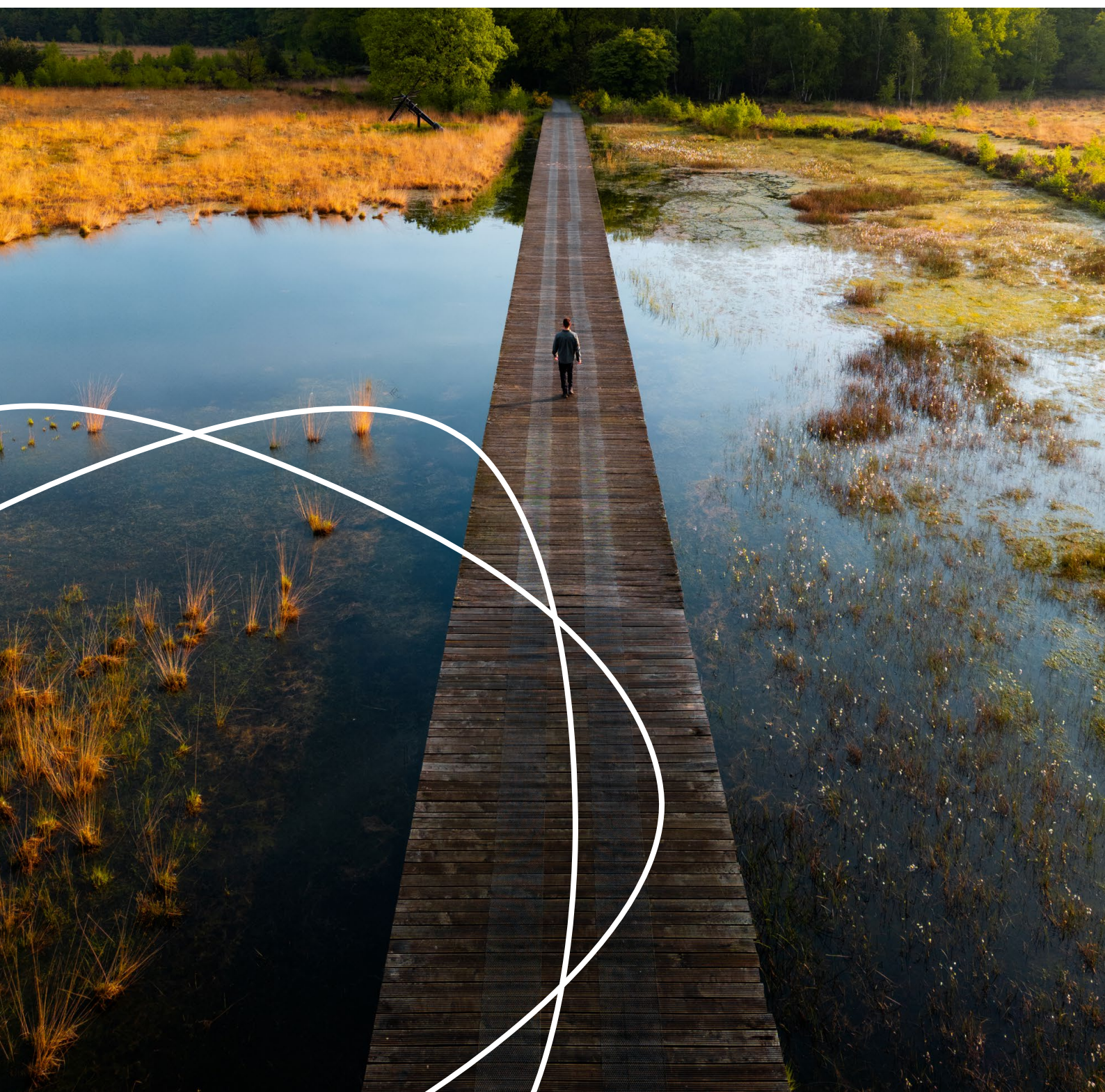


Regionaliseren toeristische statistieken

Rapportage Data & Development Lab



Colofon

Deze rapportage is samengesteld binnen het Data & Development Lab (DDL), de samenwerking van het Centraal Bureau voor de Statistiek (CBS), het Nederlands Bureau voor Toerisme & Congressen (NBTC) en het Centre of Expertise Leisure, Tourism & Hospitality (CELTH).

© 2025, NBTC en CELTH

Alle teksten en gegevens in deze uitgave mogen vermenigvuldigd of openbaar gemaakt worden, mits met bronvermelding. Hoewel wij bij de totstandkoming van deze uitgave de grootst mogelijke zorgvuldigheid hebben betracht, bestaat altijd de mogelijkheid dat bepaalde informatie niet klopt, of na verloop van tijd verouderd, of niet meer juist is. Waar bekend zijn de bronnen vermeld.

Auteurs

Diana Korteweg Maris (CELTH / HZ Kenniscentrum Kusttoerisme)

Evelien Visser-Jonker (NBTC)

Evelot Westerink-Duijzer (CELTH / HZ University of Applied Sciences)

Hans Hoekstra (NBTC)

Fotografie

Voorzijde: Marvin Kuhr

Achterzijde: Sjoerd Bracke & Cuno de Bruin

Grafische vormgeving

Kees Hoendervangers, www.dtp-plus.nl



NBTC

Nederlands Bureau
voor Toerisme &
Congressen



UNIVERSITY
OF APPLIED SCIENCES



CELTH
Centre of Expertise
leisure, tourism & hospitality



**Centraal Bureau
voor de Statistiek**

Inhoudsopgave

Managementsamenvatting	4
1. Inleiding	6
1.1 Toelichting Statistiek Logiesaccommodaties	6
1.2 Regionaliseren van data	7
1.3 Leeswijzer	7
2. Verschillende regionale niveaus	8
2.1 Huidige statistieken	8
2.2 Inventariseren toeristische gebieden	8
2.3 Eerste verkenning voor regionalisering	9
3. Methodologie	11
3.1 CBS methode: directe schatter	11
3.1.1 Analysestrata	11
3.1.2 Samenvoegen van strata	11
3.1.3 Implementatie	12
3.2 Betrouwbaarheid van de statistieken	12
3.3 Alternatieve methode: klein domein schatters	13
3.4 Gekozen methode voor regionaliseren van toeristische data	14
3.4.1. Voorgestelde methode	14
3.4.2 Gebruik van CBS data	14
3.5 Conclusie	16
4. Resultaten	17
4.1 Pilotgebieden	17
4.1.1 Resultaten Zeeland	18
4.1.2 Resultaten Drenthe	19
4.2 Heel Nederland	20
4.3 Effect van de steekproefgrootte	22
5. Conclusie en discussie	25
5.1 Conclusie	25
5.2 Discussie	26
Bijlage A – Begrippenlijst	27
Bijlage B – Toeristische gebieden	29
Bijlage C – CBS methode	31
Bijlage D – Marges	34
Bijlage E – Implementatie	37

Managementsamenvatting

Om te komen tot goede beleidsvorming in het toeristisch-recreatieve domein is data nodig. Er is grote behoefte aan data op regionaal en lokaal niveau over het aantal gasten en overnachtingen in logiesaccommodaties. Middels de Statistiek Logiesaccommodatie geeft CBS inzicht in het aantal gasten en overnachtingen in Nederland. Dit doen zij standaard op landelijk niveau, op provincie niveau, voor de 5 grote steden en voor toeristische gebieden. De vraag is in hoeverre het mogelijk is de data van de CBS Statistiek Logiesaccommodaties verder te regionaliseren naar gemeenten of toeristisch relevante regio's (anders dan de CBS indeling van toeristengebieden). In het DDL-project 'Regionaliseren toeristische statistieken' zijn de mogelijkheden verkend.

Allereerst is de achterliggende methode van de Statistiek Logiesaccommodaties (SLA) grondig in beeld gebracht. De SLA is gebaseerd op een steekproef onder logiesaccommodaties, waarbij de antwoorden van deelnemende bedrijven worden opgehoogd naar de totale populatie van aanwezige logiesaccommodaties in een gebied. Deze methode van ophoging moet ook gebruikt worden bij het samenstellen van cijfers voor kleinere gebieden en/of gemeenten.

Vervolgens is de dataset voor 2023 via de microdata-omgeving van het CBS nauwkeurig in beeld gebracht. Per gemeente en per toeristisch relevante regio is bekeken hoeveel respondenten in de dataset zijn (steekproef) en hoe zich dit verhoudt tot de totale populatie. Zo is nagegaan of de statistieken gepubliceerd mogen worden volgens de outputrichtlijnen van het CBS. Op basis hiervan bleek dat het overgrote deel van de COROP-gebieden en toeristisch relevante gebieden voldoet aan de CBS outputrichtlijnen en voldoende respondenten heeft. Op gemeentelijk niveau beschikt een kleine 20% van de gemeenten over voldoende respondenten. Daarom leken er voldoende mogelijkheden om de verkenning voort te zetten.

Daarna is meer inhoudelijk naar de data gekeken. Betrouwbare statistieken zijn namelijk niet alleen afhankelijk van het aantal respondenten, maar ook van de spreiding in de respons. Gezamenlijk bepaalt dit de betrouwbaarheidsmarge van de statistieken. Als een gebied een heel gelijk aanbod heeft (allemaal vergelijkbare accommodaties, zelfde type, zelfde karakteristieken), dan zijn er naar verwachting niet zulke hoge respondentenaantallen nodig om tot lage marges te komen. Vermoedelijk is de spreiding in je respons dan namelijk ook al laag. Maar als er in een gebied heel verschillende accommodaties zijn, dan is waarschijnlijk de spreiding in de antwoorden ook groot en kan er ondanks redelijke aantallen respondenten toch niet met voldoende betrouwbaarheid een totaalcijfer worden gegeven van de bezetting.

Het is gebruikelijk om in statistieken te werken met een 95% relatieve betrouwbaarheidsmarge van 10%. Ook het CBS hanteert deze marge in de Statistiek Logiesaccommodaties. Daarom is nagegaan voor welke gemeenten en welke toeristisch relevante regio's er voldoende respondenten zijn én de resultaten voldoen aan deze betrouwbaarheidsmarge. Onderstaand

voorbeeld laat zien hoe een relatieve betrouwbaarheidsmarge van 10% geïnterpreteerd moet worden.

Voorbeeld: Stel dat het aantal overnachtingen in hotels in 2023 in provincie X wordt ingeschat op 17.000. Praktisch betekent dit dat met 95% zekerheid het daadwerkelijke aantal overnachtingen in hotels in provincie X ligt tussen 15.300 en 18.700 in de genoemde periode. Deze grenzen zijn gelijk aan 10% onder en 10% boven het geschatte aantal van 17.000.

Kijkend naar de betrouwbaarheidsmarges van de statistieken, bleek dat slechts één op de acht COROP-gebieden voldoet aan de gewenste marge van onder de 10%. Bij toeristisch relevante gebieden en gemeenten is dit nog veel minder. Hoewel dus het aantal respondenten voor een aantal grote gemeenten en toeristisch relevante gebieden voldoende lijkt te zijn, zijn de marges toch te hoog om betrouwbare statistieken te kunnen berekenen. We moeten daarom concluderen dat het met de huidige data niet mogelijk is om betrouwbare statistieken te presenteren op gemeentelijk niveau. Zelfs het samenvoegen van gemeenten tot toeristisch relevante gebieden biedt nauwelijks extra mogelijkheden.

Deze verkenning had ook een onbedoeld bijeffect. Door de marges op de toeristische statistieken in kaart te brengen voor verschillende regionale niveaus, werd duidelijk dat zelfs op COROP-niveau marges soms behoorlijk groot kunnen zijn. Daarom wordt nog verkend in hoeverre het wenselijk is deze COROP-cijfers te blijven publiceren.

Ten slotte is in de verkenning nog bekeken hoe groot de steekproef zou moeten zijn om wel aan de gewenste betrouwbaarheidsmarge van maximaal 10% te komen. Om dit te bereiken voor alle COROP-regio's zou de steekproef moeten groeien van 2832 nu naar meer dan 6000 deelnemende accommodaties. Zo'n uitbreiding is niet aan de orde.

