



Vlaanderen
is sterk in cijfers

HANDBOEK STATISTIEKPRODUCTIE

4 september 2025

VLAAMSE
STATISTISCHE
AUTORITEIT

Inleiding	3
1 Wat is een statistiek?	4
1.1 Statistiek als een cijfer	4
1.1.1 Gebruikersbehoeften	4
1.1.2 Conceptuele definitie	7
1.1.3 Operationele definitie	8
1.1.4 Conceptuele versus operationele definitie	11
1.1.5 Kwaliteit	12
1.1.6 Attributen	16
1.1.7 Samengevat	17
1.2 Statistiek als een cijfertabel	18
1.2.1 Wat is een cijfertabel?	18
1.2.2 Concepten, dimensies en attributen	19
1.2.3 Conceptuele en operationele definitie	21
1.2.4 Samengevat	22
1.3 Statistiek als een tabelreeks	25
1.4 Experimentele statistieken	26
Bibliografie	28

INLEIDING

Één van de hoofdtaken van het netwerk Statistiek Vlaanderen en de Vlaamse Statistische Autoriteit is de productie en publicatie van openbare statistieken, ook wel Vlaamse Openbare Statistieken (VOS'en) genoemd. Maar wat verstaan we eigenlijk onder een statistiek en hoe produceren we die dan concreet? Om het netwerk goed te laten functioneren, is het cruciaal om met een gedeelde woordenschat te werken. Momenteel bestaan er echter grote verschillen in hoe termen en concepten binnen het netwerk worden gebruikt. Deze nota heeft daarom als doel enkele basisconcepten in de Vlaamse statistiekproductie te verduidelijken.

In deze handleiding wordt terminologie uit de GSIM-, GSBPM en de SDMX standaarden gebruikt, al verwijzen we voorlopig slechts sporadisch in enkele kanttekeningen naar deze standaarden. De integratie van beide standaarden kan in een volgende versie opgenomen worden.

Hoofdstuk 1

WAT IS EEN STATISTIEK?

Voor we 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? Welke informatie is essentieel om statistieken te produceren en te publiceren? Hoe bepalen we de kwaliteit van statistieken? 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. Statistiek gaat 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 aan de basis liggen van de kwaliteit van de statistiek.

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 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 (Wild, Utts en Horton 2018). Beide zijn even belangrijk en vormen één geheel. In dit geval betekent het cijfer 525 935: “het aantal inwoners van de gemeente Antwerpen in 2019”.



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 eenvoud verwijzen we in deze tekst enkel naar statistieken als enkelvoudige cijfers. Cijfers zijn immers veruit de meest voorkomende vorm van statistieken.

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. Een Vlaams decreet kan bijvoorbeeld verwijzen naar het aantal inwoners in de Vlaamse gemeenten en verplicht ons, Statistiek Vlaanderen, dit aantal te publiceren zodat de Vlaamse regering hierop beleid kan uitstippelen. Antwerpse beleidsmakers weten daarnaast waarschijnlijk ook graag hoeveel inwoners hun gemeente telt zodat ze hun diensten hierop kunnen afstemmen. Wetenschappelijke onderzoekers gebruiken de Antwerpse bevolkingsgrootte in 2019, samen met gegevens uit andere jaren en gemeenten, dan weer om bevolkingsgroei en spreiding in Vlaanderen te onderzoeken. Burgers willen misschien wel graag weten hoeveel inwoners Antwerpen telt om te beoordelen of bepaalde Vlaamse subsidies aan de stad gerechtvaardigd zijn of niet.

Gebruikers hebben dus nood aan inhoudelijk relevante en nauwkeurige statistieken. *Relevant* betekent dat we een statistiek produceren als daar behoefte aan is maar ook omgekeerd dat we geen statistieken produceren waar helemaal geen behoefte aan is. *Nauwkeurig* betekent dat de statistieken ook meten wat ze moeten meten. Beide criteria zijn sterk aan elkaar gerelateerd, je kan geen relevante statistieken opleveren als ze niet nauwkeurig zijn, maar ze bevatten toch een nuanceverschil die verder in de tekst zal worden uitgelegd. Beide criteria vormen ook een belangrijke beginselen in de “Praktijkcode voor Europese statistieken” (zie Eurostat

2017). Deze praktijkcode vormt de hoeksteen van het kwaliteitskader van het Europees statistisch systeem (ESS) en legt 16 beginselen vast voor de ontwikkeling, productie en verspreiding van Europese statistieken.

