

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. 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? Waarom publiceren we dat getal? Hoe werd het getal gemeten? Hoe kwalitatief is deze meting? Wie is verantwoordelijk voor dit cijfer? Statistiek gaat dus niet alleen over het berekenen van cijfers, het gaat ook over het beheer van allerhande informatie die bij het cijfer hoort. Deze informatie noemen we *attributen* van het cijfer. Een *statistiek* bestaat dus altijd uit twee onlosmakelijke elementen: het cijfer zelf (ook wel *data* genoemd), én de attributen van dat cijfer (ook wel *metadata* genoemd). Beide zijn even belangrijk en vormen één geheel (Wild, Utts en Horton 2018). In dit geval betekent het cijfer 525 935 trouwens “het aantal inwoners van de gemeente Antwerpen in 2019”.

In deze paragraaf gaan we dieper in op verschillende belangrijke attributen van een statistiek. Het eerste belangrijk attribuut is de gebruikersbe-

1. Wat is een statistiek?

hoeft die vertelt waarom een statistiek wordt geproduceerd en gepubliceerd. Vanuit die gebruikersbehoefte volgt steeds de conceptuele definitie die betekenis geeft aan de statistiek. De operationele definitie vertelt dan weer hoe een statistiek exact werd gemeten en de relatie tussen de gebruikersbehoefte, conceptuele en operationele definitie bepaalt de kwaliteit van de statistiek. Al deze brokken informatie moeten worden voorzien om kwaliteitsvolle statistieken te produceren.



Merk op dat statistieken niet altijd cijfers zijn, 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 echter enkel naar statistieken als eenvoudige cijfers. Cijfers zijn immers veruit de meest voorkomende vorm van statistieken.

1.1.1 Gebruikersbehoeften

Als we een statistiek publiceren, 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 behoefte aan inhoudelijk relevante en nauwkeurige statistieken. *Relevant* betekent dat de statistiek beantwoordt aan de behoeften van een gebruiker. *Nauwkeurig* betekent dan weer dat de statistiek ook meet wat ze moet meten. Beide criteria zijn sterk aan el-

kaar 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.

De productie van relevante en nauwkeurige statistieken vormt één van de beginselen van de "Praktijkcode voor Europese statistieken" (zie Eurostat 2017). Deze praktijkcode schept een kwaliteitskader voor het Europees statistisch systeem (ESS) en legt 16 beginselen vast voor de ontwikkeling, productie en verspreiding van Europese statistieken. Naast relevantie en nauwkeurigheid, somt de praktijkcode echter nog andere verwachtingen op van gebruikers. Gebruikers kunnen bijvoorbeeld ook verwachten dat statistieken *tijdig* en *punctualiteel* zijn. Dit betekent dat de statistieken snel en op duidelijke tijdstippen worden gepubliceerd. Dit laat beleidmakers 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 *toegangkelijk* 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.



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



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

1. Wat is een statistiek?

afstemmen en daardoor relevante en nauwkeurige statistieken opleveren.

Beginsel 2 bepaalt dat een statistische autoriteit een wettelijk mandaat krijgt voor de verzameling van en toegang tot gegevens. Beginsel 3 stelt op haar beurt dat statistische autoriteiten over voldoende personele, financiële en technische middelen moeten beschikken. Deze beginselen zorgen voor garantie dat gebruikersbehoeften optimaal in kaart worden gebracht en dat aan deze behoeften voldaan wordt met relevante en nauwkeurige statistieken.

Beginsel 4 vraagt om een sterk kwaliteitsbewustzijn bij statistische autoriteiten. Dit beginsel hangt nauw samen met beginsel 7, dat vereist dat statistische autoriteiten deugdelijke methoden gebruiken om kwaliteit voortdurend te optimaliseren, en beginsel 8, dat vraagt dat geschikte statistische procedures worden gebruikt door de statistische autoriteiten. Deze beginselen zeggen niet meer en niet minder dan dat statistische autoriteiten moeten streven naar maximale nauwkeurigheid in hun statistieken.

1.1.2 Conceptuele definitie

Om gebruikersbehoeften te voldoen, moet een statistische autoriteit cijfers publiceren met een bepaalde betekenis. 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 stukje 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 vaag 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*.

De operationalisering van een conceptuele definitie betekent in de praktijk dat elk concept in de definitie vertaald wordt van een abstract idee naar een concrete richtlijn om het te meten. 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 buiten gesloten 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-

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

naliseerbaar terwijl een ander concept redelijk concreet is en vrij direct geoperationaliseerd en geobserveerd kan worden.