Helaas blijkt dat de SLA nu dus geen mogelijkheden biedt om meer regionale toeristische statistieken samen te stellen. Om te voldoen aan de databehoeftes op regionaal en lokaal niveau zullen andere mogelijkheden moeten worden onderzocht.

1 Inleiding

Om te komen tot goede beleidsvorming in het toeristisch-recreatieve domein is basisdata nodig. Tot deze basisdata behoren onder andere de cijfers over het aantal gasten en overnachtingen in logiesaccommodaties. De Statistiek Logiesaccommodaties van het Centraal Bureau voor de Statistiek (CBS) is een belangrijke databron voor deze cijfers. Deze data zijn beschikbaar op landelijk en provinciaal niveau. Toeristisch-recreatief beleid wordt echter primair op regionaal en lokaal niveau vormgegeven. Er is dan ook grote behoefte aan data op regionaal en lokaal niveau.

1.1 Toelichting Statistiek Logiesaccommodaties

De Statistiek Logiesaccommodaties (SLA) heeft als doel het aantal gasten en hun overnachtingen per type Nederlandse logiesaccommodatie in beeld te brengen. Gasten zijn bezoekers die één of meer nachten achtereen verblijven in een logiesaccommodatie. Overnachtingen zijn alle nachten die gasten doorbrengen in een logiesaccommodatie. Een gezelschap van 4 personen die samen 3 nachten in een accommodatie verblijven, tellen zo voor 12 overnachtingen.

Het aantal gasten en overnachtingen wordt onderzocht door enquêtering van logiesaccommodaties. De populatie van logiesaccommodaties bestaat uit de volgende soorten accommodaties:

- Hotels, motels, pensions, appartementen met hoteldienstverlening, jeugdaccommodaties en bed & breakfasts met minimaal 5 slaapplekken;
- Huisjesterreinen met minimaal 10 slaapplekken;
- Kampeesterreinen met minimaal 4 toeristische standplaatsen. Het gaat hierbij om toeristische standplaatsen voor kortdurend verblijf. Plaatsen die voor langere duur verhuurd worden, zoals jaarplaatsen of seizoenplaatsen, maken geen deel uit van de statistiek;
- Groepsaccommodaties met minimaal 10 slaapplekken

Een gedetailleerde omschrijving van de begrippen gehanteerd in de SLA staat in bijlage A. Geen onderdeel van de populatie zijn logiesaccommodaties voor eigen gebruik, denk hierbij aan tweede woningen of stacaravans op jaarplaatsen op campings. Jachthavens en cruiseschepen maken ook geen deel uit van de populatie. Door de ondergrens blijven zeer kleinschalige aanbieders eveneens buiten beschouwing, denk aan particuliere verhuur / verhuur van woonruimte.

Jaarlijks wordt een steekproef van circa 30% getrokken uit de totale populatie van geregistreerde logiesaccommodaties. Deze steekproef ontvangt vervolgens maandelijks een vragenlijst over het aantal gasten en overnachtingen, gespecificeerd naar herkomstland. De

resultaten worden onder andere gecontroleerd op uitschieters en eventuele afwijkingen ten opzichte van vorig jaar op individueel niveau en samengevoegde cijfers per gebied. Vervolgens worden de resultaten gewogen naar de totale populatie, per provincie en voor heel Nederland.

De SLA-data zijn als open data beschikbaar via het dataportaal Statline van het CBS, voor Nederland als geheel, voor provincies en voor gegroepeerde toeristengebieden. De beschikbare tijdreeks start vanaf 2012, met cijfers per maand, per kwartaal en per jaar. De data worden maandelijks gepubliceerd, meestal zo'n 2 maanden na afloop van de verslagmaand. De SLA-data komen overeen met de richtlijnen van de Europese statistische organisatie Eurostat en zijn input voor Europese statistieken over gasten en overnachtingen.

1.2 Regionaliseren van data

Zoals eerder benoemd, is er een grote behoefte aan data op regionaal en lokaal niveau. In het Data Center Toerisme hebben NBTC en CBS samen al gezorgd voor een regionalisering van de data naar het niveau van COROP-gebieden, clusters van gemeenten ten behoeve van regionaal onderzoek. Deze gebiedsindeling sluit echter niet aan op de informatiebehoefte van partijen in het toeristisch-recreatieve werkveld, zoals gemeenten en Destinatie Management Organisaties (DMO's). De wens is om data op gemeentelijk niveau te hebben of voor toeristische gebieden. Om te verkennen in hoeverre aan de behoefte voor verder geregionaliseerde statistieken tegemoet kan worden gekomen is begin 2024 het DDL-project 'Regionaliseren toeristische statistieken' gestart. In dit rapport worden de bevindingen van dit project gepresenteerd.

1.3 Leeswijzer

Het vervolg van dit rapport is als volgt opgebouwd. We starten in Hoofdstuk 2 met een bespreking van de verschillende regionale niveaus: we kijken welke data er op dit moment beschikbaar is en verkennen de mogelijkheden tot regionaliseren. Vervolgens bespreken we in Hoofdstuk 3 de methodologie die het CBS gebruikt voor het berekenen van de toeristische statistieken. We leggen uit hoe deze methode toegepast kan worden op lagere regionale niveaus. Hierbij kijken we niet alleen naar het berekenen van de statistieken zelf, maar ook naar de precisie van deze statistieken. Aansluitend presenteren we in Hoofdstuk 4 de resultaten. Het rapport besluit met een discussie en conclusie in Hoofdstuk 5.

2 Verschillende regionale niveaus

In dit hoofdstuk verkennen we de mogelijkheden om vanuit de Statistiek Logiesaccommodaties data te creëren voor gemeenten of toeristisch relevante gebieden.

2.1 Huidige statistieken

Het uitgangspunt in dit project is de Statistiek Logiesaccommodaties zoals die door het CBS beschikbaar wordt gesteld via StatLine. Deze statistiek geeft inzicht in het aantal gasten en overnachtingen per provincie.

Deze statistieken zijn berekend aan de hand van geschatte bezettingen, waarvoor het CBS gebruik maakt van een steekproef. Deze steekproef bevat ongeveer een derde van alle geopende accommodaties. Ieder jaar wordt een andere steekproef getrokken en de accommodaties in deze steekproef wordt gevraagd te reageren op een enquête. In deze enquête geven ze per maand en herkomstland aan hoeveel gasten bij hen hebben overnacht en wat het bijbehorende aantal overnachtingen is. De resultaten van de enquête worden vervolgens gebruikt om een inschatting te maken van het aantal gasten en overnachtingen bij accommodaties die niet in de steekproef zitten of niet hebben gereageerd op de enquête.

Naast de provinciale toeristische statistieken van het CBS, publiceert NBTC op hun website sinds 2020 ook statistieken op COROP-niveau. Deze cijfers zijn eveneens gebaseerd op de steekproef van het CBS. Bij cijfers op COROP-niveau loopt NBTC er soms tegenaan dat de betrouwbaarheid van data in het geding komt en dat er te veel risico is dat data van individuele bedrijven onthuld wordt. Bij verder regionaliseren naar gemeentelijk niveau wordt dit effect groter.

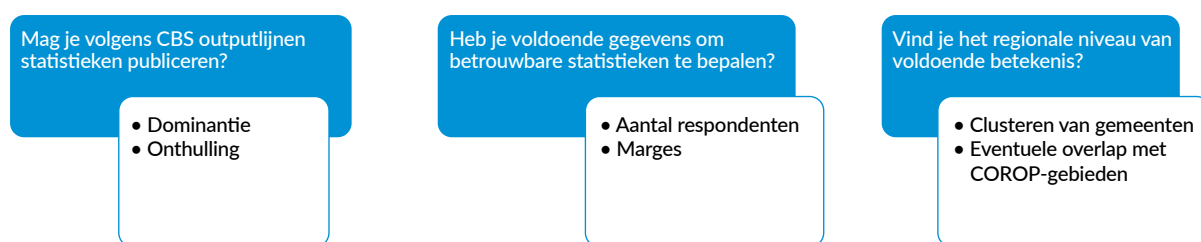
2.2 Inventariseren toeristische gebieden

Om betrouwbare statistieken op regionaal niveau te kunnen presenteren, zijn er voldoende accommodaties en respondenten nodig per regio. We beseffen dat dit met name voor kleine gemeenten waarschijnlijk niet het geval zal zijn. Daarom beschouwen we niet alleen het gemeentelijk niveau, maar bekijken we ook de optie om bepaalde gemeenten samen te voegen tot toeristisch relevante gebieden. Dit zijn andere toeristische gebieden dan de toeristengebieden die het CBS hanteert bij de publicatie van statistieken op StatLine. Bij provincies en DMO's is in beeld gebracht welke wensen er zijn als het gaat om gebiedsindelingen.

Deze uitvraag heeft reactie opgeleverd van zeven provincies: Fryslân, Drenthe, Overijssel, Gelderland, Utrecht, Noord-Brabant en Zeeland. De gewenste indeling in toeristische gebieden voor deze provincies is te vinden in Bijlage B. Het verschilt per provincie in hoeverre deze indeling afwijkt van de COROP-indeling. Voor bijvoorbeeld Drenthe en Gelderland is de overlap tussen COROP-gebieden en toeristische gebieden vrij groot. Bij de provincie Utrecht is er veel verschil, aangezien de provincie uit slechts één COROP-gebied bestaat, maar werkt met zes toeristische gebieden.

2.3 Eerste verkenning voor regionalisering

Om te bepalen of het mogelijk is toeristische statistieken op regionaal niveau weer te geven, moeten verschillende vragen worden beantwoord. Deze vragen zijn weergegeven in het onderstaande schema en om deze vragen te beantwoorden maken we gebruik van de microdata van het CBS in een RA omgeving.



Figuur 1 – Stappenplan om te komen tot betrouwbare en betekenisvolle regionale toeristische statistieken.

Zoals te zien is in Figuur 1 is de eerste vraag of het toegestaan is volgens de CBS outputrichtlijnen om statistieken te publiceren op een bepaald niveau. Deze outputrichtlijnen zijn de criteria die het CBS hanteert en houden rekening met onthullingsrisico en dominantie van bepaalde accommodaties in een bepaald gebied. Om deze vraag te kunnen beantwoorden, is informatie over het aantal accommodaties per regio en hun capaciteit voldoende. Wanneer het is toegestaan om statistieken te publiceren, is de volgende vraag of het ook mogelijk is om betrouwbare statistieken te bepalen. Hierbij speelt het aantal respondenten een belangrijke rol en de marges op de berekende statistieken. Deze marges bieden een soort bandbreedte van onzekerheid rondom de berekende bezetting. Wanneer de marges te groot zijn, is het niet mogelijk om met voldoende zekerheid iets te zeggen over de statistieken. De laatste vraag is of het regionale niveau van de statistieken ook van voldoende betekenis is. Hierbij speelt bijvoorbeeld mee of de samenvoeging van gemeenten bruikbaar is voor partijen in het toeristisch recreatieve werkveld. Ook de eventuele overlap met COROP-gebieden kan bepalen of de regionale statistieken voldoende meerwaarde hebben.

In onze eerste verkenning analyseren we de eerste twee vragen, waarbij we voor de tweede vraag alleen naar het aantal respondenten kijken. In latere hoofdstukken zullen ook de marges berekend worden. Door middel van toegang tot de microdata in de RA omgeving van het CBS is de dataset van het jaar 2023 nauwkeurig geanalyseerd voor wat betreft de omvang van de populatie en omvang van de steekproef op verschillende gebiedsniveaus. De resultaten zijn weergegeven in Tabel 1.

Regio's	Aantal	Aantal dat voldoet aan CBS outputrichtlijnen (dominantie en onthulling)		Aantal met minstens 10 respondenten	
		Januari	Juli	Januari	Juli
Provincies*	13	13	13	13	13
COROP-gebieden	40	40	40	36	38
Toeristische gebieden	32	32	32	28	31
Gemeenten	342	155	209	40	80

Tabel 1 – Analyse van de populatie en steekproef van 2023 op verschillende regionale niveaus. We kijken hierbij op totaalniveau en nemen alle accommodatietypen samen. (*) De hotels in Amsterdam worden apart genomen en losgekoppeld van de provincie Noord-Holland. Zij vormen daarmee als het ware het een dertiende provincie.