Naast relevantie en nauwkeurigheid kunnen gebruikers echter ook nog andere verwachtingen hebben. Deze worden eveneens gedefinieerd in de “Praktijkcode voor Europese statistieken”. Gebruikers kunnen bijvoorbeeld verwachten dat statistieken *tijdigheid en punctualiteit* zijn. Dit betekent dat de statistieken snel en op duidelijke tijdstippen worden gepubliceerd. Dit laat beleidsmakers bijvoorbeeld toe snelle beleidsbeslissingen te nemen in tijden van crisis of wanneer bepaalde beleidsbeslissingen gebonden zijn aan duidelijke juridische deadlines. Gebruikers kunnen ook verwachten dat de statistieken *vergelijkbaar* zijn. Dit betekent dat de gebruiker de statistieken kan vergelijken met eerder verschenen cijfers of met cijfers uit andere landen of regio's. Gebruikers verwachten meestal ook dat de statistieken *toegankelijk* zijn. Dit betekent dat de gebruikers de statistieken op een gebruiksvriendelijke manier kunnen raadplegen en dat de statistieken gemakkelijk te interpreteren zijn. Als laatste moeten statistieken ook steeds *wettelijk verantwoord* zijn. Dit betekent dat de statistieken worden gepubliceerd conform de wet en regelgeving.



Naast criteria voor statistische output, die gebruikersbehoeften weerspiegelen en hierboven werden beschreven, bevat de “Praktijkcode voor Europese statistieken” nog tien andere beginselen rond het institutioneel kader en statistische processen. Deze “beginselen” kan je echter bekijken als hulpmiddelen om de beginselen rond statistische output waar te maken.

Zo vereist beginsel 1 dat statistische autoriteiten professioneel onafhankelijk opereren. Dit betekent dat er geen politieke of andere externe invloeden zijn op het werk van de autoriteiten. Beginsel 6 is hier nauw mee verbonden en vraagt dat statistische autoriteiten onpartijdig en objectief zijn. Deze beginselen zorgen er dus voor dat de statistische autoriteiten alle gebruikersbehoeften maatschappijbreed in kaart brengen, hun statistiekproductie hierop afstemmen en daardoor relevante en nauwkeurige statistieken opleveren.

Beginsel 2 bepaalt dat de statistische autoriteit een wettelijk mandaat heeft voor het verzamelen en verkrijgen van toegang tot gegevens. Beginsel 3 stelt dat statistische autoriteiten over voldoende personele, financiële en technische middelen moeten beschikken. Deze beginselen garanderen dat gebruikersbehoeften niet alleen in kaart worden gebracht, maar dat de statistische autoriteit ook deze behoeften kan voldoen met relevante en nauwkeurige statistieken.

Beginsel 4 vraagt om een sterk kwaliteitsbewustzijn bij statistische autoriteiten. Zo'n kwaliteitsbewustzijn betekent niet meer en niet minder dan dat statistische autoriteiten moeten streven naar maximale nauwkeurigheid. Beginsel 4 focust op het institutioneel kader dat moet worden gecreëerd binnen een statistische autoriteit, maar hangt eigenlijk nauw samen met heel wat andere beginselen over de organisatie van statistische processen. Beginsel 7 vereist dat statistische autoriteiten deugdelijke methoden gebruiken om kwaliteit voortdurend te optimaliseren. Beginsel 8 vraagt dat geschikte statistische procedures worden gebruikt door de statistische autoriteit.



Dat statistische output wettelijk verantwoord moet zijn staat niet expliciet vermeld in de Europese praktijkcode als vereiste. Deze vereiste wordt wel impliciet vermeld via beginselen over een institutioneel kader rond statistische geheimhouding en gegevensbescherming. Dit is een duidelijk hiaat in de praktijkcode.

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* van de statistiek (**refs toevoegen**; Neuman 2014). In ons voorbeeld is “het aantal inwoners van de gemeente Antwerpen in 2019” dus de conceptuele definitie van de statistiek “525 935”.

Een conceptuele definitie bevat doorgaans verschillende 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 het jaar 2019. Ten derde kijken we naar een grootheid die we willen meten, namelijk de bevolkingsgrootte. Deze componenten noemen we *concepten* (Clark e.a. 2021). 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 deel 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 juist of fout antwoord. Om het

productieproces voor deze statistieken in de praktijk te organiseren zullen we echter wel afspraken 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*. Die 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. De vertaling van de conceptuele definitie naar de operationele definitie noemen we *operationaliseren*.