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 consulair 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.

Verder bevat het Rijksregister ook een wachtregister voor EU burgers in afwachting van woonstcontrole. Deze personen worden evenmin meegeteld bij de wettelijke bevolking.

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

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 enquête-design, de steekproef trekking, het ontwerp van de vragenlijst, de opvolging van respondenten, statistische correcties voor meet- en selectiefouten en de statistische analysemethoden.

Bovenstaande 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 toont ook dat de concepten in een conceptuele definitie doorgaans abstract zijn. 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 om het abstractere concept “bevolkingsgrootte” te meten.

Tijdens de vertaalslag van conceptuele naar operationele definitie moeten keuzes worden gemaakt. Die keuzes kunnen ervoor zorgen dat andere

1. Wat is een statistiek?

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.

Conceptuele definities kunnen dus vaak vertaald worden naar meerdere operationele definities, elk met eigen operationele keuzes, en 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.



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 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 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

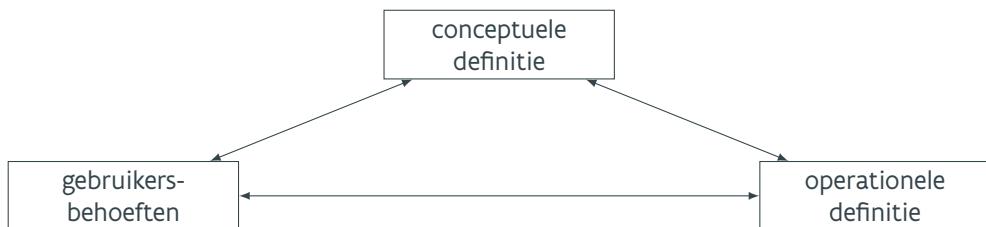
We zagen reeds dat een statistiek wordt beschreven door drie attributen. Ze ontstaat vanuit gebruikersbehoeften en deze behoeften worden door de statistische autoriteit vertaald naar een conceptuele definitie en vervolgens een operationele definitie (zie Figuur 1.2). In een ideale situatie zijn de conceptuele en operationele definitie afgestemd op de gebruikersbehoeften. Wanneer dat niet het geval is heeft de statistiek een *kwaliteitsprobleem*. Dit kan zich voordoen bij alle soorten gebruikerscriteria binnen de gebruikersbehoeften.

Relevante

Er is een probleem op het vlak van *relevantie* als de conceptueel gedefinieerde statistieken niet in overeenstemming zijn met de gebruikersbehoeften. Dit probleem kan optreden in twee richtingen. Op de eerste plaats kan er behoefte zijn aan bepaalde statistieken maar produceren we ze

1. Wat is een statistiek?

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



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 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. In dit soort situaties moeten we onderzoeken of de operationele definitie aangepast kan worden aan de behoeften. In ons voorbeeld is het 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 gebruikersbehoeften 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 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 verschil-

lende 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 ééne 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 energieuinigheid 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. 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 “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.

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

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 argumenteren 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 conditio 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 statistieken. 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 innova-

tieve 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 gebruikeronderzoek 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 gebruikeronderzoek. Daarom moeten we voor elke statistiek in de VSP lijst we ten minste volgende informatie voorzien:

1. een overzicht van de gebruikersbehoeften,
2. de conceptuele definitie inclusief een beschrijving van alle concepten
3. de operationele definitie inclusief een operationalisering van alle concepten, en
4. een kwaliteitsevaluatie van de conceptuele en operationele definitie in functie van de gebruikersbehoeften.

1.1.5 Andere administratieve attributen

De discussie hierboven heeft ondertussen aangetoond dat een statistiek bestaat uit een cijfer en een heleboel attributen zoals de gebruikersbehoeften, conceptuele en operationele definitie, en de kwaliteitsevaluatie. Daarnaast kunnen echter nog een heel wat andere attributen worden vastgesteld. Een belangrijk bijkomend attribuut is bijvoorbeeld de vertrouwelijkheid van een 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 registreren en te bewaren.

1. Wat is een statistiek?

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.

Samengevat, voor elke statistiek die we produceren berekenen we uiteraard een cijfer maar registreren we ook alle relevante attributen die bij dat cijfer horen zoals:

1. de gebruikersbehoeften,
2. een conceptuele definitie op basis van de gebruikersbehoeften met afbakening van alle individuele concepten,
3. een operationele definitie met de operationalisering van alle concepten,
4. informatie over de kwaliteit van het cijfer, en
5. bijkomende administratieve attributen.