In Tabel 1 zien we dat alle provincies voldoen aan de CBS outputrichtlijnen én minstens 10 respondenten hebben. Deze grenswaarde van 10 respondenten is gekozen in overeenstemming met de regels van het CBS om alleen statistieken publiceren die gebaseerd zijn op minstens 10 respondenten. Ook COROP-gebieden en toeristische gebieden voldoen vrijwel altijd aan de criteria. Bij verder regionaliseren naar gemeentelijk niveau verandert het beeld. Slechts een deel van de gemeenten voldoet aan de CBS outputrichtlijnen of aan het aantal van 10 respondenten. Ook zien we dat er in juli duidelijk meer gemeenten voldoen dan in januari, wat te maken heeft met accommodaties die niet het gehele jaar geopend zijn. Niettemin voldoet bijna de helft van de gemeenten aan de CBS outputrichtlijnen en 12-23% van de gemeenten voldoet ook aan het aantal van minstens 10 respondenten.

Op basis van deze resultaten is er voldoende reden om het onderzoek voort te zetten. Het lijkt voor voldoende grote gemeenten en toeristische gebieden mogelijk betrouwbare toeristische statistieken op regionaal niveau te bepalen. Om met zekerheid te kunnen stellen of voldoende betrouwbaarheid gegarandeerd is, zullen we de betrouwbaarheidsmarges op de berekende statistieken moeten berekenen. In het volgende hoofdstuk gaan we hier verder op in.

3 Methodologie

In dit hoofdstuk bespreken we op welke manier toeristische statistieken berekend kunnen worden. We evalueren hiervoor de methode die het CBS gebruikt en bespreken welke aanpassingen nodig zijn om toeristische statistieken te regionaliseren. We bespreken eveneens de betrouwbaarheidsmarges van de statistieken.

3.1 CBS methode: directe schatter

Het CBS schat de bezetting met een directe schatter op basis van de gegevens uit de steekproef. De accommodaties uit de steekproef rapporteren per maand en herkomstland het aantal gasten dat bij hen overnacht heeft en het aantal overnachtingen. Met behulp van deze informatie wordt vervolgens ook het aantal gasten en overnachtingen geschat voor accommodaties die niet in de steekproef zitten.

3.1.1 Analysestrata

Voor het berekenen van de directe schatter wordt gebruik gemaakt van analysestrata. Alle accommodaties worden ingedeeld in een bepaald stratum (groep) op basis van het logiestype, de grootteklasse, de regio, de maand en het herkomstland waarover gerapporteerd wordt. Om de geschatte bezetting te bepalen van een accommodatie in een bepaald stratum, wordt bij voorkeur alleen gebruik gemaakt van de gegevens van de accommodaties uit de steekproef in datzelfde stratum. Dit is alleen mogelijk als een stratum voldoende respondenten heeft. Er wordt in dat geval een gemiddelde bezetting genomen over alle accommodaties in dat stratum en deze gemiddelde bezetting wordt toegepast op de accommodaties die niet in de steekproef zitten.

3.1.2 Samenvoegen van strata

Wanneer een stratum te weinig respondenten heeft, zijn de gegevens uit de steekproef voor dat stratum onvoldoende betrouwbaar om een schatting te maken voor de overige accommodaties. Het CBS gebruikt in dat geval een methode waarbij strata worden samengevoegd. Deze methode is nauwkeuriger beschreven in Bijlage C.

Bij het samenvoegen van strata worden bepaalde prioriteiten gehanteerd om te komen tot betrouwbare gegevens op basis van voldoende respondenten. Samenvoegen op basis van grootteklasse heeft de eerste voorkeur. Als dat nog te weinig respondenten oplevert, wordt samengevoegd op basis van regio. Op basis van de andere eigenschappen, d.w.z. logiestype, maand en herkomstland van gasten, wordt niet samengevoegd. De reden hiervoor is waarschijnlijk dat maanden en logiestypen dermate verschillend zijn dat samenvoegen zou leiden tot verlies van belangrijke informatie. Samenvoegen van strata met verschillende herkomstlanden levert geen extra respondenten op, aangezien elke accommodatie in de

steekproef voor ieder herkomstland gegevens rapporteert (ook als het aantal gasten of overnachtingen uit dat herkomstland 0 is).

3.1.3 Implementatie

Bij het implementeren van de methode van het CBS, zoals beschreven in Bijlage C, liepen we tegen een aantal onduidelijkheden aan. De beschrijving van de methode is soms op meerdere manieren te interpreteren, waardoor het onduidelijk is welke implementatiekeuzes het CBS zelf gemaakt heeft. Onduidelijkheid ontstond met name rond de vragen wanneer analysestrata worden samengevoegd en hoe deze samenvoeging dan plaatsvindt.

De projectgroep heeft deze implementatievragen in juli 2024 aan het CBS voorgelegd. Door omstandigheden duurde het helaas tot december 2024 voordat CBS helderheid kon verschaffen over hun implementatie. We hebben deze reactie verwerkt en beschrijven in Bijlage E op welke punten de implementatie afwijkt van de gedocumenteerde methode.

3.2 Betrouwbaarheid van de statistieken

Naast het bepalen van de geschatte bezetting, berekenen we ook de precisie van de verkregen schatting. We doen dit door middel van de relatieve 95%-betrouwbaarheidsmarge. In Bijlage D is in detail beschreven hoe deze marges berekend kunnen worden. De relatieve betrouwbaarheidsmarge kan als volgt worden geïnterpreteerd.

Stel dat het aantal overnachtingen in hotels in januari 2023 in provincie X wordt ingeschat op 17.000 en dat de 95% relatieve betrouwbaarheidsmarge gelijk is aan 10%. Dit betekent dat met 95% zekerheid het daadwerkelijke aantal overnachtingen in hotels in provincie X ligt tussen de 15.300 en de 18.700 in de genoemde periode. Deze grenzen zijn respectievelijk gelijk aan 10% onder en 10% boven het geschatte aantal van 17.000.

Als een schatting een hoge marge heeft, is het interval waarin het daadwerkelijke cijfer met 95% zekerheid ligt ook groot. Dit betekent dat het daadwerkelijke cijfer mogelijk veel afwijkt van de schatting. Publiceren van schattingen met een hoge marge is daarom niet wenselijk, omdat dit kan leiden tot verwarring bij gebruikers en er verkeerde conclusies getrokken kunnen worden.

Er kunnen verschillende oorzaken zijn voor een hoge marge. Allereerst speelt het **aantal respondenten** een belangrijke rol. Bij het berekenen van geregionaliseerde statistieken bestaat het risico dat de marges toenemen, omdat statistieken in iedere regio worden geschat op basis van minder respondenten. Hoe minder respondenten, hoe lastiger het kan zijn om een betrouwbare schatting te maken. Toch is niet alleen het aantal respondenten van invloed. Nog belangrijker is de **spreiding in de antwoorden** die respondenten geven in de enquête. In een

gemeente waar de accommodaties heel erg op elkaar lijken (bijvoorbeeld allemaal middelgrote hotels in een stedelijke omgeving), kan een klein aantal respondenten voldoende zijn om een goede schatting te maken van de bezetting. Wanneer accommodaties binnen een gemeente sterk van elkaar verschillen (denk aan een combinatie van bijvoorbeeld grote en kleine hotels, campings en/of groepsaccommodaties), dan zijn er veel meer respondenten nodig om tot betrouwbare bezettingscijfers te komen. Echter, zelfs wanneer de accommodaties in een gemeente sterk op elkaar lijken maar heel verschillende bezettingen rapporteren, kunnen toch veel respondenten nodig zijn. Door grote verschillen in gerapporteerde bezettingen is het moeilijk in te bepalen of de accommodaties die niet in de steekproef zitten meer lijken op de accommodaties met een hoge of met een lage bezetting. De totaalcijfers voor die gemeente kunnen dan flink variëren en hebben dus een grote marge.

Om er zeker van te zijn dat we betrouwbare cijfers rapporteren, is het dus noodzakelijk om de marges goed in kaart te brengen en cijfers alleen te rapporteren als we marges laag genoeg vinden.

3.3 Alternatieve methode: klein domein schatters

Tijdens een overleg met het CBS op 16 juli 2024 werd nog een alternatieve methode voorgesteld voor het bepalen van schattingen op gemeentelijk niveau. Het gebruik van klein domein schatters wordt door het CBS als de meest veelbelovende richting gezien voor gemeentelijke statistieken, aangezien hierbij gebruik wordt gemaakt van zoveel mogelijk data en marges verkleind kunnen worden. Deze klein domein schatters combineren een directe schatter op gemeentelijk niveau met een schatting op basis van een model over een grote regio. Wanneer een gemeente voldoende respondenten heeft om een betrouwbare schatting te maken, zal de klein domein schatter meer neigen naar de directe schatter. Bij een gemeente met te weinig respondenten voor een betrouwbare schatting, wordt de klein domein schatter voornamelijk bepaald door de schatting op basis van het model. Door gebruik te maken van klein domein schatters kunnen de marges verkleind worden en kan een betrouwbaardere schatting worden bepaald op gemeentelijk niveau.

Om klein domein schatters toe te passen, moet een model worden opgesteld. Dit model voorspelt op basis van data wat de bezetting is in een bepaalde gemeente. Voor het opstellen van een dergelijk model moeten verschillende keuzes worden gemaakt. Allereerst moet worden bepaald op welk niveau modellen worden gemaakt. Zo kun je een model maken voor iedere maand apart, maar ook voor het hele jaar. Je kunt een model maken per provincie, maar ook voor heel Nederland. Daarnaast moet gekozen worden welke informatie wordt meegenomen in het model. Naast de beschikbare data uit de steekproef kan hulpinformatie worden toegevoegd die voorspellingen kan verbeteren. Concluderend kunnen we stellen dat het werken met klein domein schatters een aantal belangrijke keuzes vereist voor het te ontwikkelen model en mogelijk ook enig experimenteren met verschillende modellen.

Klein domein schatters hebben als voordeel dat ze over het algemeen resulteren in lagere marges en daarmee in betrouwbaardere statistieken. Hoe sterk de marges gereduceerd kunnen worden, verschilt per situatie, al zeiden de deskundigen van het CBS dat de reductie in de marges vaak beperkt is.

3.4 Gekozen methode voor regionaliseren van toeristische data

In deze sectie zullen we bespreken hoe we de keuze hebben gemaakt voor de methode voor het bepalen van geregionaliseerde data.

3.4.1 Voorgestelde methode

In de voorgaande secties zijn twee methoden besproken: de huidige CBS methode via directe schatters en de alternatieve methode via klein domein schatters. Na overleg met het projectteam is in juli 2024 besloten om binnen het DDL-project te kiezen voor directe schatters. Een belangrijke reden hiervoor was dat met de keuze voor directe schatters wordt aangesloten op de door het CBS gebruikte methode, die ook als basis dient voor de COROP-cijfers die het NBTC publiceert. We zijn ons ervan bewust dat het gebruik van directe schatters mogelijk in hogere marges resulteert, waardoor gemeenten mogelijk gegroepeerd moeten worden om tot betrouwbare statistieken te komen. We zien het echter als belangrijk doel van dit project om de grenzen te onderzoeken van wat haalbaar is op basis van de huidige steekproef.

Een andere reden voor dit besluit was dat naar verwachting de resterende looptijd van het project en het nog beschikbare budget onvoldoende waren om te komen tot een goede implementatie van klein domein schatters. Ook speelde mee dat het gebruik van klein domein schatters niet het probleem oplost van de CBS outputrichtlijnen, waardoor alsnog voor veel gemeenten geen statistieken vrijgegeven zouden mogen worden, ongeacht hoe klein de marges zijn.

3.4.2 Gebruik van CBS data

Om directe schatters te berekenen voor de toeristische statistieken, moeten we bepalen hoe we gebruik maken van de microdata van het CBS. Deze microdata bevat voor alle respondenten hun gerapporteerde bezetting, waardoor schattingen voor deze accommodaties niet nodig zijn. In de microdata zijn daarnaast geschatte bezettingen te vinden voor de overige accommodaties, zowel in termen van gasten als overnachtingen. Met behulp van de geschatte bezetting kunnen de toeristische statistieken rechtstreeks worden berekend voor niet-respondenten.