Operationalisering van een conceptuele definitie betekent in de praktijk meestal dat elk concept in de definitie geoperationaliseerd wordt. Concepten variëren echter in hoe concreet of abstract ze zijn (**referentie toevoegen**). Een *concreet concept* verwijst naar een fenomeen dat direct observeerbaar of eenvoudig meetbaar is 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 moet daarom eerst geoperationaliseerd worden naar concretere concepten. Een voorbeeld van een abstract concept is sociaal isolement. Sociaal isolement verwijst immers naar een toestand waarin iemand weinig of geen sociale contacten heeft of zich sociaal buitengesloten voelt. Dit kan je niet rechtstreeks meten; je moet het operationaliseren, bijvoorbeeld via een langere vragenlijst over, onder andere, het aantal sociale contacten, subjectieve gevoelens van eenzaamheid en deelname aan sociale activiteiten.

In de praktijk zijn concepten niet of concreet of abstract, maar kan je ze op een continue schaal plaatsen op het vlak van abstractheid en concreetheid. Het éne concept is al abstracter en moeilijk meetbaar en operatio-

naliseerbaar terwijl een ander concept redelijk concreet is en vrij direct geoperationaliseerd en geobserveerd kan worden. In de praktijk zijn de meeste concepten in een conceptuele definitie al een klein stukje abstract en zal je duidelijk moeten uitleggen in de operationele definitie wat je precies bedoelt met elk concept.

Neem opnieuw het cijfer 525 935 als voorbeeld 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 consulaire personeel) zijn vrijgesteld van inschrijving in de bevolkingsregisters. 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.

1. Wat is een statistiek?

Tabel 1.1: Operationaliseren betekent dat alle abstracte concepten in de conceptuele definitie worden vertaald naar concrete en observeerbare concepten in de operationele definitie

	Concept 1	Concept 2	Concept 3
conceptuele definitie	aantal inwoners	in Antwerpen	in 2019
	↓	↓	↓
operationele definitie	aantal inwoners volgens de wettelijke bevolking	in Antwerpen volgens NIS-code 11002 in 2019	op 1 januari 2019, 0.00 uur

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, deze 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 enquêtedesign, de steekproeftrekking, het ontwerp van de vragenlijst, de opvolging van respondenten, statistische correcties voor meet- en selectiefouten en de statistische analysemethoden.

De operationele definitie bevat een duidelijke operationalisering van elk concept afzonderlijk (zie Tabel 1.1). 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.

Het voorbeeld van de Antwerpse bevolkingsgrootte illustreert ook dat de meeste concepten in een conceptuele definitie al een stukje abstract zijn. Zelfs een eenvoudig concept als “bevolkingsgrootte” bevat al een zekere abstractie. De “grootte van de wettelijk bevolking” is duidelijk een con-

creter concept en kan worden gebruikt om het abstractere concept “bevolkingsgrootte” te meten.

1.1.4 Conceptuele versus operationele definitie

Zoals de voorbeelden hierboven duidelijk maakten, moeten tijdens de vertaalslag van conceptuele naar operationele definitie verschillende keuzes worden gemaakt. Die keuzes kunnen ervoor zorgen dat andere cijfers worden bekomen. Het cijfer 525 935 verwijst bijvoorbeeld naar de Antwerpse bevolking in 2019, maar wordt wel gemeten 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 wachtregisters 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 wachtregisters 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 Antwerps 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 en Vlaamse cijfers gewoon als de “bevolkingsgrootte”.

Het eerste doel van de operationele definitie is daarom om 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.

1. Wat is een statistiek?



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.

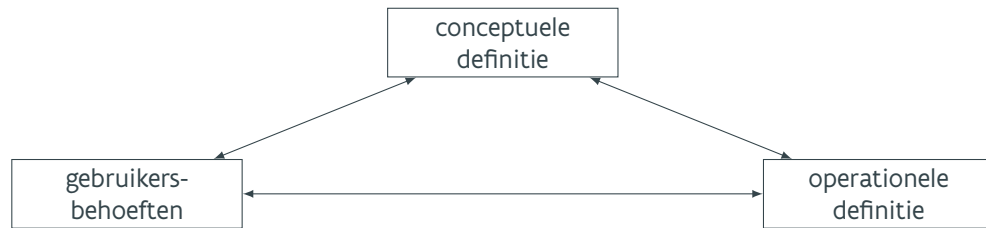
Het tweede doel van de operationele definitie is duidelijkheid te scheppen over de vergelijkbaarheid van cijfers. Je kan bijvoorbeeld niet zomaar de grootte van de wettelijke Antwerpse bevolking in 2019 vergelijken met de grootte van de verblijvende Antwerpse bevolking in 2020. Dit zou immers kunnen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

Je zou je nu kunnen 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 om de cijfers te rapporteren. 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 afratelen. 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.

