

Hoofdstuk 1

WAT IS EEN STATISTIEK?

Voor we kunnen bespreken hoe we statistieken produceren, moeten we eerst afspreken wat we precies verstaan onder een statistiek. Wat beschouwen we als een statistiek en wat niet? Wat hebben we allemaal nodig om te spreken van een statistiek? Welke informatie is essentieel om een statistiek te produceren? Hoe bepalen we de kwaliteit van een statistiek? Het antwoord op al deze vragen komt aan bod in dit hoofdstuk.

1.1 STATISTIEK ALS EEN CIJFER

Als je mensen vraagt wat ze verstaan onder een statistiek, zullen velen een statistiek op de eerste plaats definiëren als een cijfer. Die definitie is echter niet volledig, een statistiek houdt meer in dan enkel een cijfer. Statistiek gaat namelijk niet alleen over het berekenen van cijfers, het draait ook om de interpretatie die eraan wordt gegeven. In deze paragraaf zullen we zien dat de interpretatie van een statistiek wordt aangestuurd door gebruikersbehoeften en dat dit leidt tot twee soorten definities van een statistiek: de conceptuele en de operationele definitie. We zullen ook zien dat deze definities de kwaliteit van de statistiek bepalen en hoe dat gebeurt. Tot slot helpt deze discussie om af te bakenen wat experimentele statistieken zijn.

1.1.1 Gebruikersbehoeften

Een statistiek verwijst steeds naar een cijfer, maar een cijfer alleen bepaalt nog geen statistiek. Neem bijvoorbeeld het cijfer “525 935”. Dit getal is een voorbeeld van een gepubliceerde statistiek in Vlaanderen. Toch zegt dit cijfer op zichzelf zeer weinig. Waarvan zijn er 525 935? Wat stelt dit getal

1. Wat is een statistiek?

voor? Zonder aanvullende informatie heeft het cijfer geen betekenis. Een *statistiek* bestaat daarom altijd uit twee elementen: het cijfer zelf én de betekenis die eraan wordt gegeven (**Wild2018What**). Beide zijn even belangrijk en vormen één geheel. In dit geval betekent het cijfer 525 935: “het aantal inwoners van de gemeente Antwerpen op 1 januari 2019”.

Als we een statistiek publiceren met een bepaalde betekenis, is dat steeds als antwoord op specifieke *gebruikersbehoeften*. Als we een cijfer publiceren over het aantal inwoners in Antwerpen in 2019, betekent dit dat sommige beleidsmakers, onderzoekers of burgers daadwerkelijk graag willen weten hoeveel inwoners Antwerpen telde in 2019. Bijvoorbeeld, een Vlaams decreet kan verwijzen naar het aantal inwoners in de Vlaamse gemeenten en zijn we als Statistiek Vlaanderen verplicht het aantal inwoners in Antwerpen te publiceren zodat de Vlaamse regering hierop beleid kan uitstippen. Antwerpse beleidsmakers weten daarnaast waarschijnlijk ook graag hoeveel inwoners hun gemeente telt zodat ze hun diensten hierop kunnen afstemmen. Demografen gebruiken dit cijfer, samen met gegevens uit andere jaren en gemeenten, om bevolkingsgroei en spreiding in Vlaanderen te onderzoeken. Burgers willen weten hoeveel inwoners Antwerpen telt om te beoordelen of bepaalde Vlaamse subsidies aan de stad gerechtvaardigd zijn of niet.



Statistieken zijn niet altijd cijfers, het kunnen ook kwalitatieve observaties zijn of zelfs complexere “informatie-objecten”. Zo publiceert Agentschap Binnenlands Bestuur bijvoorbeeld per gemeente een statistiek over de top 10 andere gemeenten waarnaar kinderen heen pendelen voor school. Voor de stad Antwerpen kan de observatie op deze statistiek bijvoorbeeld de volgende tekstuele vector zijn:

“Antwerpen, Brasschaat, Lier, Brussel, Beveren, Mechelen, ...”.

Voor de duidelijkheid verwijzen we hieronder naar enkelvoudige cijfers. Cijfers zijn immers veruit de meest voorkomende vorm van statistieken.

1.1.2 Conceptuele definitie

Elke statistiek bestaat dus uit een cijfer en de betekenis die een doorsnee gebruiker aan dat cijfer geeft. Die betekenis noemen we de *conceptuele definitie* (**neuman2014Social; refs toevoegen**). In ons voorbeeld is “het aantal inwoners van de gemeente Antwerpen op 1 januari 2019” dus gewoon de conceptuele definitie.

Een conceptuele definitie bevat doorgaans meerdere delen of componenten. Bij de statistiek over het aantal inwoners in Antwerpen in 2019 kunnen we bijvoorbeeld drie componenten onderscheiden. Ten eerste kijken we naar een gemeente, namelijk Antwerpen. Ten tweede, kijken we naar een bepaald jaar, namelijk 2019. Ten derde kijken we naar een grootheid

die we willen meten, namelijk de bevolkingsgrootte. Deze componenten noemen we *concepten* (**clark2021Social**). Een conceptuele definitie bevat dus meestal meerdere concepten die we willen meten of gebruiken.

De afbakening van concepten in een conceptuele definitie is niet altijd duidelijk en is voor een stuk arbitrair. Zo kan je bijvoorbeeld cijfers publiceren over het aantal inwoners in Antwerpen per bevolkingsgroep (mannen, vrouwen, 0- tot 18-jarigen, 19- tot 64- jarigen, 65+'ers, mensen met Belgische nationaliteit, mensen met buitenlandse nationaliteit, ...). Spreken we in deze definitie van één concept “bevolkingsgroep” of hebben we hier te maken met meerdere concepten zoals “geslacht”, “leeftijd” en “nationaliteit”? Op deze vraag is er geen éénduidig goed of fout antwoord. Om het productieproces voor deze statistieken te organiseren in de praktijk zullen we echter wel een afspraak moeten maken over welke optie we volgen binnen onze organisatie. Dit wordt verder uitgelegd in hoofdstuk ??.

1.1.3 Operationele definitie

De conceptuele definitie is meestal niet concreet over wat een cijfer precies weergeeft. In de zin “de Antwerpse bevolkingsgrootte neemt toe” is bijvoorbeeld “bevolkingsgrootte” enkel een abstract idee. Het woord vertelt immers niet wie je nu precies meetelt en wie niet om de grootte van de bevolking te bepalen. De conceptuele definitie van een cijfer vertelt dus weinig over hoe het cijfer precies werd berekend en exact geïnterpreteerd kan worden. Naast de conceptuele definitie heeft elke statistiek daarom ook een *operationele definitie*. Deze operationele definitie beschrijft tot in detail hoe het cijfer exact is gemeten en geeft zo de precieze betekenis ervan weer. De operationele definitie is daarom meestal veel uitgebreider dan de conceptuele definitie.