In Sectie 3.1 hebben we toegelicht hoe de bezetting wordt geschat volgens de methode van het CBS. In hoofdlijnen geldt dat er een gewogen gemiddelde wordt genomen over vergelijkbare accommodaties in een bepaalde regio. Het CBS berekent twee versies van de geschatte bezetting, waarbij het regionale niveau van de geschatte bezetting verschilt. Er

zijn voor iedere accommodatie schattingen op provinciaal (PROV) niveau en op het niveau van toeristengebieden (TG)¹. Dit geeft ons de volgende mogelijkheden om data op een lager regionaal niveau te berekenen.

Niveau waarop de geschatte bezetting wordt berekend	Niveau waarop de data wordt gepubliceerd	Opmerking
PROV	PROV	De provinciale statistieken op StatLine.
PROV	COROP	De COROP-statistieken van het NBTC.
TG	COROP	Een mogelijk alternatief voor de COROP-statistieken van het NBTC.
COROP*	COROP	Een mogelijk alternatief voor de COROP-statistieken van het NBTC.
PROV	GEM	Mogelijkheid in DDL project
TG	GEM	Mogelijkheid in DDL project
GEM*	GEM	Mogelijkheid in DDL project

Tabel 2 – Overzicht van mogelijke combinaties van het regionale niveau van de schatting en het niveau waarop de data wordt gepubliceerd. (*) Geschatte bezetting op dit niveau wordt niet door het CBS geleverd en moet zelf worden berekend.

Tabel 1 laat grofweg twee manieren zien om regionale statistieken te berekenen. Ten eerste, kan er rechtstreeks gebruik worden gemaakt van de door het CBS geschatte bezetting op niveau van provincie of toeristengebied. Ten tweede is het mogelijk om zelf de geschatte bezetting op een bepaald regionaal niveau te bepalen. De eerste methode is eenvoudig toe te passen, maar heeft als groot nadeel dat grote verschillen binnen een provincie of toeristengebied niet aan het licht zullen komen in de gemeentelijke statistieken. Er wordt immers aangenomen dat de bezetting gemiddeld per provincie of toeristengebied gelijk is. De tweede methode vereist complexere berekeningen op basis van de enquêtegegevens van het CBS, maar geeft wel een betrouwbaarder beeld van de situatie per gemeente. Een ander mogelijk nadeel van de tweede aanpak is dat er niet voor alle gemeenten genoeg respondenten zijn om tot betrouwbare schattingen te komen. Hierdoor zal in sommige gevallen toch gebruik gemaakt moeten worden van geschatte bezettingen op hoger regionaal niveau (COROP of provincie).

Na overleg met het projectteam kiezen we voor de tweede aanpak, waarbij we de geschatte bezettingen zelf berekenen op gemeentelijk niveau volgens de CBS methode die beschreven is in Bijlage C. Op deze manier wordt de gemeentelijke situatie zo goed mogelijk weergegeven en blijven de statistieken qua methode in lijn met die op hogere regionale niveaus. Het is goed om op te merken dat wanneer we hierboven spreken over ‘gemeentelijk niveau’ we een regionaal niveau bedoelen dat gedetailleerder is dan COROP-niveau, maar niet noodzakelijk overeenkomt met daadwerkelijke gemeenten. We houden de mogelijkheid open om bepaalde

1 De door het CBS gehanteerde indeling van toeristengebieden is niet gelijk aan de indeling van toeristische gebieden in Bijlage B. Er zijn in totaal 17 toeristengebieden, terwijl het aantal toeristische gebieden groter is.

gemeenten samen te voegen tot toeristisch relevante gebieden, wanneer gemeenten zelf over te weinig respondenten beschikken of de gemeentelijke statistieken tot te hoge marges leiden.

3.5 Conclusie

In dit hoofdstuk hebben we een voorstel gedaan voor de methode voor het bepalen van toeristische statistieken op regionaal niveau. We kiezen voor directe schatters en voegen strata waar nodig samen volgens de methode van het CBS. Wanneer een regio over te weinig respondenten beschikt, zullen we strata samenvoegen op een hoger regionaal niveau. We publiceren statistieken alleen als de bijbehorende marges laag genoeg zijn. Op die manier voorkomen we het vrijgeven van onbetrouwbare informatie die bijvoorbeeld gebaseerd is op te weinig respondenten. De gekozen methode is in principe toepasbaar op geheel Nederland.

4 Resultaten

In het projectplan is ervoor gekozen om een eerste analyse uit te voeren op een aantal pilotgebieden. We hebben gekozen voor de provincies Zeeland en Drenthe als pilotgebieden. In Sectie 4.1 zullen de resultaten van deze pilotgebieden worden besproken. Aansluitend geven we in Sectie 4.2 de resultaten voor heel Nederland. We besluiten dit hoofdstuk met een bespreking van het effect van de steekproefgrootte op de betrouwbaarheid van de statistieken in Sectie 4.3.

De resultaten in dit hoofdstuk zijn gebaseerd op analyses in de microdata omgeving van het CBS. Aangezien volgens de regels van het CBS alleen statistieken mogen worden weergegeven die gebaseerd zijn op minstens 10 respondenten, zullen in sommige tabellen gegevens ontbreken voor de regio's waarin niet aan dit criterium is voldaan.

4.1 Pilotgebieden

Voor de pilotgebieden Zeeland en Drenthe onderzoeken we de volgende onderzoeksvragen op vier regionale niveaus: provinciaal, COROP-gebied, toeristische gebied en gemeentelijk.

Voldoet de regio aan de outputrichtlijnen van het CBS? Dat wil zeggen, zijn er in de regio minstens 10 accommodaties en is de capaciteit van de grootste accommodatie minder dan 50% van de totale capaciteit?

Is het mogelijk om betrouwbare statistieken te bepalen voor deze regio? We bekijken hierbij naar het aantal respondenten en naar de betrouwbaarheidsmarges van de statistieken, waarbij we een betrouwbaarheidsmarge van maximaal 10% als acceptabel beschouwen.

4.1.1 Resultaten Zeeland

In Tabel 2 zijn de antwoorden op bovenstaande vragen weergegeven voor de provincie Zeeland, waarbij we als statistiek uitgaan van het totale aantal gasten in de maand juli 2023.

Indeling	Regio	1. CBS outputrichtlijnen	2. Betrouwbaarheid	
		Voldaan?	# respondenten	Betrouwbaarheidsmarge
Provincie	Zeeland	Ja	277	0.06
COROP	Zeeuws-Vlaanderen	Ja	57	0.23
	Overig Zeeland	Ja	220	0.06
Toeristisch gebied	Beveland & Tholen	Ja	27	0.38
	Schouwen-Duiveland	Ja	62	0.26
	Walcheren	Ja	131	0.20
	Zeeuws-Vlaanderen	Ja	57	0.23
Gemeente	Borsele*			
	Goes*			
	Hulst*			
	Kapelle*			
	Middelburg	Ja	14	0.27
	Noord-Beveland	Ja	13	0.40
	Reimerswaal*			
	Schouwen-Duiveland	Ja	62	0.26
	Sluis	Ja	46	0.50
	Terneuzen*			
	Tholen*			
	Veere	Ja	106	0.26
	Vlissingen	Ja	11	0.57

Tabel 3 – Overzicht van de regio's in Zeeland met daarbij aangegeven of ze voldoen aan de CBS outputrichtlijnen en of er betrouwbare statistieken kunnen worden weergegeven. (*) Deze gemeenten hebben minder dan 10 respondenten, waardoor geen cijfers kunnen worden vrijgegeven.

Uit bovenstaande tabel kunnen we afleiden dat op vrijwel alle regionale niveaus is voldaan aan de CBS outputrichtlijnen. Op provinciaal niveau is ook de betrouwbaarheid gegarandeerd, gezien de lage marges. Deze marges nemen echter snel toe als we verder regionaliseren. Al op het niveau van COROP-gebieden en toeristische gebieden zien we marges die behoorlijk hoog zijn, vermoedelijk als gevolg van het beperkte aantal respondenten per gebied. Op gemeentelijk niveau is dat nog sterker het geval. Slechts een beperkt aantal gemeenten heeft minstens 10 respondenten, waardoor cijfers kunnen worden vrijgegeven. Daarnaast zijn de marges voor geen van de toeristisch gebieden of gemeenten laag genoeg om tot betrouwbare statistieken te komen, uitgaande van een acceptabele marge van maximaal 10%.

We merken op dat de resultaten gebaseerd zijn op totaalcijfers voor de maand juli 2023. Voor andere maanden, zoals bijvoorbeeld januari, kunnen deze cijfers er anders uitzien en ligt het aantal respondenten buiten het zomerseizoen vermoedelijk nog lager. De statistieken

zijn gebaseerd op totaalcijfers en zijn dus niet uitgesplitst naar logiestype of herkomst van de gasten. Verder uitsplitsen kan enerzijds zorgen voor lagere marges, doordat soortgelijke accommodaties bij elkaar worden geplaatst. Tegelijkertijd betekent verder uitsplitsen dat er minder respondenten zijn per deelgroep, wat weer nadelig werkt voor de marges.

4.1.2 Resultaten Drenthe

Een soortgelijke analyse voeren we ook uit voor de provincie Drenthe.

Indeling	Regio	1. CBS outputrichtlijnen	2. Betrouwbaarheid	
		Voldaan?	# respondenten	Betrouwbaarheidsmarge
Provincie	Drenthe	Ja	162	0.06
COROP	Noord-Drenthe	Ja	68	0.15
	Zuidoost-Drenthe	Ja	42	0.17
	Zuidwest-Drenthe	Ja	52	0.34
Toeristisch gebied	Hondsrug	Ja	67	0.09
	Zuidwest Drenthe	Ja	52	0.34
	Midden-Drenthe	Ja	18	0.42
	Kop van Drenthe	Ja	25	0.33
Gemeente	Aa en Hunze	Ja	25	0.25
	Assen*			
	Borger-Odoorn	Ja	19	0.21
	Coevorden	Ja	13	0.23
	De Wolden	Ja	44	0.33
	Emmen	Ja	10	0.52
	Hoogeveen*			
	Meppel*			
	Midden-Drenthe	Ja	18	0.42
	Noordenveld	Ja	12	0.93
	Tynaarlo*			
	Westerveld*			

Tabel 4 – Overzicht van de regio's in Drenthe met daarbij aangegeven of ze voldoen aan de CBS outputrichtlijnen en of er betrouwbare statistieken kunnen worden weergegeven. (*) Deze gemeenten hebben minder dan 10 respondenten, waardoor geen cijfers kunnen worden vrijgegeven.

De resultaten van de provincie Drenthe zijn vergelijkbaar met die van Zeeland. Betrouwbare cijfers met lage marges zijn vrijwel alleen te garanderen op provinciaal niveau. Naast deze algemene conclusie, is er nog een opvallende resultaat. We zien namelijk een lage marge in het toeristische gebied Hondsrug. Dit gebied bestaat uit het COROP-gebied Zuidoost-Drenthe, aangevuld met één gemeente uit het COROP-gebied Noord-Drenthe (Aa en Hunze). Blijkbaar is de bezetting in Aa en Hunze vergelijkbaar met die in het COROP-gebied Zuidoost-Drenthe en zorg het toevoegen van deze extra gemeente voor voldoende extra respondenten om tot een lagere marge te komen. Dit resultaat laat zien dat het groeperen van vergelijkbare gemeenten ervoor kan zorgen dat je lagere marges vindt. Soortgelijke resultaten kunnen

ook optreden als je verder uitsplitst in je resultaten, door bijvoorbeeld voor hotels aparte statistieken te rapporteren. Toch is dit niet noodzakelijk de oplossing voor het probleem van hoge marges. Hoewel je accommodaties kunt selecteren die erg op elkaar lijken (in een bepaalde regio of van een bepaald logiestype), houd je ook de resterende accommodaties over. Voor deze accommodaties zijn de marges niet altijd lager. Ook dit zien we terug in Tabel 4. Het COROP-gebied Noord-Drenthe wordt verdeeld over de toeristische gebieden Hondsrug, Midden-Drenthe en Kop van Drenthe. Alleen Hondsrug heeft een lage marge. De andere toeristische gebieden hebben juist hogere marges dan op COROP-niveau.