1.1.5 Kwaliteit

We zagen dat een statistiek wordt beschreven door drie elementen. Ze ontstaat vanuit gebruikersbehoeften en deze behoeften worden door de statistische autoriteit vertaald naar een conceptuele definitie en vervolgens een operationele definitie (zie Figuur 1.2). In een ideale situatie zijn de conceptuele en operationele definitie afgestemd op de gebruikersbe-

Figuur 1.2: De kwaliteit van een statistiek wordt gedefinieerd door de verhouding tussen drie elementen.



hoeften. Wanneer dat niet het geval is heeft de statistiek een *kwaliteitsprobleem*. Dit kan zich voordoen bij alle soorten gebruikerscriteria binnen de gebruikersbehoefte.

Relevantie

Er is een probleem op het vlak van *relevantie* als onze conceptueel gedefinieerde statistieken niet in overeenstemming zijn met de gebruikersbehoefte. Dit probleem kan optreden in twee richtingen. Op de eerste plaats kan er behoefte zijn aan bepaalde statistieken maar produceren we ze niet. We vervullen onze rol als onafhankelijke openbare statistiekdienst dan uiteraard onvoldoende. Op de tweede plaats kunnen we echter ook statistieken produceren waar helemaal geen behoefte aan is. Ook in die situatie kan de kwaliteit van onze statistiekproductie verbeterd worden. We verliezen dan immers tijd en energie aan irrelevante statistieken terwijl we die tijd en energie ook kunnen inzetten op een andere manier.

Een relevantieprobleem kan ook ontstaan als er een wanverhouding ontstaat tussen gebruikersbehoefte en de operationele definitie van een statistiek. In ons Antwerps 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. In dit soort situaties moeten we onderzoeken of de operationele definitie aangepast kan worden aan de behoefte. In ons voorbeeld is misschien aangeraden 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 gebruikersbehoefte echter geen zwart-wit-verhaal. Er kunnen zich immers situaties voordoen waarin de gebruikte operationele definitie voldoet voor de ene gebruiker maar niet voor een andere gebruiker. Zo kan onze statistiek over bevolkingsgrootte op basis van de wettelijke bevol-

king 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 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 wat het oplevert.

Nauwkeurigheid

Er is een probleem op het vlak van *nauwkeurigheid* als een statistiek niet meet wat ze moet meten of, anders verwoord, de operationele definitie onvoldoende aansluit bij de conceptuele definitie. De nauwkeurigheid van statistieken hangt steeds af van gebruikte methoden en technieken om data te verzamelen en te analyseren.

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 genoeg nauwkeurig om een beeld te krijgen van de algemene energiezuinigheid van woningen in Vlaanderen.

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 nauwkeurigheid van de cijfers. In-

dien 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 “energiescores van geregistreerde woningen”. Uiteraard kan de aanpassing van de conceptuele definitie er toe leiden dat de statistiek haar relevantie verliest omdat de conceptuele definitie niet meer in lijn is met de gebruikersbehoeften. Beleidsmakers weten immers 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 gebruikte analysemethoden en hun beperkingen, en om de keuze voor de analysemethoden zo goed mogelijk te beargumenteren op basis van theoretische en empirisch wetenschappelijke studies.

Tijdigheid en Punctualiteit, Vergelijkbaarheid, Toegankelijk, en Wettelijk Verantwoord

Wanneer er problemen ontstaan rond *tijdigheid en punctualiteit, vergelijkbaarheid, toegankelijk, en wettelijk verantwoordheid* van de statistiek is dit ook doorgaans omwille van een verkeerd gekozen operationele definitie om de statistiek te meten. Zo kan een operationele definitie inhoudelijk aansluiten bij een gebruikersbehoefte, maar als de berekening te lang duurt en de statistiek te laat beschikbaar is, verlaagt dit alsnog de kwaliteit. Hetzelfde geldt voor toegankelijkheid en vergelijkbaarheid, statistiekgebruikers hebben uiteraard geen boodschap aan nauwkeurig gemeten statistieken als deze helemaal niet toegankelijk zijn of vergelijkbaar met eerder gepubliceerde cijfers. De wettelijke verantwoordheid van een statistiek is uiteraard een sine qua non voordat een gebruiker de statistiek kan raadplegen. Statistieken die verzameld werden op illegale manieren kunnen per definitie niet ontsloten worden en hebben zo geen enkele waarde voor de gebruikers.