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.

- *Cijfer:*
525 935
- *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 ...
- *Kwaliteit:*
 - Relevantie: De gebruikers hebben ook nood aan de grootte van de verblijvende bevolking naast de wettelijke bevolking
 - Tijdelijkheid en puncualiteit: Dit cijfer wordt elk jaar in maart gepubliceerd volgens een afgesproken kalender. Er treedt geen vertraging op.
 - ...
- *Andere attributen:*
 - Vertrouwelijkheid: Dit cijfer mag publiek worden gedeeld
 - Eigenaar: Statbel

1.2 STATISTIEK ALS COMBINATIE VAN CIJFERS

In de vorige paragraaf werd uitgelegd hoe een statistiek verwijst naar één enkel cijfer met bijhorende informatie-attributen waaronder de conceptuele en operationele definitie ontstaan vanuit specifieke gebruikersbehoeften. In de praktijk produceren we echter vaak verschillende 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. Het is in dat geval uiteraard niet efficiënt om attributen cijfer per cijfer te beheren en te ontsluiten. Om efficiëntie te verhogen groeperen we daarom cijfers bij elkaar. Dat doen we via cijfertabellen of zelfs reeksen van cijfertabellen zoals uitgelegd in deze paragraaf.

1.2.1 Cijfertabellen

Een *cijfertabel* (of cijferreeks) verwijst naar een verzamelingen van cijfers die grotendeels dezelfde conceptuele en operationele definitie hebben. Tabel 1.3 toont bijvoorbeeld een cijfertabel waarin we niet alleen het inwoneraantal van Antwerpen in 2019 zien, maar ook de inwoneraantallen van alle andere Vlaamse gemeenten in 2019. Deze tabel bevat voor elk cijfer een aparte conceptuele definitie maar je merkt al gauw dat er veel overlap zit tussen deze definities. Om die reden zullen we de attributen

1. Wat is een statistiek?

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

van een cijfertabel anders organiseren dan de attributen van een individueel cijfer.

Concepten, dimensies, attributen

Om te beginnen kunnen we binnen de conceptuele en operationele definities van cijfers in een cijfertabel een onderscheid maken tussen twee soorten concepten. Sommige concepten variëren over de cijfers en helpen elk cijfer binnen de tabel uniek te identificeren, terwijl andere concepten constant blijven over alle cijfers. De concepten die variëren over de cijfers noemen we de *dimensies* van de cijfertabel, maar 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 in 2019. De cijfers verschillen dus enkel van elkaar door de gemeente waarnaar ze verwijzen.

De concepten die constant zijn over alle cijfers blijven we daarentegen eenvoudigweg *attributen* noemen. Omdat deze attributen enkel bijkomende informatie geven zonder de cijfers verder te identificeren worden ze 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.

Merk op dat attribuutconcepten van een cijfertabel dimensies kunnen worden wanneer verschillende cijfertabellen worden samengevoegd en omgekeerd. Als we bijvoorbeeld een tabel met inwoneraantallen definiëren over verschillende jaren heen in plaats van enkel 2019, wordt jaartal

een bijkomende dimensie van die cijfertabel. Omgekeerd kan een dimensie ook een attribuutconcept 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.

Conceptuele en operationele definitie

Aangezien een cijfertabel bestaat uit cijfers met sterk overlappende conceptuele en operationele definities, kunnen we verder ook deze definities veralgemenen naar het niveau van de hele tabel. Dit maakt informatiebeheer efficiënter.

De *conceptuele definitie* van een cijfertabel verwijst nog steeds naar de interpretatie die een doorsnee gebruiker aan de cijfers toekent. Die conceptuele definitie bevat nog steeds alle concepten, al kunnen sommige concepten verwijzen naar dimensies in plaats van louter attributen. De cijfertabel in Tabel 1.3 kan bijvoorbeeld conceptueel worden gedefinieerd als het “aantal inwoners in gemeenten in het Vlaamse Gewest in 2019”. De concepten zijn in deze definitie nog steeds de grootheid inwoneraantal, het jaartal 2019 en gemeente. Gemeente is echter een dimensie in deze tabel. Voor de exacte interpretatie van de individuele cijfers moet de gebruiker daarom naar de waarden op deze dimensie kijken.

De *operationele definitie* van een cijfertabel beschrijft hoe de cijfertabel precies werd verzameld en gereproduceerd kan worden. Bovendien preciseert de operationele definitie steeds welke waarden de attributen en dimensies binnen de tabel 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...”