4.2 Heel Nederland

We breiden vervolgens onze resultaten uit naar heel Nederland. Omdat we ons in dit onderzoek specifiek richten op het regionaliseren van de toeristische statistieken, geven we hier alleen de resultaten weer op de regionale niveaus van COROP, toeristische gebieden en gemeenten. Omdat het aantal regio's behoorlijk groot kan zijn, zeker als we kijken naar gemeentelijk niveau, beperken we ons in de weergave van de resultaten. We kiezen ervoor gebieden alleen in de resultaten op te nemen als de marges op de doelvariabele maximaal 20% zijn. Als doelvariabele hanteren we opnieuw het totaal aantal gasten.

COROP	Januari 2023			Juli 2023		
	Voldaan?	# respondenten	Marge	Voldaan?	# respondenten	Marge
Noord-Drenthe				Ja	68	0.15
Zuidoost-Drenthe	Ja	19	0.14	Ja	42	0.17
Twente				Ja	121	0.15
Veluwe	Ja	106	0.16	Ja	181	0.10
Achterhoek				Ja	111	0.15
Utrecht	Ja	85	0.07			
Kop van Noord-Holland				Ja	112	0.20
Zaanstreek				Ja	10	0.12
Groot-Amsterdam	Ja	260	0.04	Ja	269	0.04
Agglomeratie Leiden en Bollenstreek				Ja	54	0.20
Agglomeratie 's Gravenhage	Ja	51	0.15	Ja	59	0.11
Groot-Rijnmond	Ja	79	0.15	Ja	98	0.09
Overig Zeeland	Ja	90	0.10	Ja	220	0.06
Zuidoost-Noord-Brabant	Ja	53	0.19	Ja	112	0.19
Noord-Limburg				Ja	74	0.14
Zuid-Limburg	Ja	111	0.12	Ja	159	0.07
Flevoland	Ja	27	0.20	Ja	44	0.18

Tabel 5 – Overzicht van COROP-gebieden met daarbij aangegeven of ze voldoen aan de CBS outputrichtlijnen en de bijbehorende maatstaven voor het bepalen van betrouwbare statistieken. Alleen gebieden met een marge van maximaal 20% in één van beide perioden zijn opgenomen in de tabel.

In bovenstaande tabel zien we dat 17 van de 40 COROP-gebieden een relatieve betrouwbaarheidsmarge hebben van maximaal 20% in januari of juli. Dit betekent dat meer dan de helft van de COROP-gebieden hier niet aan voldoet en dus hogere marges heeft. We zien in Tabel 5 dat slechts een beperkt aantal COROP-gebieden een marge van maximaal 10% heeft. Dit betreft drie gebieden in januari en vijf gebieden in juli. Het zijn met name COROP-gebieden met veel respondenten die resulteren in een lagere marge. Een enkele uitzondering vormt Zaanstreek die met 10 respondenten toch een relatief lage marge heeft in juli. Dit kan twee mogelijke oorzaken hebben. Aan de ene kant kunnen de respondenten in dit gebied erg op elkaar lijken, waardoor de spreiding in de bezetting laag is. Anderszins kan het ook toevallig zijn dat de gerapporteerde bezetting van de respondenten in dit gebied voor juli 2023 weinig varieert. Alleen wanneer er een aanwijsbare verklaring is, bijvoorbeeld omdat accommodaties in die gemeente sterk op elkaar lijken, kan verwacht worden dat ook voor de data van andere maanden en jaren soortgelijke lage marges worden gevonden voor dit gebied.

We presenteren vervolgens ook de resultaten op het niveau van toeristische gebieden.

Toeristische gebieden	Januari 2023			Juli 2023		
	Voldaan?	# respondenten	Marge	Voldaan?	# respondenten	Marge
Hondsrug	Ja	28	0.20	Ja	67	0.09
IJsseldelta				Ja	15	0.18
Twente				Ja	91	0.20
Veluwe	Ja	110	0.16	Ja	188	0.10
Achterhoek				Ja	104	0.16
Utrecht	Ja	22	0.17	Ja	25	0.10
Walcheren				Ja	131	0.20
Midden-Brabant				Ja	50	0.17
Zuid-Oost Brabant	Ja	53	0.19	Ja	112	0.19
Waddengebied				Ja	177	0.16

Tabel 6 - Overzicht van toeristische gebieden met daarbij aangegeven of ze voldoen aan de CBS outputrichtlijnen en de bijbehorende maatstaven voor het bepalen van betrouwbare statistieken. Alleen gebieden met een marge van maximaal 20% in één van beide perioden zijn opgenomen in de tabel.

De resultaten uit Tabel 6 zijn vergelijkbaar met de resultaten op COROP-niveau. In bovenstaande tabel zijn 10 toeristische gebieden opgenomen. Dit is minder dan een derde van de 32 toeristische gebieden, wat betekent dat de niet getoonde gebieden een marge boven de 20% hebben. Als we de tabel nauwkeuriger bekijken, zien we dat vrijwel geen enkel toeristisch gebied een marge onder de 10% heeft. We zien hier dus nog sterker dan bij de COROP-gebieden dat verreweg de meeste toeristische gebieden een te hoge marge om tot betrouwbare resultaten te komen.

Laten we tenslotte de resultaten op gemeentelijk niveau bekijken.

Gemeenten	Januari 2023			Juli 2023		
	Voldaan?	# respondenten	Marge	Voldaan?	# respondenten	Marge
Ameland				Ja	32	0.16
Amstelveen	Ja	21	0.19			
Amsterdam	Ja	206	0.04	Ja	210	0.05
Arnhem				Ja	18	0.15
Haarlem				Ja	13	0.16
Haarlemmermeer	Ja	27	0.11			
Loon op Zand				Ja	15	0.17
Maastricht	Ja	27	0.14	Ja	30	0.19
Rotterdam	Ja	41	0.20	Ja	47	0.19
's Gravenhage	Ja	41	0.12	Ja	41	0.14
Texel				Ja	49	0.13
Utrecht	Ja	22	0.17	Ja	25	0.10
Voorst	Ja	21	0.19			
Zandvoort				Ja	16	0.17

Tabel 7 - Overzicht van gemeenten met daarbij aangegeven of ze voldoen aan de CBS outputrichtlijnen en de bijbehorende maatstaven voor het bepalen van betrouwbare statistieken. Alleen gemeenten die voldoen aan de CBS outputrichtlijnen én een marge hebben van maximaal 20% in één van beide perioden zijn opgenomen in de tabel.

In bovenstaande tabel zijn 14 gemeenten terug te vinden, waarvan vijf grote steden (Utrecht, Amsterdam, Rotterdam, 's-Gravenhage en Maastricht). Alleen de gemeenten Utrecht en Amsterdam hebben een betrouwbaarheidsmarge van maximaal 10% in januari of juli, wat we beschouwen als acceptabel. Dit is zo beperkt dat we moeten concluderen dat het niet mogelijk is om op gemeentelijk niveau betrouwbare toeristische statistieken te berekenen op basis van de huidige steekproef.

4.3 Effect van de steekproefgrootte

In voorgaande secties hebben we gezien dat de marges op de statistieken meestal toenemen naarmate cijfers op een lager regionaal niveau worden berekend. Dit wordt deels verklaard door het afnemende aantal respondenten per regio. Het is daarom interessant om te onderzoeken of het vergroten van de steekproef leidt tot marges van een acceptabel niveau. We analyseren in deze sectie welke toename in de omvang van de steekproef nodig is om marges

te bereiken van 5% of 10%. We gebruiken hierbij een benadering om in grote lijnen in kaart te brengen wat het effect is van de steekproefgrootte².

Regionaal niveau	Huidige steekproef	Totaal aantal gasten		Totaal aantal overnachtingen		Beide statistieken	
		Marges ≤ 10%	Marges ≤ 5%	Marges ≤ 10%	Marges ≤ 5%	Marges ≤ 10%	Marges ≤ 5%
Provincies	2832	3353	4120	3809	5441	3809	5441
COROP-gebieden	2817*	5394	6972	5690	7440	5999	7640
Toeristische gebieden	1737**	3854	5028	4092	5128	4234	5212

Tabel 8 – Een ondergrens op het benodigde aantal respondenten om de 95%-betrouwbaarheidsmarges klein genoeg te krijgen voor twee statistieken, uitgaande van de steekproef van juli 2023. (*) Dit aantal ligt lager dan op provinciaal niveau, omdat twee COROP-gebieden niet zijn meegenomen, vanwege te weinig respondenten om gegevens over vrij te geven. () De toeristische gebieden omvatten niet geheel Nederland en één toeristisch gebied had te weinig respondenten om gegevens over te publiceren.**

De huidige steekproef heeft in juli 2023 een omvang van 2832 respondenten op een totaal van 9058 geopende accommodaties. Op provinciaal niveau zijn de marges vaak al laag. De steekproef hoeft slechts met 12% te stijgen om in alle provincies een 95% betrouwbaarheidsmarge van maximaal 10% te behalen op beide statistieken. Om op beide statistieken een marge te hebben van maximaal 5% is een grotere steekproef nodig en zal het aantal respondenten moeten stijgen met 92%.

Op COROP-niveau missen we de gegevens van twee COROP-gebieden die niet kunnen worden meegenomen vanwege een te laag aantal respondenten. De cijfers uit Tabel 8 voor de COROP-gebieden vormen daarom een ondergrens, al kunnen we aannemen dat de benodigde procentuele stijging voor de twee ontbrekende gebieden vergelijkbaar is met de andere COROP-gebieden. Om marges van maximaal 10% te bereiken op COROP-niveau voor beide statistieken, zal het aantal respondenten moeten stijgen tot 6000. Dit betekent een toename van ruim 110%, dus meer dan een verdubbeling. Wanneer de gewenste marge gelijk is aan maximaal 5% is een substantieel grotere toename van de steekproef nodig. Wil dit voor beide statistieken bereikt worden op COROP-niveau, dan zal de steekproef ruim 2.7 keer zo groot moeten zijn.

Voor toeristische gebieden is de omvang van de huidige steekproef kleiner, aangezien we niet van alle provincies input hebben ontvangen bij het inventariseren van de toeristische gebieden. Deze gebieden omvatten dus niet geheel Nederland. Ook voor deze gebieden zien we echter dat de steekproef respectievelijk 2.4 of 3 keer zo groot moet zijn om te komen tot relatieve betrouwbaarheidsmarges van maximaal 10% of 5%. Merk op dat de huidige

2 We gaan hierbij uit van de formules voor de marges en de variantie zoals te vinden zijn in Bijlage D. Onder de aannamen dat de (1) steekproef-variantie niet verandert bij het vergroten van de steekproef en (2) de gemiddelde grootte van accommodaties in de steekproef en populatie aan elkaar gelijk zijn, kan de benodigde steekproefomvang als volgt worden benaderd: $m = \frac{n}{f + (1-f)\frac{n}{N}}$. Hierbij zijn n en N de omvang van respectievelijk de huidige steekproef en populatie en is de factor f gelijk aan het kwadraat van de gewenste marge gedeeld door de huidige marge.

steekproef ongeveer een derde van alle geopende accommodaties bevat. Wanneer de steekproef 2 tot 3 keer zo groot moet zijn, betekent dit dat de steekproef vrijwel alle geopende accommodaties moet bevatten. Gezien de eerdere resultaten van de marges op gemeentelijk niveau, zal het patroon van de groeiende gewenste steekproefgrootte zich doorzetten op gemeentelijk niveau om acceptabele marges te bereiken.

We merken tenslotte op dat we in het analyseren van de gegevens uit Tabel 8 geen rekening hebben gehouden met non-response. Wanneer een deel van de accommodaties in de steekproef niet reageert op de enquête van het CBS, zal een nog grotere steekproef nodig zijn om de gewenste marges te bereiken.