Kwaliteitsevaluatie

Concluderend is het dus belangrijk om dat we als statistiekorganisatie diepgaand onderzoek uitvoeren naar gebruikersbehoeften en deze grondig in kaart brengen via *kwaliteitsevaluatie* (**referenties**). 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? Daarnaast is er ook voortdurend onderzoek nodig naar de kwaliteit van de statistieken en innovatie. Zijn de statistieken nog steeds nauwkeurig en kosten-efficiënt? Kunnen we de statistieken voldoende tijdig opleveren? Kunnen we de operationele definitie niet verbeteren door gebruik te maken van nieuwe dataverzamelings- en analysemethoden? Al dit soort onderzoek valt onder de term *kwaliteitsevaluatie*.

Hoewel onderzoek naar bovenstaande vragen niet altijd een exacte wetenschap is, vormen ze een belangrijke stap in de productie en publicatie van openbare statistieken. Voor zo'n onderzoek moeten we bovendien verder kijken dan de kwantitatieve methoden waarmee statistiek vaak geassocieerd worden. Binnen gebruikersonderzoek kan ook veel kennis worden vergaard door tekstuele analyse van wetteksten en beleidsdocumenten of door kwalitatief onderzoek via diepte-interviews, focusgroepen, participerende observaties en case studies bij gebruikersgroepen. De literatuur rond evaluatie-onderzoek biedt hierbij een kapstok (**referenties**).

We moeten hierbij ook opmerken dat het in de praktijk natuurlijk 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.

1.1.6 Attributen

Voor we deze paragraaf afsluiten moeten we ook kort stilstaan bij attributen. Naast de operationale definitie van een cijfer, kan er ook informatie over een cijfer beschikbaar zijn die geen betekenis geeft aan het cijfer. Dit worden *attributen* van de cijfers genoemd.

In principe zijn de kwaliteitskenmerken zoals de relevantie, nauwkeurigheid, tijdigheid en punctualiteit van statistieken voorbeelden van zo'n at-

tributen. Naast deze kwaliteitskenmerken zijn er echter nog andere belangrijke attributen.

Een belangrijk attribuut is de vertrouwelijkheid van het cijfer. Kan het cijfer worden gedeeld met het brede publiek of met collega's of is het vertrouwelijk en kan het niet zomaar gedeeld worden? Hoewel de vertrouwelijkheid van een cijfer geen betekenis geeft aan het cijfer zelf, is het uiteraard belangrijk om dit soort informatie ook nauwgezet te bewaren.

Een ander belangrijk attribuut is het eigenaarschap van een cijfer. Het eigenaarschap van een officiële statistiek verwijst naar de verantwoordelijkheid en het mandaat dat een bepaalde instantie heeft over het verzamelen, produceren, publiceren en beheren van die statistiek. In de context van officiële statistieken houdt dit eigenaarschap doorgaans ook de verantwoordelijkheid in om kwaliteit en betrouwbaarheid te garanderen, een mandaat om het cijfer te publiceren, en de verantwoordelijkheid om definities, classificaties, bronnen en methoden te beheren die bij de statistiek horen. Het is belangrijk steeds duidelijkheid te hebben over het eigenaarschap van cijfers omdat dit coherentie en vertrouwen garandeert in officiële cijfers, dubbel werk en tegenstrijdige cijfers voorkomt tussen instellingen en duidelijk maakt wie aanspreekbaar is bij vragen of kritiek.

1.1.7 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. informatie over attributen inclusief een evaluatie van de kwaliteit.

Hieronder vatten we nog eens alle informatie samen voor de bevolkingsgrootte van gemeente Antwerpen in 2019 bij wijze van voorbeeld. In de praktijk zal er uiteraard veel meer informatie beschikbaar zijn dan wat hieronder staat opgesomd.

- *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 ...

1. Wat is een statistiek?

- 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
- *Attributen.*
 - Evaluatie relevantie: De gebruikers hebben ook nood aan de grootte van de verblijvende bevolking naast de wettelijke bevolking
 - Evaluatie tijdelijkheid en punctualiteit: Dit cijfer wordt elk jaar in maart gepubliceerd volgens een afgesproken kalender. Er treedt geen vertraging op.
 - Vertrouwelijkheid: Dit cijfer mag publiek worden gedeeld
 - Eigenaar: Statbel

1.2 STATISTIEK ALS EEN CIJFERTABEL

1.2.1 Wat is een cijfertabel?

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 inwonersaantallen van alle Vlaamse gemeenten.