Neem opnieuw het cijfer 525 935 om het aantal inwoners van Antwerpen in 2019 weer te geven. De volledige operationele definitie van dit cijfer luidt:

“De grootte van de wettelijke bevolking op 1 januari 2019, 0.00 uur, van de gemeente Antwerpen (NIS 11002 in 2019). De wettelijke bevolking telt alle inschrijvingen in het bevolkingsregister en het vreemdelingenregister. Het bevolkingsregister bevat alle Belgen en buitenlanders die gemachtigd zijn tot vestiging op het Belgisch grondgebied. Het vreemdelingenregister bevat alle buitenlanders die toegelaten of gemachtigd zijn tot een verblijf van meer dan 3 maanden op het Belgisch grondgebied, hetzij voor bepaalde of onbepaalde duur. Bepaalde categorieën buitenlanders (vb. diplomatiek en consulair personeel) zijn vrijgesteld van inschrijving in de bevolkingsregis-

[1. Wat is een statistiek?](#)

ters. In sommige gevallen kunnen zij op eigen vraag wel ingeschreven worden. Enkel in dat geval worden zij meegerekend in de bevolkingscijfers.

Deze bevolkingsgrootte wordt gedefinieerd en aangeleverd door Statbel op basis van het Rijksregister van de natuurlijke personen, waar het bevolkingsregister en het vreemdelingenregister deel van uitmaken.

Het Rijksregister omvat verder ook een wachtregister voor asielzoekers en een wachtregister voor EU burgers. Het wachtregister voor asielzoekers bevat alle verzoekers om internationale bescherming die worden ingeschreven door de Dienst Vreemdelingenzaken (DVZ). In 1995 besliste Statbel de personen in dit wachtregister niet meer mee te tellen bij de wettelijke bevolking. Pas nadat asielzoekers worden overgeschreven van het wachtregister naar het bevolkingsregister of het vreemdelingenregister, worden zij opgenomen in de bevolkingsstatistieken van Statbel. Zo'n overschrijving naar het bevolkingsregister of het vreemdelingenregister gebeurt na erkenning als vluchteling, na toekenning van een statuut subsidiaire bescherming, of na verwerving van een verblijfsvergunning om een andere reden.

Verder bevat het Rijksregister ook een wachtregister voor EU burgers in afwachting van woonstcontrole. Deze personen worden evenmin meegeteld bij de wettelijke bevolking. Pas na woonstcontrole worden deze personen overgeschreven naar het vreemdelingenregister en worden zij meegeteld in de wettelijke bevolking.”

Zoals je ziet, de operationele definitie is al een hele mond vol, zelfs voor een redelijk eenvoudige statistiek als de grootte van een bevolking. De lengte van een operationele definitie kan dan ook sterk variëren. Voor een eenvoudig cijfer afgeleid uit officiële registers, zoals de bevolkingsgrootte in Vlaamse gemeenten, kunnen enkele korte paragrafen volstaan. Bij complexere statistieken, zoals bijvoorbeeld de gemiddelde opinie van inwoners gemeten via een bevraging, is daarentegen een uitgebreid rapport nodig met alle details over, onder andere, het enquête-design, de steekproef trekking, het ontwerp van de vragenlijst, de opvolging van respondenten, statistische correcties voor meet- en selectiefouten en de statistische analysemethoden.

Operationalisering van de concepten

Merk op dat de operationele definitie een duidelijke operationalisering bevat van elk concept afzonderlijk. Het concept “gemeente” wordt geoperationaliseerd als de gemeenten volgens de NIS-code indeling in 2019. Het concept “jaar” wordt geoperationaliseerd als de toestand op 1 januari, 0.00 uur, van een kalenderjaar. Het concept “bevolkingsgrootte” wordt geoperationaliseerd als het aantal personen in de wettelijke bevolking met bijhorende precieze definitie van wie meegerekend wordt in de wettelijke bevolking en wie niet.

De eenvoud van een statistiek wordt bepaald door hoe concreet of abstract de concepten in de conceptuele definitie zijn. Over het algemeen wordt er een onderscheid gemaakt tussen abstracte concepten en concrete concepten (**referentie toevoegen**). Een *concreet concept* verwijst naar een fenomeen dat direct observeerbaar of eenvoudig meetbaar is, vaak zonder complexe interpretatie of afleiding. Een voorbeeld van een concreet concept is het aantal huishoudens in een gemeente. Dit is een concreet concept want je kan het rechtstreeks vaststellen op basis van administratieve gegevens (bv. bevolkingsregister). Er is bovendien weinig discussie over wat een huishouden is, zeker als er een duidelijke juridische of administratieve definitie wordt gehanteerd.

Een *abstract concept* verwijst daarentegen naar een idee of fenomeen dat niet rechtstreeks waarneembaar is. Het heeft een zekere mate van theoretische of interpretatieve lading en vereist meestal operationalisatie om meetbaar te worden. Een voorbeeld van een abstract concept is sociaal isolement. Het verwijst immers naar een toestand waarin iemand weinig of geen sociale contacten heeft of zich sociaal buiten gesloten voelt. Dit kan je niet rechtstreeks meten; je moet het operationaliseren, bijvoorbeeld via een vragenlijst over, onder andere, het aantal sociale contacten, subjectieve gevoelens van eenzaamheid en deelname aan sociale activiteiten.

In de praktijk kan je concepten eerder op een continue schaal plaatsen op het vlak van abstractheid en concreetheid. Het ééne concept is al abstracter en vraagt om een diepgaandere operationalisering terwijl een ander concept redelijk concreet is en met weinig moeite geoperationaliseerd kan worden. Zoals ons voorbeeld over de bevolkingsgrootte in Antwerpen aantoonde hebben de meeste concepten al een zekere mate van abstractie. Zelfs een eenvoudig concept als “bevolkingsgrootte” bevat al een zekere abstractie. De “grootte van de wettelijk bevolking” is duidelijk een concreter concept en kan worden gebruikt als de operationalisering van het abstractere concept “bevolkingsgrootte”.

1. Wat is een statistiek?

Conceptuele versus operationele definitie

Operationalisering betekent eigenlijk niets meer dan een vertaling van abstracte concepten naar concretere concepten. Tijdens die vertaalslag moeten echter verschillende keuzes worden gemaakt. Die keuzes kunnen er echter voor zorgen dat andere cijfers worden bekomen. Het cijfer 525 935 verwijst bijvoorbeeld naar de bevolking op 1 januari 2019. Een vergelijkbaar maar ander cijfer kon worden berekend als “jaar” op een andere wijze werd geoperationaliseerd, bijvoorbeeld door een andere dag in het jaar te kiezen in plaats van 1 januari of door het jaargemiddelde te berekenen over alle dagen. Bovendien verwijst 525 935 naar de wettelijke bevolking gerapporteerd door Statbel en in die wettelijke bevolking worden personen uit de wachtrechtregisters niet meegeteld. Een vergelijkbaar maar ander cijfer had ook kunnen worden berekend als bevolking op een andere manier werd geoperationaliseerd, bijvoorbeeld als de verblijvende bevolking waarin personen uit de wachtrechtregisters wel worden meegeteld. Verder verwijst het cijfer 525 935 naar de gemeente Antwerpen zoals gedefinieerd door NIS code 11002 in 2019. Door fusies, splitsingen of grensaanpassingen verandert het grondgebied van sommige gemeenten over de tijd; bijvoorbeeld, in 2025 fuseerde Antwerpen met Borsbeek, waardoor het grondgebied veranderde. We zouden opnieuw een vergelijkbaar maar ander cijfer kunnen berekenen voor alle personen in 2019 die op dat moment woonden in wat we nu als de gemeente Antwerpen beschouwen, inclusief Borsbeek dus.