Concluderend kunnen we stellen dat de enorme uitbreidingen die nodig zijn in de omvang van de steekproef om te komen tot betrouwbare cijfers op het niveau van COROP-gebieden of toeristische gebieden niet haalbaar lijken. Dergelijke uitbreidingen brengen hoge kosten met zich mee. Daarnaast zijn geregionaliseerde cijfers voor het CBS niet noodzakelijk, omdat zij met het publiceren van cijfers op landelijk en provinciaal niveau al voldoen aan de verplichtingen van Eurostat. De steekproef zodanig ophogen dat betrouwbaarheid gegarandeerd is voor alle gebieden op een bepaald niveau, is daarmee onwaarschijnlijk. Het is wellicht wel mogelijk om de steekproef in beperkte mate te vergroten zodat acceptabele marges behaald kunnen worden voor meer gebieden dan nu het geval is.

5 Conclusie en discussie

5.1 Conclusie

Dit rapport onderzoekt of het mogelijk is om toeristische statistieken te regionaliseren. Uitgangspunt is de Statistiek Logiesaccommodaties die op dit moment op provinciaal niveau beschikbaar wordt gesteld door het CBS en op COROP-niveau door het NBTC. Vanwege de wens naar meer regionale statistieken, onderzoeken we of deze statistiek ook op gemeentelijk niveau kan worden gepubliceerd. We nemen hierbij de optie mee om gemeenten samen te voegen tot toeristisch relevante gebieden, wanneer betrouwbare cijfers op gemeentelijk niveau niet mogelijk zijn.

Het is alleen mogelijk betrouwbare cijfers te publiceren als aan de volgende voorwaarden is voldaan. Allereerst moet het zijn toegestaan volgens de CBS outputrichtlijnen om statistieken te publiceren. Daarnaast moet het mogelijk zijn om betrouwbare statistieken te berekenen. Hierbij speelt het aantal respondenten een belangrijke rol en de marges op de berekende statistieken. Deze marges bieden een soort bandbreedte van onzekerheid rondom de berekende bezetting.

Qua methode kiezen we in ons onderzoek voor directe schatters en we volgen daarmee de methode van het CBS. De gekozen methode is geïmplementeerd in SPSS en is in principe toepasbaar op geheel Nederland.

Eerste resultaten laten zien dat het overgrote deel van de provincies, COROP-gebieden en toeristisch relevante gebieden voldoet aan de CBS output richtlijnen en lijkt te beschikken over voldoende respondenten. Op gemeentelijk niveau geldt dit niet altijd, maar toch lijkt een kleine 20% van de gemeenten over voldoende respondenten te beschikken. Wanneer we echter naar de marges op de statistieken kijken, zien we dat al op het niveau van COROP-gebieden betrouwbare statistieken niet altijd te garanderen zijn. Dit effect wordt nog versterkt op gemeentelijk niveau. Het aandeel COROP-gebieden dat een betrouwbaarheidsmarge heeft van maximaal 10% ligt rond de 12%. Bij toeristische gebieden is er slechts één gebied en op gemeentelijk niveau vinden we slechts twee gemeenten in heel Nederland die hieraan voldoen. Hoewel dus het aantal respondenten voor een aantal grote gemeenten en toeristisch relevante gebieden behoorlijk groot lijkt te zijn, zijn de marges toch te hoog om betrouwbare statistieken te kunnen berekenen. We moeten daarom concluderen dat het op basis van de huidige steekproef niet mogelijk is om betrouwbare statistieken te presenteren op gemeentelijk niveau. Zelfs het samenvoegen van gemeenten tot toeristisch relevante gebieden biedt nauwelijks extra mogelijkheden. Alleen wanneer deze gebieden in samenstelling en omvang vrijwel overeenkomen met de COROP-gebieden, kunnen af en toe betrouwbare cijfers worden bepaald.

5.2 Discussie

Het onderzoeken van de mogelijkheid om de toeristische data te regionaliseren heeft een onbedoeld bijeffect gehad. Door de marges op de toeristische statistieken in kaart te brengen voor verschillende regionale niveaus, wordt duidelijk dat zelfs op COROP-niveau marges soms behoorlijk groot kunnen zijn. Het is daarom aan te bevelen te onderzoeken of het nog wenselijk is deze COROP-cijfers te blijven vrijgeven zonder daarbij een onzekerheidsmarge te vermelden.

Verder onderzoek zou zich eveneens kunnen richten op mogelijkheden om via andere methoden te komen tot betrouwbare cijfers op een gedetailleerder regionaal niveau dan COROP-niveau. Hiervoor zou onderzocht kunnen worden wat er bereikt kan worden met klein domein schatters. Dit onderzoek zou bij voorkeur in nauwe samenwerking met het CBS moeten gebeuren, omdat zij over ruime expertise beschikken op dit gebied.

Tenslotte bieden ook onze resultaten over de benodigde steekproefgrootte aanknopingspunten voor vervolgstappen. Als er mogelijkheden zijn om de steekproef in beperkte mate te vergroten, wordt het wellicht wel mogelijk om betrouwbare statistieken te presenteren voor meer gemeenten of toeristisch relevante gebieden dan op basis van de huidige steekproef het geval is.

Bijlage A – Begrippenlijst

Hieronder staat een overzicht van alle begrippen die worden gehanteerd binnen de Statistiek Logiesaccommodaties.

Gasten: Bezoekers die één of meer nachten achtereen verblijven in een logiesaccommodatie.

Een gast die meer dan twee maanden aaneengesloten verblijft in een logiesaccommodatie, wordt als vaste gast aangemerkt en telt niet mee voor de statistiek.

Asielzoekers en seizoenarbeiders worden niet als gast meegeteld, ook niet als ze korter dan twee maanden aaneengesloten verblijven in een logiesaccommodatie.

Per maand worden de in die maand vertrokken gasten waargenomen, ongeacht in welke maand de gast aangekomen is. Het is mogelijk dat een persoon twee of meerdere malen in een maand in dezelfde accommodatie of in verschillende accommodaties als gast wordt geteld.

Hotel: Een accommodatie met slaappleatsen voor logiesverstrekking in overwegend een- en tweepersoonskamers tegen boeking per nacht, waar afzonderlijke maaltijden, kleine etenswaren en dranken kunnen worden verstrekt aan gasten en aan passanten. Hierbij is een gast een persoon die overnacht in de desbetreffende accommodatie en een passant iemand die niet blijft overnachten. Verder kunnen nog andere diensten ter beschikking worden gesteld zoals receptie, room- en telefoonservice. De statistiek hanteert een ondergrens van minimaal 5 slaappleatsen voor deze accommodaties.

Motel: Een locatie langs de snelweg met aan elkaar geschakelde kamers met de deuren aan een parkeerplaats of gemeenschappelijke ruimte voor logiesverstrekking tegen boeking per nacht. De statistiek hanteert een ondergrens van minimaal 5 slaappleatsen voor deze accommodaties.

Pension: Een accommodatie met slaappleatsen voor logiesverstrekking in overwegend een- en tweepersoonskamers, waar afzonderlijke maaltijden, kleine etenswaren en dranken kunnen worden verstrekt aan gasten maar niet aan passanten. De statistiek hanteert een ondergrens van minimaal 5 slaappleatsen voor deze accommodaties.

Appartement met hoteldienstverlening: Appartement waarbij gedurende het verblijf het appartement schoon wordt gehouden en de bedden opgemaakt. De statistiek hanteert een ondergrens van minimaal 5 slaappleatsen voor deze accommodaties.

Jeugdaccommodatie: Jeugdhotel en jeugdherberg. Een jeugdhotel is een hotel voor overwegend jeugdige gasten met slaapgelegenheid die men niet met 'vreemden' hoeft te delen. Een jeugdherberg is een accommodatie voor overwegend jeugdige gasten met slaapgelegenheid in kamers en/of zalen die men mogelijk met 'vreemden' moet delen. De statistiek hanteert een ondergrens van minimaal 5 slaappleatsen voor deze accommodaties.

Bed & breakfast: Particuliere woning waar men kan overnachten en ontbijten. De statistiek hanteert een ondergrens van minimaal 5 slaappleatsen voor deze accommodaties.

Kampeerterrein: Een terrein of een deel van een terrein met toeristische slaappleatsen, waarop kan worden overnacht in tenten, toercaravans, kampeerauto's, tenthuisjes of trekkershutten. De statistiek hanteert een ondergrens van minimaal 4 toeristische standplaatsen voor deze accommodaties.

Huisjesterrein: Een terrein met een aantal zomerhuisjes, stacaravans, (vakantie-)bungalows of (vakantie-)appartementen, die hoofdzakelijk voor verhuur door de exploitant of beheerder van het complex beschikbaar zijn. Appartementen die verhuurd worden met hoteldienstverlening, worden niet als huisjescomplex beschouwd maar als (appartementen) hotel. Appartementen zonder hoteldienstverlening, die vaak deel uitmaken van een groter gebouw, worden beschouwd als bungalow of zomerhuisje. De statistiek hanteert een ondergrens van minimaal 10 slaappleatsen voor deze accommodaties.

Groepsaccommodatie: Accommodatie met logiesverstrekking overwegend aan personen in groepsverband (geen gezinsverband) met slaapgelegenheid in kamers, zalen, huisjes, tenthuisjes, appartementen en/of tenten die gasten mogelijk met vreemden moeten delen. De statistiek hanteert een ondergrens van minimaal 10 slaappleatsen voor deze accommodaties. Onder groepsaccommodaties worden verstaan:

- kampeerboerderijen
- (kinder)vakantiehuizen
- kamphuizen/scoutinghuizen
- natuurvriendenhuisen
- tentenkampen
- logiesaccommodaties behorende tot watersportcentra of maneges.

Bijlage B – Toeristische gebieden

Provincie	Toeristische gebieden	Gemeenten
Drenthe	Hondsrug	Aa en Hunze, Borger-Odoorn, Coevorden, Emmen
	Zuidwest Drenthe	Hoogeveen, Meppel, Westerveld, De Wolden
	Midden-Drenthe	Midden-Drenthe
	Kop van Drenthe	Assen, Noordenveld, Tynaarlo
Overijssel	Weerribben-Wieden	Steenwijkerland
	IJsseldelta	Kampen, Zwartewaterland, Zwolle
	Vechtdal	Dalfsen, Hardenberg, Ommen, Staphorst
	Salland	Deventer, Olst-Wijhe, Raalte
	Sallandse Heuvelrug	Hellendoorn, Rijssen-Holten
	Twente	Almelo, Borne, Dinkelland, Enschede, Haaksbergen, Hengelo, Hof van Twente, Losser, Oldenzaal, Tubbergen, Twenterand, Wierden
Gelderland	Rivierenland	Buren, Culemborg, Druten, Maasdriel, Neder-Betuwe, Tiel, West-Betuwe, West Maas en Waal, Zaltbommel
	Arnhem-Nijmegen	Arnhem, Beuningen, Berg en Dal, Doesburg, Duiven, Heumen, Lingewaard, Nijmegen, Overbetuwe, Renkum, Rheden, Rozendaal, Westervoort, Wijchen, Zevenaar
	Achterhoek	Aalten, Berkelland, Bronckhorst, Doetinchem, Lochem, Montferland, Oost Gelre, Oude IJsselstreek, Winterswijk, Zutphen
	Veluwe	Apeldoorn, Barneveld, Brummen, Ede, Elburg, Epe, Ermelo, Harderwijk, Hatterum, Heerde, Nijkerk, Nunspeet, Oldebroek, Putten, Scherpenzeel, Voorst, Wageningen
Utrecht	Amersfoort e.o.	Amersfoort, Bunschoten, Eemnes
	De Utrechtse Heuvelrug	Baarn, De Bilt, Leusden, Renswoude, Rhenen, Soest, Utrechtse Heuvelrug, Veenendaal, Woudenberg, Zeist
	Het Groene Hart	IJsselstein, Lopik, Montfoort, Nieuwegein, Oudewater, Vijfheerenlanden, Woerden
	Gooi & Vecht	De Ronde Venen, Stichtse Vecht
	Kromme Rijnstreek	Bunnik, Houten, Wijk bij Duurstede
	Utrecht (stad)	Utrecht
Noord-Brabant	Noord-Oost Brabant	Bernheze, Boekel, Boxtel, 's-Hertogenbosch, Land van Cuijk, Maashorst, Meierijstad, Oss, Sint-Michielsgestel, Vucht
	Noord-West Brabant	Altena, Drimmelen, Geertruidenberg, Oosterhout
	West-West Brabant	Bergen op Zoom, Halderberge, Moerdijk, Roosendaal, Steenbergen, Woensdrecht
	Zuid-Oost Brabant	Asten, Bergeijk, Best, Bladel, Cranendonck, Deurne, Eersel, Eindhoven, Geldrop-Mierlo, Gemert-Bakel, Heeze-Leende, Helmond, Laarbeek, Nuenen Gerwen en Nederwetten, Oirschot, Reusel - De Mierden, Someren, Son en Breugel, Valkenswaard, Veldhoven, Waalre

Provincie	Toeristische gebieden	Gemeenten
	Zuid-West Brabant	Alphen-Chaam, Baarle-Nassau, Breda, Etten-Leur, Rucphen*, Zundert
	Midden-Brabant	Dongen, Gilze en Rijen, Goirle, Heusden, Hilvarenbeek, Loon op Zand, Oisterwijk, Tilburg, Waalwijk
Zeeland	Schouwen-Duiveland	
	Walcheren	Middelburg, Veere, Vlissingen
	Beveland en Tholen	Borsele, Goes, Kapelle, Noord-Beveland, Reimerswaal, Tholen
	Zeeuws-Vlaanderen	Hulst, Sluis, Terneuzen

* Rucphen is gerekend tot het toeristische gebied Zuid-West Brabant. In de aangeleverde informatie vanuit de provincie Noord-Brabant was het niet duidelijk tot welke grootregio de gemeente Rucphen behoorde. Op basis van de DMO indeling is ervoor gekozen Rucphen te plaats en onder Zuid-West Brabant.