Tabel 1.3: We publiceren statistieken doorgaans niet als individuele cijfers maar als cijfertabellen.

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

1.2.2 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 statistiekreeks 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 de gemeente waarnaar ze verwijzen.

De concepten die enkel bijkomende informatie geven zonder de cijfers verder te identificeren worden daarentegen *attributen* genoemd. Attributen

1. Wat is een statistiek?

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 ??.

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 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 voorzetsels 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 ge-

1. Wat is een statistiek?

bied (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 expliciet 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 aantallen 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.

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 gebruikersbe-

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

hoeften 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.

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.

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

1. Wat is een statistiek?

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, opgelijste gebruikersbehoeften en een kwaliteitsevaluatie.

<p>Aantal mannen en vrouwen, versie 1</p> <p><i>gebruikersvoorwaarden:</i></p> <ul style="list-style-type: none">→ 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 ...→ ... <p><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 wachtregister en ambassadeurs) op 1 januari van elk kalenderjaar vanaf 2005 per Vlaamse gemeente volgens de NIS-code-indeling in 2019 = 600 cijfers per jaar.</p> <p><i>Kwaliteit:</i> De gebruikers hebben eerder nood aan de grootte van de verblijvende bevolking in plaats van de feitelijke bevolking.</p> <p><i>Nota:</i> Stopgezet in 2022, wegens aanpassing definitie.</p>
<p>Aantal mannen en vrouwen, versie 2</p> <p><i>gebruikersvoorwaarden:</i></p> <ul style="list-style-type: none">→ 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 ...→ ... <p><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 wachtregister 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.</p> <p><i>Kwaliteit:</i> Onderzoek XX toont aan dat er geen problemen zijn met deze statistiekreeks.</p> <p><i>Nota:</i> Stopgezet in 2025, wegens aanpassing gemeenten door fusies.</p>
<p>Aantal mannen en vrouwen, versie 3</p> <p><i>gebruikersvoorwaarden:</i></p> <ul style="list-style-type: none">→ 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 ...→ ... <p><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 wachtregister 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.</p> <p><i>Kwaliteit:</i> Onderzoek XX toont aan dat er geen problemen zijn met deze statistiekreeks.</p>
<p>Tewerkstelling in hoogtechnologische sector, versie 1</p> <p><i>gebruikersvoorwaarden:</i></p> <ul style="list-style-type: none">→ Decreet XX.XX 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.→ ... <p><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 omhelst ...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.</p> <p><i>Kwaliteit:</i> Onderzoek YY toont aan dat er geen problemen zijn met deze statistiekreeks.</p>
<p>Drinkwaterkwaliteit, versie 1</p> <p><i>gebruikersvoorwaarden:</i></p> <ul style="list-style-type: none">→ Decreet XX.XX verwijst naar drinkwaterkwaliteit in kader van ...→ ... <p><i>Operationele definitie:</i> Conformiteitspercentage van het kraantjeswater in heel Vlaanderen per jaar. Het conformiteitspercentage wordt berekend door Vlaamse Milieumaatschappij (VMM) op basis van het totale aantal analyses en het totale aantal vastgestelde normoverschrijdingen voor volgende parameters: ...= 1 cijfer per jaar.</p> <p><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.</p>
<p>....., versie ...</p> <p><i>gebruikersvoorwaarden:</i> ...</p> <p><i>Operationele definitie:</i> ...</p> <p><i>Kwaliteit:</i> ...</p> <p>.....</p>

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 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.

BIBLIOGRAFIE

- Clark, Tom e.a. (2021). *Bryman's Social Research Methods*. 6th edition. Oxford: Oxford University Press.
- Eurostat (2017). *European Statistics Code of Practice – Revised Edition 2017*. Accessed: 2025-07-18. URL: <https://ec.europa.eu/eurostat/web/quality/european-quality-standards/european-statistics-code-of-practice>.
- Neuman, W. Lawrence (2014). *Social Research Methods: Qualitative and Quantitative Approaches*. 7th edition. Boston: Pearson.
- Wild, C.J., J.M. Utts en N.J. Horton (2018). "What Is Statistics?" In: *International Handbook of Research in Statistics Education*. Red. door D. Ben-Zvi, K. Makar en J. Garfield. Springer International Handbooks of Education. Cham: Springer, p. 5–36. DOI: 10.1007/978-3-319-66195-7_1.