Het Antwerpse voorbeeld illustreert dat zelfs relatief eenvoudige conceptuele definities vaak vertaald kunnen worden naar meerdere operationele definities, elk met eigen operationele keuzes. Die keuzes kunnen leiden tot verschillende cijfers. Eurostat, bijvoorbeeld, rapporteert niet de wettelijke bevolking maar de verblijvende bevolking en publiceert daarom andere cijfers dan Stabel en Statistiek Vlaanderen over bevolkingsgroottes, ook al interpreteren de meeste gebruikers zowel de Europese als de Belgische/-Vlaamse cijfers gewoon als de “bevolkingsgrootte”.

Het eerste doel van de operationele definitie is daarom om de verschillende concepten te concretiseren en duidelijkheid te scheppen over de gemaakte keuzes om een statistiek te berekenen. Hierdoor kan iedereen de cijfers reproduceren en creëer je transparantie. Als twee onderzoekers met dezelfde operationele definitie verschillende cijfers bekomen, is de definitie onvolledig of onnauwkeurig omdat bepaalde concepten nog te abstract zijn.



In theorie leidt een operationele definitie steeds tot één exact resultaat. In de praktijk is dit echter niet steeds het geval. Als het productieproces bijvoorbeeld een willekeurige steekproef of simulatie bevat, kan het cijfer

variëren. Toch blijft de operationele definitie als idee ook in zo'n situaties overeind. Twee onderzoekers zouden in zo'n situatie op basis van dezelfde operationele definitie gemiddeld steeds tot hetzelfde cijfer moeten komen wanneer zij hun productieproces blijven herhalen. Het gebied van de inferentiële statistiek biedt hiervoor het theoretisch kader, maar dat gaat voorbij het doel van deze handleiding.

Het tweede doel van de operationele definitie is duidelijkheid te schaffen over de vergelijkbaarheid van cijfers. Je kan bijvoorbeeld niet zomaar de grootte van de wettelijke Antwerpse bevolking in 2019 vergelijken met de grootte van de verblijvende Antwerpse bevolking in 2020. Dit zou immers kunnen leiden tot onjuiste conclusies over bevolkingsgroei of -krimp, ook al worden beide cijfers conceptueel geïnterpreteerd als de Antwerpse bevolkingsgrootte.

Je kan je afvragen waarom we eigenlijk conceptuele definities opstellen als ze, in tegenstelling tot de operationele definitie, niet concreet en voldoende accuraat zijn? Toch zijn er goede redenen om voor elke statistiek een goede conceptuele definitie te bepalen. Ten eerste is de conceptuele definitie een pak handiger in gebruik dan de operationele definitie wanneer de cijfers worden gebruikt voor beleidsbeslissingen of te rapporteren over onderzoek. Statistiekgebruikers beperken zich doorgaans tot de conceptuele definitie om cijfers te benoemen en te interpreteren, ook al zijn er binnen dezelfde conceptuele definitie meerdere operationele definities mogelijk. Iemand die praat over het aantal inwoners in Antwerpen in 2019 zal dit eenvoudigweg de "Antwerpse bevolkingsgrootte" noemen en niet de hele operationele definitie afraten. Ten tweede hebben de nuances in de operationele definitie vaak weinig invloed op beleidskeuzes of onderzoeksconclusies die gebruikers maken op basis van de cijfers. Of je nu de feitelijke of de verblijvende bevolking gebruikt om de bevolkingsomvang in Antwerpen te meten, en of je dit nu doet aan het begin of in het midden van het kalenderjaar, de meeste gebruikers zullen min of meer dezelfde conclusies trekken over bevolkingsgroei als ze deze cijfers over de jaren heen vergelijken.

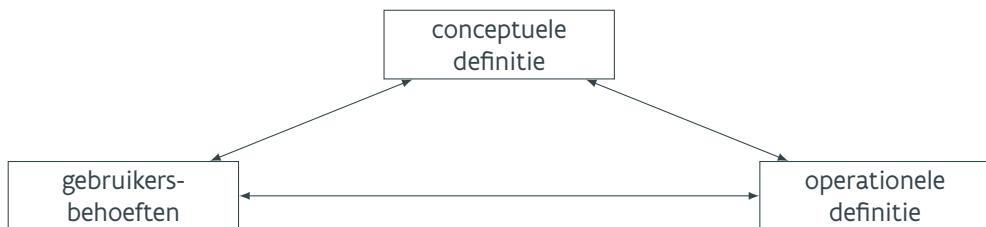
1.1.4 Kwaliteit

We zagen dat een statistiek wordt bescheven door drie elementen. Ze ontstaat vanuit gebruikersbehoeften en deze behoeften worden vertaald naar een conceptuele definitie en een operationele definitie (zie Figuur 1.1). In een ideale situatie zijn deze drie elementen goed op elkaar afgestemd en kunnen we zeggen dat we kwaliteitsvolle statistieken produceren. Wanneer deze drie elementen echter niet in lijn liggen met elkaar daalt de *kwaliteit* van de statistiek. Dit kan gebeuren in twee situaties:

- De conceptuele definitie is niet in lijn met de gebruikersbehoeften.

1. Wat is een statistiek?

Figuur 1.1: De kwaliteit van een statistiek wordt gedefinieerd door de verhouding tussen drie elementen.



- De operationele definitie is niet in lijn met de gebruikersbehoeften.
- De operationele definitie is niet in lijn met de conceptuele definitie.

We bespreken de drie situaties hieronder.

De conceptuele definitie is niet in lijn met gebruikersbehoeften

Als conceptuele definities niet in lijn zijn met de gebruikersbehoeften betekent dit eigenlijk dat we (1) geen statistieken produceren die bepaalde gebruikersbehoeften beantwoorden en/of dat we (2) statistieken produceren waar helemaal geen vraag naar is. In beide situaties hebben we een kwaliteitsprobleem binnen onze organisatie en vervullen we onze rol als onafhankelijke openbare statistiekdienst niet.

Het is daarom belangrijk dat we als statistiekorganisatie diepgaand onderzoek uitvoeren naar gebruikersbehoeften en deze grondig in kaart brengen. Wie heeft er nood aan statistieken? Wie gebruikt deze statistieken? Welk decreet dwingt ons ertoe deze statistieken te verzamelen? Voor welke beleidsbeslissingen of welk onderzoek zijn deze statistieken relevant? ...? Hoewel onderzoek naar deze vragen niet altijd een exacte wetenschap is, is het noodzakelijk om de kwaliteit van statistieken te kunnen beoordelen. Voor zo'n onderzoek moeten we bovendien verder kijken dan de kwantitatieve methoden waarmee statistiek vaak geassocieerd worden. Binnen gebruikeronderzoek kan ook veel kennis worden vergaard door diepgaande tekstanalyse via documentanalyse, hermeneutiek, juridische exegese of discoursanalyse, of door kwalitatief onderzoek via diepte-interviews, focusgroepen, participerende observaties en case studies.