Bijlage C – CBS methode

1 Ophoging

Bij de ophoging worden de vier logiestypen (hotels, campings, huisjesterreinen en groepsaccommodaties) onderscheiden, maar worden binnen de hotels de Amsterdamse en overige apart behandeld. Voor de output worden de resultaten van deze hotels samengevoegd. Voor ieder van deze vijf logiestypen vindt een aparte ophoging plaats.

Voor heel Nederland worden individuele ophooggewichten w_{hi} bepaald per combinatie van logiestype, grootteklasse en maand, d.w.z. per (analyse)stratum h . De ophoogfactoren worden berekend als het quotiënt van het aantal accommodaties N_h dat in die maand geopend is en het aantal respondenten r_h . Er wordt binnen de strata geen onderscheid gemaakt tussen ketenbedrijven, combi's en overige accommodaties, d.w.z. $w_{hi} = w_h = N_h/r_h$ voor alle accommodaties i uit stratum h . De verschillen in insluitkansen worden dus verwaarloosd (dit gebeurde in het verleden al voor de ketenbedrijven, maar niet voor de combi's). De reden hiervoor is dat de bezettingsgraad per land van herkomst weinig systematisch verschilt tussen deze drie soorten accommodaties. We nemen aan dat de bias (mate van onzuiverheid) door deze samenvoegingen voor de schattingen van aantallen gasten en overnachtingen kleiner is dan de extra variantie die de verschillen in gewichten anders zouden opleveren. Het gevolg is dat alle accommodaties in een analysestratum 'logiestype x regio x maand x grootteklasse' hetzelfde ophooggewicht N_h/r_h krijgen. Alleen bij kleine aantallen steekproefaccommodaties per stratum worden de ophoogfactoren gebruikt, vanwege samenvoeging van grootteklassen (stap 3 in paragraaf 1.2). Bij uitzondering is een gewicht w_h groter dan 10. Het gewicht wordt dan afgekapt op 10. Ook dit heeft alleen invloed op de uitkomsten bij samenvoeging van grootteklassen [Yuri].

2 Aantallen gasten en overnachtingen per logiestype x maand x regio

We geven hieronder de schattingsmethode voor het aantal gasten per analysestratum logiestype x grootteklasse x regio x maand. Regio kan hierbij zowel de regio Nederland zijn als de provincies, de COROP-gebieden e.d.

Per analysestratum logiestype x grootteklasse x regio x maand (h):

y_{hi} = score op doelvariabele y (aantal gasten of overnachtingen per land van herkomst) van accommodatie (respondent) i van stratum h ;

x_{hi} = capaciteit van accommodatie i van stratum h (per dag);

D_h = aantal dagen in de maand bij stratum h ;

$R_{hi} = \frac{y_{hi}}{x_{hi}D_h}$ = bezetting in accommodatie i van stratum h .

Per analysestratum h vinden de volgende stappen plaats.

- 1 Voor iedere doelvariabele y wordt de gemiddelde bezetting per bed geschat als het quotiënt van het aantal gasten of overnachtingen en de capaciteit bij de respondenten.

$$\hat{R}_h = \frac{\sum_i w_{hi} y_{hi} \frac{1}{D_h}}{\sum_i w_{hi} x_{hi} \frac{1}{D_h}}. \quad (1)$$

De accommodaties worden dus in teller en noemer gewogen met de ophoogfactoren $w_{hi} = N_{(h)}/r_h$. Dit heeft geen invloed op de uitkomst omdat binnen het analysestratum geen onderscheid wordt gemaakt tussen ketenbedrijven, combi's en overige accommodaties, en alle accommodaties dus hetzelfde gewicht $w_{hi} = w_h$ hebben. Er geldt dus

$$\hat{R}_h = \frac{\sum_i w_{hi} y_{hi} \frac{1}{D_h}}{\sum_i w_{hi} x_{hi} \frac{1}{D_h}} = \frac{\sum_i y_{hi} \frac{1}{D_h}}{\sum_i x_{hi} \frac{1}{D_h}} \quad (2)$$

Alleen bij samenvoeging van grootteklassen vervalt het laatste gelijkteken.

- 2 Voor de niet-respondenten (niet-geënquêteerden en non-respondenten) worden de schattingen $\hat{R}_h$ als bezetting $\hat{R}_{hi}$ geïmputeerd. Vervolgens worden de aantallen gasten en overnachtingen berekend als product van de corresponderende bezetting en de capaciteit

$$\hat{y}_{hi} = \hat{R}_{hi} x_{hi} D_h. \quad (3)$$

De respondenten behouden de door hun opgegeven aantallen gasten en overnachtingen.

- 3 Wanneer een analysestratum minder dan 5 respondenten heeft, wordt het samengevoegd met het stratum met een hogere of lagere grootteklasse: GK 1 en 2 kunnen worden samengevoegd, en GK 3 en 4. De te imputeren bezetting voor de niet-waargenomen accommodaties wordt dan berekend over de beide grootteklassen (binnen logiestype x regio x maand), waarbij de accommodaties de in paragraaf 5.1 berekende ophoogfactoren per oorspronkelijke GK als gewichten krijgen. In dat geval wordt formule (1) toegepast in plaats van (2). Wanneer echter minstens 40% van de accommodaties zijn waargenomen, wordt het stratum niet samengevoegd.
Als GK1+2 of GK3+4 minder dan 5 respondenten heeft, worden alle grootteklassen samengevoegd.
- 4 Indien binnen een regio (per type x maand) minder dan 8 accommodaties zijn waargenomen, dan wordt de bezetting overgenomen van heel Nederland.

5 De resultaten worden per logiestype x regio x maand gesommeerd over de grootteklassen h:

$$\hat{Y} = \sum_{st,pr} y_{hi} + \sum_{pop \setminus st} \hat{y}_{hi} . \quad (4)$$

Het eerste somteken betreft de in de steekproef waargenomen aantallen overnachtingen (gasten), en het tweede somteken de geïmputeerde aantallen overnachtingen (gasten) voor de niet-waargenomen accommodaties. Deze formule levert het geschatte totaal aantal overnachtingen (en gasten) per maand per logiestype voor iedere regio (Nederland, provincie, COROP etc.).

Kwartaal- en jaarcijfers worden van formule (4) afgeleid door sommatie over de maanden.

De bovenstaande procedure wordt ook gevolgd voor het aantal sterren in hotels.

3 Aggregatie over regio's en correctiefactoren

De provincies vormen de basis voor de landelijke cijfers. De totalen voor Nederland worden berekend door sommatie over de 12 provincies. Amsterdam is hierbij als onderdeel van de Provincie Noord-Holland apart geanalyseerd. Na berekening van de geschatte provincietotalen volgens formule (4) volgt dus

$$\hat{Y}_{Ned} = \sum_{prov} \hat{Y}_{prov} \quad (5)$$

Het volgens (4) berekende provincietotaal $\hat{Y}_{prov}$ (gasten en overnachtingen naar herkomstland) zal doorgaans afwijken van de som van de geschatte totalen voor de tot de provincie behorende COROPs, $\sum_{corop \in prov} \hat{Y}_{corop}$. Om consistentie te verkrijgen wordt een correctiefactor berekend als quotiënt van beide, en worden de COROP-totalen $\hat{Y}_{corop}$ hiermee vermenigvuldigd:

$$\hat{\hat{Y}}_{corop} = \hat{Y}_{corop} (\hat{Y}_{prov} / \sum_{corop \in prov} \hat{Y}_{corop}) . \quad (6)$$

De COROP-totalen worden zo geijkt aan de provincietotalen.

Voor de meeste andere kenmerken gaan de categorieën over de grenzen van provincies heen. De categorietotalen worden dan geijkt aan totaal Nederland, d.w.z. aan de som over de provincies.

Bijlage D – Marges

Schriftelijke toelichting door CBS, augustus 2024

Er wordt bij de statistiek Logiesaccommodaties gebruik gemaakt van de bezetting bij het maken van schattingen. Hieronder wordt ingegaan op het maken van schattingen in zo'n scenario en het bepalen van de bijbehorende marges op basis van een quotiëntschatting. De theorie die hier beschreven wordt, is afgeleid van de theorie in Särndal et al (1992).

Belangrijke doelvariabelen voor het onderzoek naar logiesaccommodaties zijn het populatie-totaal per maand voor het aantal gasten en het aantal overnachtingen per accommodatietype. Daarnaast is het gebruikelijk om een onderscheid te maken tussen Nederlandse en niet-Nederlandse gasten (met eventueel nog een verdere uitsplitsing naar herkomstland van de gasten). Deze doelvariabelen geven we aan met de letter Y , bijvoorbeeld het aantal overnachtingen per accommodatietype in Nederland. Laat $\hat{y}$ een schatting zijn voor Y . Y wordt geschat aan de hand van diverse analysestrata, gebaseerd op grootteklasse, accommodatietype, regio en maand. Hierbij wordt per analysestratum gebruik gemaakt van een quotiëntschatting met als hulpvariabele de capaciteit. Er is namelijk een sterk verband tussen het aantal overnachtingen en de beschikbare capaciteit van een accommodatie. Het quotiënt van de twee variabelen lijkt bovendien minder te fluctueren dan de doelvariabele zelf, waardoor het gebruik van de quotiëntschatting een positieve invloed heeft op het maken van de schattingen.

In het vervolg van dit document wordt met stratum het analysestratum bedoeld. Laat $h = 1, \dots, H$ de verschillende strata zijn en laat U_h de populatie van de accommodaties in stratum h weergeven. Laat ρ_h de verzameling van responderende accommodaties uit stratum h weergeven. Laat n_h het aantal responderende accommodaties in stratum h zijn en laat N_h het aantal accommodaties in de populatie in stratum h zijn. Laat y_{hk} de waarde van doelvariabele voor de (responderende) accommodatie k in stratum h zijn. Laat x_{hk} de capaciteit (per dag) van accommodatie k in stratum h zijn, laat D_h het aantal dagen in de maand zijn behorend bij de maand corresponderend met stratum h en laat R_{hk} de gemiddelde bezetting per dag bij accommodatie k in stratum h zijn, i.e.

$$R_{hk} = \frac{y_{hk}}{x_{hk} D_h},$$

waarbij y_{hk} het aantal overnachtingen voor respondent k op maandbasis is en x_{hk} de capaciteit per dag. Merk op dat R_{hk} alleen voor de respondenten bepaald kan worden. Als met $\hat{R}_h$ de schatting voor de gemiddelde bezetting per dag voor stratum h wordt genoteerd, dan geldt dat

$$\hat{R}_h = \frac{\sum_{k \in \rho_h} y_{hk} w_{hk}}{\sum_{k \in \rho_h} x_{hk} D_h w_{hk}},$$

waarbij w_{hk} een ophoogfactor is met $w_{hk} = \min(N_h/n_h, 10)$. Over het algemeen geldt dat

$$\hat{R}_h = \frac{\sum_{k \in \rho_h} y_{hk} w_{hk}}{\sum_{k \in \rho_h} x_{hk} D_h w_{hk}} = \frac{\sum_{k \in \rho_h} y_{hk}}{\sum_{k \in \rho_h} x_{hk} D_h},$$

tenzij er strata worden samengevoegd, zoals bij het samenvoegen van grootteklassen.

