende groot is om uitspraken over te doen. Bij een kleine, ongewogen N kunnen de resultaten aanzienlijk beïnvloed worden door de antwoorden van één of slechts enkele respondenten. Het samenvoegen van bijvoorbeeld enkele meetjaren of het kiezen van een hoger aggregatieniveau kan dit probleem enigszins verhelpen maar er zullen binnen een hogeschool altijd kleine opleidingen blijven met slechts een klein aantal respondenten.

Herkenbaarheid

Behalve het statistische aspect betekenen kleine aantallen ook dat individuen makkelijker herkenbaar zijn. Zeker wanneer cijfers binnen een instelling verspreid worden bestaat de mogelijkheid dat zelfs zonder naw informatie resultaten naar individuen herleidbaar kunnen zijn.

Inhoud tekstvelden

Alvorens het bestand verder te verspreiden verdient het aanbeveling om de ingevulde tekstvelden te 'screenen' op inhoud. Het is iedere deelnemer aan de enquête vrij om in te vullen wat hij of zij wil en er vindt verder geen correctie of cleaning plaats op deze tekstvelden. Het kan dus heel incidenteel voorkomen dat er te persoonlijke, of bijvoorbeeld docenten met naam en toenaam genoemd worden in positieve of negatieve zin. Het verdient daarom aanbeveling om bij het verder verspreiden van het databestand deze tekstvelden er of uit te halen of even door te lopen op inhoud.

Invloed selecties op resultaten

Resultaten kunnen beïnvloed worden door het maken van selecties. Zo hebben werkenden die op minmaal hbo niveau werkzaam zijn en/of in de eigen of verwante richting een veel positiever oordeel over de aansluiting tussen hun opleiding en de arbeidsmarkt dan werkenden die dat niet hebben. Dit doet verder niets af aan de correctheid van deze resultaten, maar het is van belang dit altijd bij de interpretatie in het achterhoofd te houden. Bij het presenteren van de resultaten aan derden is het eveneens van belang dat altijd wordt aangegeven dat de resultaten betrekking hebben op een bepaalde subgroep.