We moeten hierbij natuurlijk opmerken dat het in de praktijk zo goed als onmogelijk is om alle gebruikersbehoeften in kaart te brengen, laat staan om aan al deze behoeften te voldoen. Dat is de reden waarom we als Statistiek Vlaanderen een officiële lijst aanleggen van statistieken die we produceren, namelijk de lijst van het Vlaams Statistisch Programma of VSP-lijst. Uiteraard is zo'n lijst nooit af en moet deze voortdurend geëvalueerd worden. Als Statistiek Vlaanderen hebben we dan ook de opdracht

om de VSP-lijst elke legislatuur te herzien en dit doen we natuurlijk best door te blijven investeren in gebruikersonderzoek.

De operationele definitie is niet in lijn met de gebruikersbehoeften

In sommige situaties kan er een wanverhouding ontstaan tussen gebruikersbehoeften en de operationele definitie van een statistiek. In ons Antwerpse voorbeeld kunnen gebruikers bijvoorbeeld het cijfer 525 935 systematisch interpreteren als de grootte van de verblijvende Antwerpse bevolking in plaats van de wettelijke bevolking omdat ze gewoonweg beleidsmatig voornamelijk de grootte van de verblijvende bevolking nodig hebben. Met andere woorden, de operationele definitie van onze cijfers is niet in overeenstemming met de noden van de gebruikers.

In situaties waarbij de operationele definitie niet beantwoordt aan de gebruikersbehoeften, moeten we onderzoeken of we onze operationele definitie kunnen aanpassen aan de behoeften. In bovenstaande situatie zou het bijvoorbeeld aangeraden zijn om als Statistiek Vlaanderen voortaan Eurostat te volgen en de grootte van de verblijvende bevolking te publiceren in plaats van de wettelijke bevolking.

In vele situaties is een evaluatie van de operationele definitie in functie van gebruikersbehoeften echter geen zwart-wit-verhaal. Er kunnen zich immers situaties voordoen waarin de voorgestelde operationele definitie voldoet voor de ene gebruiker maar niet voor een andere gebruiker. Zo kan onze statistiek over bevolkingsgrootte op basis van de wettelijke bevolking voldoende zijn voor heel wat gebruikers, behalve voor een specifieke beleidsmaker die het woonbeleid in Antwerpen moet bepalen. Deze beleidsmaker heeft misschien nood aan exacte cijfers over de verblijvende bevolking in plaats van de wettelijke bevolking. Ook in zo'n geval is er een gebruikersbehoefte waar we met de statistiek "Aantal inwoners in Antwerpen" niet aan voldoen, ook al voldoet deze statistiek wel voor de behoeften van de meeste andere gebruikers.

In situaties met verschillende niet-overlappende gebruikersbehoeften kunnen we besluiten niet langer één statistiek te publiceren maar verschillende statistieken. Zo kunnen we besluiten niet één statistiek te publiceren over "het aantal inwoners in Antwerpen", maar wel twee afzonderlijke statistieken: "de wettelijke bevolkingsgrootte van Antwerpen" en "de verblijvende bevolkingsgrootte van Antwerpen". We voldoen dan aan de specifieke noden van alle gebruikers. Merk op dat we in zo'n situaties zowel de conceptuele definities als de bijhorende operationele definities verder verfijnen en aanpassen. Uiteraard moeten dit soort keuzes steeds gebeuren vanuit praktische en pragmatische overwegingen in functie van de middelen en het personeel die we ter beschikking hebben. We kunnen

[1. Wat is een statistiek?](#)

in de bovenstaande situatie bijvoorbeeld ook besluiten om niet te voldoen aan de noden van deze éne specifieke beleidsmaker omdat dit te veel energie zou vragen voor een te kleine return.

We moeten ook opmerken dat kwaliteit in deze context ook in de brede zin van het woord kan worden bekeken. Een operationele definitie kan inhoudelijk aansluiten bij een gebruikersbehoefte, maar als de berekening te lang duurt en de statistiek te laat beschikbaar is, verlaagt dit alsnog de kwaliteit. Het is dus belangrijk een evenwicht te vinden tussen verschillende elementen van kwaliteit zoals inhoudelijke nauwkeurigheid, gebruiksvriendelijkheid en tijdige levering.

De operationele definitie is niet in lijn met de conceptuele definitie

De berekening van statistieken hangt steeds af van beschikbare informatie en data. Daarnaast kunnen statistieken berekend worden met heel verschillende methoden en technieken, de éne al complexer en grondiger dan de andere. De keuze voor bepaalde data en analysemethoden kunnen ervoor zorgen dat geproduceerde statistieken, zoals beschreven in de operationele definitie, sterk afwijken van de beoogde cijfers volgens de conceptuele definitie.

Zo produceren we bijvoorbeeld statistieken over de energiescore van bestaande woningen. Jammer genoeg bevatten administratieve registers enkel informatie over geregistreerde energiescores, meer bepaald op het moment dat een woning verkocht of verhuurd wordt. Er is dus, onder andere, geen informatie beschikbaar over energiescores van woningen die werden gerenoveerd na verkoop of voor woningen die al heel lange tijd niet werden doorverkocht. Dit kan er toe leiden dat de geproduceerde cijfers niet voldoende accuraat zijn om te voldoen aan de gebruikersbehoeften en de conceptuele definitie die daarmee gepaard gaat.

In dit soort situaties hebben we twee taken te vervullen. Ten eerste communiceren we best zo transparant mogelijk over de manier waarop de cijfers werden verzameld en wat de invloed kan zijn van de beperkingen van de data en analysemethoden op de accuraatheid van de cijfers. Indien mogelijk verfijnen we ook de conceptuele definitie van deze statistiek zodat het risico op verwarring verkleint. Zo kunnen we in bovenstaand voorbeeld de conceptuele definitie aanpassen van “energiescore van bestaande woningen” naar “geregistreerde energiescore van woningen”. Uiteraard kan de aanpassing van de conceptuele definitie er toe leiden dat de conceptuele definitie niet meer in lijn is met de gebruikersbehoeften, beleidsmakers weten waarschijnlijk liefst hoe het zit met de energiescore van alle woningen in Vlaanderen, niet enkel diegenen die toevallig geregistreerd werden.

Ten tweede kunnen we op zoek gaan naar bijkomende informatie, analysemethoden en modeleringstechnieken die de tekortkomingen van de beschikbare data helpen recht te zetten. Indien we geschikte methoden vinden om zo'n correcties uit te voeren, is het misschien mogelijk om een statistiek toch onder de oorspronkelijke conceptuele definitie te blijven produceren, in het voorbeeld hierboven de "energiescore van bestaande woningen". Uiteraard vertrekken analysemethoden vaak van onverifieerbare aannames die alsnog ervoor kunnen zorgen dat de geproduceerde cijfers niet volledig het beoogde concept accuraat weergeven. Het blijft dan belangrijk om volledig transparant te zijn over de gebruiken analysemethoden en hun beperkingen, en om de keuze voor de analysemethoden zo goed mogelijk te argumenteren op basis van empirisch wetenschappelijke studies.

De Europese Praktijkcode

Als we over de kwaliteit van statistieken spreken kunnen we niet voorbij de "Praktijkcode voor Europese statistieken" fietsen (**Eurostat2017European**). Deze praktijkcode geeft een overzicht van alle kwaliteitseisen voor statistieken.