Vervolgens wordt voor niet-respondenten de geschatte doelvariabele $\hat{y}_{hk}$ als volgt bepaald

$$\hat{y}_{hk} = \hat{R}_h x_{hk} D_h.$$

Vervolgens kan het populatietotaal voor stratum h , $\hat{y}_h$, als volgt geschat worden

$$\hat{y}_h = \sum_{k \in \rho_h} y_{hk} + \sum_{k \in U_h \setminus \rho_h} \hat{y}_{hk},$$

waaruit een schatter voor Y eenvoudig volgt: $\hat{y} = \sum_{h=1}^H \hat{y}_h$.

Het is verder belangrijk om de precisie van de verkregen schatter te kennen. De precisie kan worden gekwantificeerd via de relatieve 95%-betrouwbaarheidsmarge $M(\hat{y})$. Deze kan worden berekend door:

$$M(\hat{y}) = \frac{1,96 \cdot \sqrt{\text{Var}(\hat{y})}}{\hat{y}}.$$

De marge is evenredig met de wortel van de variantie van de schatter. Hoe kleiner de variantie, hoe kleiner de marge en dus hoe preciezer de schatting. De variantie $\text{Var}(\hat{y})$ hangt af van alle waarden van de doelvariabele in de populatie. Deze kan niet worden berekend, omdat de waarde van de doelvariabele alleen bekend is voor de responderende accommodaties, en zal dus moeten worden geschat op basis van de respons. Om tot een schatting voor de variantie van het totaal over alle strata te komen, is het handig om de variantie per stratum, $\text{Var}(\hat{y}_h)$, te kunnen schatten. Deze kan, voor strata die niet samengevoegd worden, benaderd worden met de volgende formule

$$\widehat{\text{Var}}(\hat{y}_h) = \left(\frac{\sum_{k \in U_h} x_{hk} / N_h}{\sum_{k \in \rho_h} x_{hk} / n_h} \right)^2 N_h^2 \frac{1 - n_h / N_h}{n_h} S_{esh}^2,$$

zie Särndal (1992), waarbij

$$S_{esh}^2 = \frac{1}{n_h - 1} \sum_{k \in \rho_h} (y_{hk} - \hat{R}_h x_{hk})^2,$$

de som van het kwadraat van de residuen. Deze twee formules veranderen op het moment dat strata worden samengevoegd. In zulke gevallen wordt de volgende benadering gehanteerd voor de geschatte variantie per gecombineerd stratum h

$$\widehat{\text{Var}}(\hat{y}_h) = \left(\frac{\sum_{k \in U_h} x_{hk} / N_h}{\sum_{k \in \rho_h} x_{hk} / n_h} \right)^2 N_h^2 \frac{1 - n_h / N_h}{n_h} \tilde{S}_{esh}^2,$$

waarbij

$$\tilde{S}_{esh}^2 = \frac{1}{n_h - 1} \sum_{k \in \rho_h} (w_{hk} (y_{hk} - \hat{R}_h x_{hk}))^2,$$

waarbij de gewichten zowel gebruikt worden bij het bepalen van $\hat{R}_h$ als bij de som van het kwadraat van de residuen.

Met behulp van de geschatte variantie van $\hat{y}_h$ kan vervolgens de 95%-betrouwbaarheidsmarge per stratum worden geschat en deze is gelijk aan

$$\hat{M}(\hat{y}_h) = \frac{1,96 \cdot \sqrt{\widehat{Var}(\hat{y}_h)}}{\hat{y}_h}.$$

De variantie voor het totaal over alle strata samen is verder simpelweg $\widehat{Var}(\hat{y}) = \sum_{h=1}^H \widehat{Var}(\hat{y}_h)$

en de geschatte marge voor het totaal is $\hat{M}(\hat{y}) = \frac{1,96 \cdot \sqrt{\widehat{Var}(\hat{y})}}{\hat{y}}.$

Referenties

- C.-E. Särndal, B. Swensson & J. Wretman (1992) Model assisted survey sampling. Springer Verlag, New York Inc. Springer series in Statistics.

Bijlage E – Implementatie

Zoals beschreven in Sectie 3.1.3 is de methode van het CBS soms op meerdere manieren te interpreteren, waardoor het onduidelijk is welke implementatiekeuzes het CBS zelf gemaakt heeft. Onduidelijkheid ontstond met name rond de vragen *wanneer* analysestrata worden samengevoegd en *hoe* deze samenvoeging dan plaatsvindt. Deze vragen zijn aan het CBS voorgelegd, waarna enige aanpassingen in onze implementatie zijn doorgevoerd. In deze bijlage lichten we onze implementatie toe die zoveel mogelijk aansluit bij de implementatie van het CBS, maar dus op bepaalde punten wel afwijkt van de door het CBS gedocumenteerde methode uit Bijlage C.

Afwijkende respons

Wanneer een gerapporteerde bezetting van een respondent te veel afwijkt van de overige respondenten, wordt die respons beschouwd als een uitbijter. Het CBS hanteert als criterium dat uitbijters meer dan 5 standaarddeviaties afwijken van de gemiddelde bezetting voor gasten of meer dan 15 standaarddeviaties voor overnachtingen. In de microdata is aangegeven welke respondenten als uitbijter worden beschouwd. We nemen deze respons van uitbijters in onze implementatie wel mee voor hun eigen bezetting te, maar nemen deze niet mee in het bepalen van de geschatte bezetting voor niet-respondenten.

Naast afwijkende respons, valt het ook op dat in sommige strata alle of veel respondenten een respons van 0 rapporteren. Dat wil zeggen dat ze aangeven in die maand geen gasten of overnachtingen te hebben gehad. Het CBS geeft aan deze respons niet mee te nemen in hun implementatie. Wij kiezen ervoor deze respons wel te behouden, omdat ook een waarde van 0 informatie bevat, bijvoorbeeld omdat er in een bepaalde maand geen gasten uit een zeldzaam herkomstland geweest zijn. We nemen aan dat onterecht gerapporteerde waarden van 0 worden uitgefilterd in de definitie van de uitbijters. Het zou nauwkeuriger onderzoek vereisen om precies te bepalen wanneer een gerapporteerde waarde van 0 als onbruikbaar beschouwd zou moeten worden.

Terminologie introduceren; partners en familie

Wanneer een analysestratum te weinig respondenten heeft, wordt het soms samengevoegd met een ander analysestratum. Hierbij wordt gekeken naar de grootteklasse van de accommodaties en wordt onderscheid gemaakt tussen 'grote' en 'kleine' accommodaties. Wij nemen aan dat voor hotels accommodaties met grootteklassen 5 en 6 worden beschouwd als *groot* en accommodaties met grootteklassen 2, 3 of 4 als *klein*. Voor de overige logiestypen beschouwen we grootteklasse 4 als *groot* en grootteklassen 2 en 3 als *klein*.

Om nauwkeuriger te definiëren hoe we analysestrata samenvoegen, introduceren we de volgende terminologie. Zij h een willekeurig stratum. We definiëren de familiestrata van h als volgt: alle familiestrata van h gaan over dezelfde regio, logiestype, tijdsperiode en herkomstland als h , maar niet over dezelfde grootteklasse. Naast de familiestrata introduceren we ook

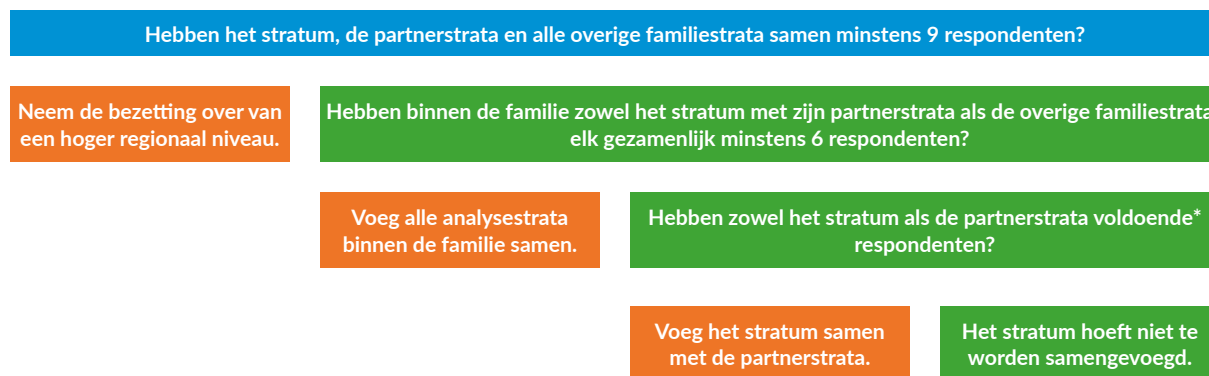
de *partnerstrata* van ieder willekeurig stratum h op de volgende manier. Partnerstrata van stratum h zijn familiestrata van h , maar hebben een grootteklasse in dezelfde categorie (groot of klein) als de grootteklasse van stratum h . Tabel A geeft een overzicht van de familierelaties tussen grootteklassen.

Logiestype	Grootteklasse van stratum	Grootteklassen van partnerstrata	Grootteklassen van overige familiestrata
Hotels	2	3, 4	5, 6
	3	2, 4	5, 6
	4	2, 3	5, 6
	5	6	2, 3, 4
	6	5	2, 3, 4
Overig (campings, huisjesterreinen en groepsaccommodaties)	2	3	4
	3	2	4
	4	-	2,3

Tabel A – De partnerrelaties en familierelaties tussen de verschillende grootteklassen .

Samenvoegen van strata

Aan de hand van de gedefinieerde partnerstrata en de familiestrata, kunnen we bespreken hoe we strata samenvoegen wanneer er te weinig respondenten zijn. We hanteren daarbij de onderstaande beslisstructuur die per analysestratum wordt doorlopen. Wordt een vraag met 'ja' beantwoord, dan vervolg je de structuur naar rechtsonder (groen). Bij het antwoord 'nee', dan ga je een stap naar linksonder (oranje).



Tabel B – Beslisstructuur voor het samenvoegen van analysestrata.{*} Een stratum heeft voldoende respondenten als er óf minstens 6 respondenten zijn, of de respondenten in het stratum dekken minstens 60% van de capaciteit in dat stratum.

We merken op dat de verschillende grenswaarden in Tabel B correcties zijn op de methode in Bijlage C. Ook de definitie wanneer een stratum voldoende respondenten is iets aangepast.

Samenvoegen op overkoepelend niveau

De beslisstructuur in Tabel B laat zien dat er minstens 9 respondenten in een familie nodig zijn om tot een betrouwbare bezetting te kunnen komen. Is dat niet het geval, dan wordt de bezetting overgenomen van een hoger regionaal niveau. Voor het hogere regionale niveau nemen we het landelijk niveau in het geval dat we statistieken op provinciaal niveau berekenen. Voor statistieken op de overige regionale niveaus (COROP, toeristische gebieden, gemeenten) nemen we het provinciale niveau als het hogere regionale niveau.

De marges moeten in dat geval ook anders berekend worden. We gaan hierbij uit van de steekproefstandaarddeviatie op provinciaal niveau en imputeren die voor alle strata binnen de betreffende familie. Dit resulteert voor al deze strata in dezelfde steekproefstandaarddeviatie, omdat we in dit geval op provinciaal niveau alle grootteklassen samennemen. We vervolgen verder de reguliere berekening voor de marges.



NBTC
Nederlands Bureau
voor Toerisme &
Congressen



UNIVERSITY
OF APPLIED SCIENCES



Centre of Expertise
leisure, tourism & hospitality



**Centraal Bureau
voor de Statistiek**