Gebruikersbehoeften, conceptuele en operationele definitie komen weliswaar aan bod in informatiestandaarden zoals de SIMS en de GSBPM maar op een zeer onduidelijke en onoverzichtelijke manier. In de SIMS worden de gebruikersbehoeften pas opgelijst in categorie 12 terwijl dit de start vormt van een statistiekproductieproces. De operationele definitie zit in de SIMS dan weer op een zeer onsamenvhengende manier verspreid over verschillende categorieën terwijl de conceptuele definitie volledig ontbreekt.

De GSBPM doet het op dit vlak beter. Het beschrijft een proces dat vertrekt vanuit gebruikersbehoeften en eindigt met een kwaliteitsevaluatie. Het expliciteert echter onvoldoende dat de evaluatie gestoeld moet zijn op de gebruikersbehoeften en alle operationele keuzes die onderweg werden gemaakt.

1.1.5 Attributen van cijfers

Naast de operationale definitie van een cijfer, kan een cijfer gekenmerkt worden door andere informatie, die geen betekenis geeft aan het cijfer. Dit worden *attributen* genoemd.

Een belangrijke informatiepunt is de vertrouwelijkheid van het cijfer. Kan het cijfer worden gedeeld met het breed publiek, met collega's of is het vertrouwelijk en moet het veilig bewaard worden.

Een ander attribuut is de eigenaar van het cijfer. De eigenaar van een cijfer verwijst naar wie verantwoordelijk is voor dat cijfer. Deze kan

1. Wat is een statistiek?

Tabel 1.2: Een statistiek bestaat uit een cijfer, een conceptuele definitie die voortvloeit uit gebruikersbehoeften, en een operationele definitie die vertelt hoe het cijfer exact wordt gemeten. Een wanverhouding tussen gebruikersbehoeften en de operationele definitie bepaalt de kwaliteit van de statistiek.

Gebruikersbehoeften:	→ Decreet XX.XX verwijst naar het aantal inwoners in Vlaamse gemeenten in de context van ... → Agentschap Binnenlands Bestuur publiceert cijfers over het aantal inwoners in gemeenten om informatie te geven over ... → Academische onderzoekers vragen cijfers over het aantal inwoners in gemeenten om onderzoek te voeren over ... → ...
Conceptuele definitie:	Aantal inwoners van de gemeente Antwerpen in 2019
Operationele definitie:	De grootte van de wettelijke bevolking op 1 januari 2019 0.00 uur in de gemeente Antwerpen (Statbel NIS-code 11002 in 2019). De data worden aangeleverd door Statbel op basis van het Rijkregister van de natuurlijke personen. De wettelijke bevolking verwijst naar ...
Cijfer:	525 935
Kwaliteitsevaluatie:	De gebruikers hebben ook nood aan de grootte van de verblijvende bevolking naast de wettelijke bevolking

1.1.6 Samengevat

Samengevat, voor elke statistiek die we produceren en publiceren hebben we nood aan:

1. de gebruikersbehoeften,
2. een conceptuele definitie op basis van de gebruikersbehoeften,
3. een operationele definitie,
4. het cijfer, en
5. een kwaliteitsevaluatie van de operationele definitie in functie van de gebruikersbehoeften.

Tabel 1.2 vat deze basisinformatie samen aan de hand van het Antwerps voorbeeld. Als de gebruikersbehoeften van een statistiek nog niet in kaart zijn gebracht en er daardoor ook geen kwaliteitsevaluatie kan worden uitgevoerd, is de statistiek een experimentele statistiek.

1.2 STATISTIEK ALS EEN CIJFERTABEL

Tabel 1.3: We publiceren doorgaans geen individuele statistieken maar statistiekreeksen.

Conceptuele definitie	Cijfer
Aantal inwoners in Aalst in 2019	86 445
Aantal inwoners in Aalter in 2019	28 906
Aantal inwoners in Aarschot in 2019	30 115
Aantal inwoners in Aartselaar in 2019	14 293
Aantal inwoners in Affligem in 2019	13 228
Aantal inwoners in Alken in 2019	11 499
Aantal inwoners in Alveringem in 2019	5 047
Aantal inwoners in Antwerpen in 2019	525 935
Aantal inwoners in Anzegem in 2019	14 716
...	...
Aantal inwoners in Zwijndrecht in 2019	19 056

1.2.1 Wat is een statistiekreeks?

In de vorige paragraaf werd uitgelegd hoe een statistiek verwijst naar één enkel cijfer met een bijbehorende conceptuele en operationele definitie ontstaan vanuit specifieke gebruikersbehoeften. In de praktijk produceren we echter veel statistieken met sterk overlappende gebruikersbehoeften en definities. Zo publiceren we niet alleen het aantal inwoners voor de gemeente Antwerpen, maar ook voor alle andere Vlaamse gemeenten, de drie gewesten en heel België. Aparte definities opstellen en beheren voor al deze cijfers is niet efficiënt.

Om efficiëntie te verhogen werken we daarom met *statistiekreeksen* of cijferreeksen: verzamelingen van statistieken die grotendeels dezelfde conceptuele en operationele definities delen. Tabel 1.3 toont bijvoorbeeld een statistiekreeks met de inwoneraantallen van alle Vlaamse gemeenten.

1.2.2 Concepten, dimensies en attributen

Aangezien de conceptuele en operationele definities van individuele cijfers in statistiekreeksen overlappen, kunnen we deze definities opsplitsen in verschillende onderdelen. Deze verschillende onderdelen van de definities worden de *concepten* genoemd. Concepten zijn eigenlijk gewoon variabelen, kenmerken waar de statistieken naar refereren en die kunnen variëren tussen statistieken. Bij de statistiekreeks in tabel 1.3 kunnen we bijvoorbeeld drie concepten onderscheiden: (1) de parameter “aantal inwoners”, (2) het jaartal “2019”, en (3) de gemeente. Niet alle statistieken die we publiceren geven immers het aantal inwoners als parameter weer, en evenmin geven ze allemaal informatie voor het jaar 2019. De stati-

1. Wat is een statistiek?

stiekreeks in tabel 1.3 toont bovendien al aan hoe gemeente varieert over verschillende statistieken.

Binnen één statistiekreeks wordt een onderscheid gemaakt tussen twee soorten concepten. Sommige concepten helpen elk cijfer uniek te identificeren, terwijl anderen enkel bijkomende informatie geven over de cijfers. De concepten die de cijfers helpen identificeren binnen een statistiekreeks bepalen de *dimensies* van de statistiekreeks. Ze worden ook wel *identificeerders* (identifiers) genoemd. In tabel 1.3 is er slechts één dimensie, namelijk de gemeente, aangezien elk cijfer het aantal inwoners van een andere gemeente weergeeft en de cijfers enkel van elkaar verschillen door de gemeente waarnaar ze verwijzen.

De concepten die enkel bijkomende informatie geven zonder de cijfers verder te identificeren worden daarentegen *attributen* genoemd. Attributen worden daarom ook soms *beschrijvers* (describers) genoemd. In tabel 1.3 zijn de parameter en het jaartal de attributen, omdat elk cijfer verwijst naar een inwonersaantal in 2019 en niets anders. Attributen kunnen informatie bevatten voor de hele statistiekreeks, maar kunnen evengoed enkel informatie bevatten voor één enkel cijfer of een beperkte groep van cijfers binnen de statistiekreeks. Een voorbeeld van zo'n laatste situatie is informatie over de vertrouwelijkheid van cijfers. Één groep cijfers in een tabel kan beschouwd worden als vrij te publiceren terwijl een andere groep niet openbaar mag gemaakt worden omdat ze vertrouwelijke informatie kunnen ontsluiten. De aanduiding van deze vertrouwelijkheid identificeert de cijfers op zich niet maar varieert wel van cijfer tot cijfer.

Onthoud echter dat attributen dimensies kunnen worden wanneer verschillende statistiekreeksen worden samengevoegd. Om bijvoorbeeld bevolkingsdichtheden te berekenen, moeten de cijfers uit tabel 1.3 worden gecombineerd met cijfers over de oppervlakte van alle gemeenten. In die samengestelde tabel wordt de parameter een dimensie in plaats van een attribuut, omdat sommige cijfers verwijzen naar de parameter ‘aantal inwoners’, terwijl anderen verwijzen naar de parameter “oppervlakte” van de gemeenten. Hetzelfde gebeurt met het concept jaartal wanneer we aantallen inwoners combineren over verschillende jaren heen voor een tijdreeks. Omgekeerd kan een dimensie ook een attribuut worden wanneer we een tabel opsplitsen volgens die dimensie. Bijvoorbeeld, als we longitudinale data op jaarbasis opsplitsen in aparte tabellen per jaartal, wordt jaartal in die nieuwe tabellen slechts een attribuut in plaats van een dimensie. Het verschil tussen attributen en dimensies is dus relatief want het hangt af van welke tabel je precies bekijkt. Over al mogelijke statistieken bekeken zijn alle concepten dimensies want geen enkele operationele definitie is dezelfde.



We hebben hier een hiaat in het SDMX model blootgelegd. De documentatie van SDMX bespreekt nergens hoe attributen dimensies kunnen worden of vice versa wanneer datasets met statistiekreeksen worden samengevoegd of uitgesplitst. Om deze reden is de uitwerking van attributen in het model ook veel beperkter dan de uitwerking van dimensies. Hierdoor kan SDMX moeilijker gelinkt worden aan de ideeën rond conceptuele definities, operationele definities en kwaliteit van statistiekreeksen.

Sterker nog, binnen SDMX wordt parameter nooit duidelijk beschouwd als een dimensie waardoor je verschillende parameters zoals bevolkingsgrootte en oppervlakte moeilijk in één tabel kan combineren zonder in problemen te komen over de definitie van dimensies en attributen.

1.2.3 Conceptuele en operationele definitie

Net zoals elke afzonderlijke statistiek worden statistiekreeksen ook beschreven via een conceptuele en een operationele definitie. Dit vergemakkelijkt informatiebeheer over de statistiekreeksen.

De *conceptuele definitie* verwijst naar de interpretatie die een doorsnee gebruiker aan de statistiekreeks toekent. De statistiekreeks in tabel 1.3 kan bijvoorbeeld conceptueel worden gedefinieerd als het “aantal inwoners in gemeenten in het Vlaamse Gewest in 2019”. Deze definitie sluit aan bij specifieke gebruikersbehoeften, zoals bijvoorbeeld beschreven in tabel 1.2.

De *operationele definitie* van een statistiekreeks beschrijft hoe de cijferreeks precies werd verzameld en gereproduceerd kan worden. Bovendien specificeert de operationele definitie steeds welke waarden de attributen en dimensies binnen de reeks precies aannemen. Voor tabel 1.3 luidt de operationele definitie bijvoorbeeld: “De grootte van de wettelijke bevolking op 1 januari 2019 per gemeente volgens de NIS code indeling in 2019. De wettelijke bevolking omvat...” Deze definitie benoemt heel duidelijk ‘gemeente’ als dimensie en geeft aan hoeveel gemeenten de reeks omvat, namelijk de 300 Vlaamse gemeenten volgens de NIS code indeling van 2019, en bijvoorbeeld niet de 285 Vlaamse gemeenten vanaf 2025. Daarnaast beschrijft de definitie ook heel duidelijk de waarde van de attributen. De reeks bevat namelijk cijfers over de grootte van de wettelijke bevolking en bijvoorbeeld niet de verblijvende bevolking, en voor het jaar 2019 en geen ander kalenderjaar.

Door de exacte beschrijving van dimensies bepaalt de operationele definitie bovendien strikt uit hoeveel cijfers een statistiekreeks bestaat, zelfs als sommige cijfers een missende of versluierde waarde hebben. De reeks in tabel 1.3 bestaat bijvoorbeeld uit 300 cijfers, voor elke Vlaamse gemeente één. Als er in de reeks ook totaalcijfers zouden worden gepubliceerd voor Vlaanderen, Wallonië, Brussel en België, verandert de dimensie ‘gemeente’ naar het ruimer concept ‘geografisch gebied’ met 304 mogelijke waarden

1. Wat is een statistiek?

en dikt de reeks tot evenveel cijfers aan. Splitsen we de cijfers verder uit naar mannen en vrouwen, dan wordt de dimensie ‘geslacht’ toegevoegd met twee mogelijke waarden (man en vrouw) en bevat de reeks plots 608 cijfers. Inclusief de totalen over beide geslachten komt het totaal op 912 cijfers. Voeg daar nog het percentage mannen en vrouwen per gemeente aan toe, dan bevat de reeks 1520 cijfers.

Een kleine tip om de operationele definitie duidelijk uit te schrijven: Maak slim gebruik van voorzetels zoals “voor”, “in” of “per” om de attributen en dimensies aan te duiden. Zo kan je bijvoorbeeld een statistiekreeks definiëren als het “aantal inwoners IN het jaar 2019 PER geografisch gebied (Vlaamse gemeenten, Belgische gewesten en heel België volgens de NIS code indeling van 2019), VOOR elk geslacht (mannen, vrouwen, totaal), en PER leeftijdsgroep (0 18, 19 30, 31 65, 66+ jaar, totaal).”

Hoewel het bij de conceptuele definitie minder strikt is om alle attributen en dimensies explicet te vermelden, is dit ook een goede praktijk. Voor de statistiekreeks in tabel 1.3 kunnen we bijvoorbeeld kiezen tussen de conceptuele definities “bevolkingsomvang” of “aantal inwoners in de Vlaamse gemeenten in 2019”. De tweede definitie is uiteraard een pak duidelijker als omschrijving. De eerste definitie is anderzijds een pak beknopter en gemakkelijker te citeren maar wordt uiteraard dubbelzinnig als we ook tabellen publiceren voor andere kalenderjaren (tenzij we al deze tabellen samenvoegen natuurlijk, maar dan zitten we weer in een ander verhaal).

Net zoals bij individuele statistieken bepaalt een wanverhouding tussen gebruikersbehoeften en de operationele definitie mee de kwaliteit van een statistiekreeks. Als de operationele definitie niet in lijn ligt met de behoeften van gebruikers moet de reeks worden aangepast of worden geschrapt.

De presentatie van een statistiekreeks gebeurt doorgaans niet zoals in tabel 1.3. Attributen worden vaak uit de rijen verwijderd en opgenomen in de titel of conceptuele definitie van de statistiekreeks. Tabel 1.4 toont een meer gangbare weergave van een uitgebreidere cijferreeks. In deze tabel zie je dat de concepten “aantal inwoners” en jaartal “2019” enkel worden vermeld in de titel van de tabel. De tabel telt verder drie dimensies. Dimensies gemeente en geslacht staan in de rijen. De derde dimensie is de statistische parameter die een verschil maakt tussen aantalen en percentages. Deze dimensie staat in de kolommen.



Binnen de huidige SDMX-standaard stellen we statistiekreeksen voor met slechts één kolom voor de cijfers. De structuur in tabel 1.3 zou hiervoor gepivoteerd moeten worden. Hoogstwaarschijnlijk versoepelt de SDMX-standaard echter in de toekomst waardoor verschillende kolommen cijfers kunnen bevatten.

Tabel 1.4: De conceptuele definitie van een statistiekreeks verwijst naar de interpretatie van een doorsnee gebruiker terwijl de operationele definitie alle concepten (attributen en dimensies) van de reeks duidelijk beschrijft.

AANTAL MANNEN EN VROUWEN IN VLAAMSE GEMEENTEN IN 2019.
Het aantal en het percentage mannen en vrouwen in de wettelijke bevolking op 1 januari 2019 per gemeente volgens de NIS-code-indeling in 2019.

Gemeente	Geslacht	Aantal	Percentage
Aartselaar	man	7089	49.6
Aartselaar	vrouw	7204	50.4
Antwerpen	man	262921	50.0
Antwerpen	vrouw	263014	50.0
Boechout	man	6506	49.0
Boechout	vrouw	6760	51.0
Boom	man	9024	49.5
Boom	vrouw	9220	50.5
Borsbeek	man	5270	48.6
Borsbeek	vrouw	5584	51.4

1.2.4 Samengevat

Samengevat, een statistiekreeks is een reeks van statistieken waarbij gebruikersbehoeften, conceptuele en operationele definities overlappen. Elke statistiekreeks is nauwkeurig gedefinieerd via haar operationele definitie. Deze operationele definitie vertelt exact hoeveel cijfers de statistiekreeks bevat, wat deze cijfers betekenen en hoe de cijfers werden bekomen. Een wanverhouding tussen de operationele definitie en de gebruikersbehoeften is een van de kwaliteitsindicatoren van de statistiekreeks volgens de praktijkcode.

Met Statistiek Vlaanderen leggen we op voorhand vast welke statistiekreeksen we publiceren. Deze statistiekreeksen zijn terug te vinden in het Vlaams Statistisch Programma (VSP lijst). Voor elke statistiek in de VSP lijst voorzien we ten minste volgende informatie:

1. de conceptuele definitie (die gebruikt wordt als handig label voor de statistiekreeks),
2. het versienummer
3. de operationele definitie inclusief een beschrijving van alle concepten (attributen en dimensies) inclusief de exacte waarden van die concepten,
4. een overzicht van de gebruikersbehoeften, en
5. een kwaliteitsevaluatie van de operationele definitie in functie van de gebruikersbehoeften.

1. Wat is een statistiek?

Tabel 1.5 toont hoe de VSP-lijst er minimaal zou moeten uitzien. In de praktijk zullen we uiteraard informatie over statistiekreeksen niet bewaren in één gigantische tabel maar werken met informatiefiches per statistiekreeks gebaseerd op de single integrated metadata structure (SIMS)¹.

1.3 STATISTIEK ALS EEN TABELREEKS

Voorbeeld: bevolking naar leeftijd en geslacht – bevolkingspyramide Hieruit kunnen heel wat andere cijferreeksen worden berekend - aantal inwoners naar geslacht - percentage volgens geslacht - aantal inwoners naar leeftijd - percentage per leeftijd - percentage 65+'ers - vergrijzing - ...

Opnieuw conceptuele en operationele definitie nodig



Er bestaan verschillende soorten concepten. Op de eerste plaats kunnen concepten concreet of abstract zijn.

Op de tweede plaats kunnen concepten enkelvoudig of meervoudig zijn.

- *Een enkelvoudig concept verwijst naar een idee of kenmerk dat op één dimensie of aspect betrekking heeft. Het is relatief eenvoudig te definiëren en te meten, en het omvat één afgebakend betekenisveld. Leeftijd is bijvoorbeeld een enkelvoudig concept: het verwijst enkel naar het aantal jaren sinds iemands geboorte. Het is eenduidig te meten en wordt in de praktijk meestal als een numerieke variabele weergegeven (in jaren, maanden, ...).*
- *Een meervoudig concept verwijst naar een idee dat uit meerdere dimensies of onderliggende componenten bestaat. Je hebt meerdere indicatoren nodig om het concept volledig te vatten. Welzijn is bijvoorbeeld een meervoudig concept, want het kan verschillende onderdelen bevatten zoals fysieke gezondheid, mentale gezondheid, sociale relaties of economische situatie omvatten. Om het te meten zijn verschillende indicatoren of meetinstrumenten nodig – bijvoorbeeld via samengestelde indexen of uitgebreide vragenlijsten.*

De verschillende soorten concepten kunnen ook samen voorkomen zoals blijkt uit volgende voorbeelden:

- *Leeftijd in jaren is een concreet en enkelvoudig concept. Dit is een direct observeerbaar, meetbaar kenmerk (concreet) en het gaat slechts over één dimensie (enkelvoudig). De meting is objectief en eenduidig.*
- *Woningkenmerken (bv. oppervlakte, aantal kamers, type dak) is een concreet en meervoudig concept. De verschillende woningkenmerken zijn stuk voor stuk observeerbaar (concreet), maar samen beschrijven ze een meervoudig concept zoals de "kwaliteit van een*

¹ec.europa.eu/eurostat/web/metadata/reference-metadata-reporting-standards

1.3. Statistiek als een tabelreeks

Tabel 1.5: Het Vlaams Statistisch Programma (VSP) definieert een lijst van statistiekreeksen die Statistiek Vlaanderen publiceert. Elke statistiekreeks in deze reeks beschikt over een duidelijke conceptuele en operationele definitie, opgeliijste gebruikersbehoeften en een kwaliteitsevaluatie.

Aantal mannen en vrouwen, versie 1
<i>gebruikersvoorraarden:</i> → Decreet XX.XX verwijst naar het aantal inwoners in Vlaamse gemeenten om beleid te voeren over ... → Agentschap Binnenlands Bestuur publiceert cijfers over het aantal inwoners in gemeenten als ... → Academische onderzoekers vragen cijfers over het aantal inwoners in gemeenten om onderzoek te voeren over ... → ...
<i>Operationele definitie:</i> Het aantal en het percentage mannen en vrouwen in de feitelijke bevolking (aantal geregistreerde inwoners in het Rijksregister inclusief personen in het wachtrecht en ambassadeurs) op 1 januari van elk kalenderjaar vanaf 2005 per Vlaamse gemeente volgens de NIS-code-indeling in 2019 = 600 cijfers per jaar.
<i>Kwaliteit:</i> De gebruikers hebben eerder nood aan de grootte van de verblijvende bevolking in plaats van de feitelijke bevolking.
<i>Nota:</i> Stopgezet in 2022, wegens aanpassing definitie.
Aantal mannen en vrouwen, versie 2
<i>gebruikersvoorraarden:</i> → Decreet XX.XX verwijst naar het aantal inwoners in Vlaamse gemeenten om beleid te voeren over ... → Agentschap Binnenlands Bestuur publiceert cijfers over het aantal inwoners in gemeenten als ... → Academische onderzoekers vragen cijfers over het aantal inwoners in gemeenten om onderzoek te voeren over ... → ...
<i>Operationele definitie:</i> Het aantal en het percentage mannen en vrouwen in de verblijvende bevolking (aantal geregistreerde inwoners in het Rijksregister inclusief personen in het wachtrecht en ambassadeurs en personen die minder dan drie maanden in België verblijven) op 1 januari van elk kalenderjaar vanaf 2005 per Vlaamse gemeente volgens de NIS-code-indeling in 2019 = 600 cijfers per jaar.
<i>Kwaliteit:</i> Onderzoek XX toont aan dat er geen problemen zijn met deze statistiekreeks.
<i>Nota:</i> Stopgezet in 2025, wegens aanpassing gemeenten door fusies.
Aantal mannen en vrouwen, versie 3
<i>gebruikersvoorraarden:</i> → Decreet XX.XX verwijst naar het aantal inwoners in Vlaamse gemeenten om beleid te voeren over ... → Agentschap Binnenlands Bestuur publiceert cijfers over het aantal inwoners in gemeenten als ... → Academische onderzoekers vragen cijfers over het aantal inwoners in gemeenten om onderzoek te voeren over ... → ...
<i>Operationele definitie:</i> Het aantal en het percentage mannen en vrouwen in de verblijvende bevolking (aantal geregistreerde inwoners in het Rijksregister inclusief personen in het wachtrecht en ambassadeurs en personen die minder dan drie maanden in België verblijven) op 1 januari van elk kalenderjaar vanaf 2005 per Vlaamse gemeente volgens de NIS-code-indeling in 2025 = 570 cijfers per jaar.
<i>Kwaliteit:</i> Onderzoek XX toont aan dat er geen problemen zijn met deze statistiekreeks.
Tewerkstelling in hoogtechnologische sector, versie 1
<i>gebruikersvoorraarden:</i> → Decreet XXXX verwijst naar Tewerkstelling in hoogtechnologische sector in kader van ... → De cijferpagina met cijfers over tewerkstelling in hoogtechnologische sector wordt X aantal keer per jaar geraadpleegd. → ...
<i>Operationele definitie:</i> Percentage van de hele werkende bevolking aan de slag in de hoogtechnologische sector op 1 januari van elk kalenderjaar vanaf 2005 per Vlaamse gemeente en voor het hele Vlaamse gewest volgens de NIS-code-indeling in 2019. De werkende bevolking omvat ... De hoogtechnologische sector bestaat uit- bedrijven ... De data worden verzameld via de Enquête naar de Arbeidskrachten (EAK) door Statbel. In deze enquête wordt data verzameld door ... = 301 cijfers per jaar.
<i>Kwaliteit:</i> Onderzoek YY toont aan dat er geen problemen zijn met deze statistiekreeks.
Drinkwaterkwaliteit, versie 1
<i>gebruikersvoorraarden:</i> → Decreet XX.XX verwijst naar drinkwaterkwaliteit in kader van ... → ...
<i>Operationele definitie:</i> Conformiteitspercentage van het kraantjeswater in heel Vlaanderen per jaar. Het conformiteitspercentage wordt berekend door Vlaamse Milieuumaatschappij (VMM) op basis van het totale aantal analyses en het totale aantal vastgestelde normoverschrijdingen voor volgende parameters: ... = 1 cijfer per jaar.
<i>Kwaliteit:</i> Volgens rapport ZZ ontstaan er kleine onzekerheidsfouten in de meting van parameter x, waardoor het werkelijke conformiteitspercentage kan afwijken met 0,2%. Deze fluctuatie heeft slechts beperkte invloed op de kwaliteit waardoor geen herdefiniëring nodig is.
....., versie ...
<i>gebruikersvoorraarden:</i> ... <i>Operationele definitie:</i> ... <i>Kwaliteit:</i> ...
.....

1. Wat is een statistiek?

woning". Je hebt meerdere concrete indicatoren nodig om het volledig te vatten.

- *Arbeidstevredenheid is een abstract en enkelvoudig concept. Dit is een subjectief, niet rechtstreeks observeerbaar gevoel (abstract), maar het wordt soms in onderzoek benaderd als één dimensie, bijvoorbeeld via een schaalvraag: "Hoe tevreden bent u over uw werk?". Dan behandel je het als een enkelvoudig concept, al kan je het ook complexer opvatten.*
- *Sociaal isolement is een abstract en meervoudig concept. Dit is een abstract concept dat meerdere onderliggende aspecten omvat, zoals subjectieve eenzaamheid, frequentie van sociaal contact of deelname aan sociale activiteiten. Het vereist een set van vragen of indicatoren om het volledig te meten.*

1.4 EXPERIMENTELE STATISTIEKEN

Sommige statistieken ontstaan niet vanuit geobserveerde gebruikersbehoeften, maar vanuit veronderstellingen hierover. Dit kan voortkomen uit de inhoudelijke interesse van een collega of door nieuwe dataverzamelingen of analysemethoden. Deze statistieken noemen we *experimentele statistieken*.

Experimentele statistieken verwijzen dus niet enkel naar statistieken ontwikkeld met de allernieuwste analysetechnieken die recent op de markt werden gebracht. Eurostat definieert experimentele statistieken als "Experimental statistics use new data sources and methods to better respond to our users' needs in a timely manner." Wat "new data sources and methods" betekent in deze definitie is echter rekbaar. Zo kan een nieuwe vraag in een bevraging al beschouwd worden als een nieuwe databron en kunnen we deze vraag gebruiken om een nieuwe experimentele statistiek te ontwikkelen waarover we enkel aannames kunnen maken over de gebruikersbehoeften.

Bij gebrek aan concrete gebruikersbehoeften kan de kwaliteit van experimentele statistieken uiteraard niet meteen worden geëvalueerd. Kwaliteit verwijst immers, zoals eerder al aangehaald, naar de overeenkomst tussen de gebruikersbehoeften en de operationele definitie van een statistiek.

Daarom is het noodzakelijk om voor elke experimentele statistiek een deadline voor evaluatieonderzoek rond gebruikersbehoeften af te spreken (bijvoorbeeld vijf jaar na de eerste publicatie). Dit evaluatieonderzoek beantwoordt vragen zoals:

- Wordt de statistiek gebruikt, en zo ja, waarvoor?

- Sluit het gebruik en de interpretatie aan bij de operationele definitie?
- Zijn er suggesties van gebruikers om de statistiek beter af te stemmen op hun behoeften?

Eens gebruikersbehoeften in kaart zijn gebracht, kan de kwaliteit van de statistiek verder worden geëvalueerd. Afhankelijk van de resultaten kan de statistiek nadien:

- worden geüpgraded tot een volwaardige officiële statistiek;
- worden aangepast op basis van de bevindingen; of
- worden stopgezet indien de kwaliteit onvoldoende is.



Het groot deel van de statistieken van de VSA zijn vermoedelijk experimenteel, aangezien gebruikersbehoeften niet systematisch zijn vastgelegd. Hierdoor is een kwaliteitsbeoordeling van deze statistieken momenteel moeilijk. Onderzoek naar gebruikersbehoeften dringt zich op bij deze statistieken.