Door de expliciete en exacte beschrijving van dimensies bepaalt de operationele definitie steeds uit hoeveel cijfers een cijfertabel precies bestaat, zelfs als sommige cijfers een missende of versluierde waarde hebben. De operationele definitie van Tabel 1.3 benoemt bijvoorbeeld 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.

1. Wat is een statistiek?

Tabel 1.4: Bij cijfertabellen worden attribuutconcepten opgenomen in de conceptuele definitie in de titel. De rijen en kolommen bevatten enkel de dimensies.

Aantal en percentage inwoners per geslacht in Vlaamse gemeenten 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
...

Een kleine tip om de operationele definitie duidelijk uit te schrijven: Maak slim gebruik van voorzetters 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 Vlaamse gemeente (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).” Pas hierbij goed op dat je duidelijk bent bij percentages. “Het percentage inwoners per geslacht in elke gemeente” is niet hetzelfde als het “het percentage inwoners per gemeente voor elk geslacht”.

De presentatie van een statistiekreeks gebeurt doorgaans niet zoals in tabel 1.3. Attribuutconcepten worden vaak uit de rijen verwijderd en opgenomen in conceptuele definitie van de cijfertabel, die gebruikt wordt als titel. Tabel ?? toont een meer gangbare weergave van een uitgebreidere cijfertabel. In deze tabel zie je dat het concept jaartal “2019” enkel wordt vermeld in de titel van de tabel. De tabel telt verder drie dimensies. Dimensies gemeente en geslacht staan in de rijen. De dimensie grootheid, die hier een verschil maakt tussen aantalen en percentages, staat in de kolommen.

Soorten attributen

Een cijfertabel kan uiteraard ook nog steeds vele andere attributen hebben. Net zoals de conceptuele en operationele definities met bijhorende concepten, kan informatie over gebruikersbehoeften of de kwaliteitseva-

luatione veralgemeend worden van het niveau van individuele cijfers naar het niveau van de hele cijfertabel. Attributen kunnen in een cijfertabel dan ook worden toegewezen op drie niveaus.

Een eerste soort attributen zijn attributen op het niveau van de hele cijfertabel. Deze attributen bevatten informatie die geldt voor alle cijfers in de tabel. In Tabel 1.3 is het jaartal 2019 een voorbeeld van zo'n attribuut aangezien alle cijfers verwijzen naar bevolkingsgroottes in 2019. Dit soort attributen hoeft je uiteraard enkel op tabelleniveau te bewaren en niet apart per cijfer.

De tweede soort attributen zijn attributen op het niveau van individuele cijfers. Zij bevatten dus informatie die slechts geldt voor één enkel cijfer en niet voor alle andere cijfers in de tabel. Een voorbeeld van zo'n attribuut is informatie over de vertrouwelijkheid van elk cijfer. Bepaalde cijfers in een tabel kunnen beschouwd worden als vrij te publiceren terwijl andere cijfers niet openbaar gemaakt mogen worden omdat ze vertrouwelijke informatie kunnen ontsluiten. Opgelet, de aanduiding van deze vertrouwelijkheid identificeert de cijfers niet en is dus geen dimensie van de tabel, ook al varieert deze informatie wel van cijfer tot cijfer.

Een derde soort attributen bevinden zich op een tussenniveau, namelijk attributen die betrekking hebben op bepaalde groepen van cijfers. Zo kunnen, hypothetisch, de cijfers in Tabel 1.3 uit één bepaalde provincie misschien op een andere manier zijn gemeten. De cijfers uit de verschillende provincies vormen dan groepen en per groep heb je een andere operationele definitie als attribuut.

1.2.2 Tabelreeksen

Om statistieken te definiëren kunnen we zelfs nog een stap verder gaan dan cijfertabellen. Het kan voorkomen dat we cijfers publiceren die nog steeds sterk overlappende conceptuele en operationele definities hebben maar die niet gemakkelijk te organiseren zijn in één gestructureerde dimensionale tabel. In die situatie kunnen we ervoor kiezen om statistieken te organiseren in *reeksen van cijfertabellen*.

Zoals we in vorige voorbeelden al toonden, publiceren we bij Statistiek Vlaanderen heel wat cijfers over het aantal inwoners in Vlaamse gemeenten. Dat doen we niet alleen per gemeente en per jaar, we publiceren ook cijfers volgens verschillende demografische groepen. Zo publiceren we bijvoorbeeld cijfers over het aantal mannen en vrouwen en over het aantal inwoners in verschillende leeftijdsgroepen in de Vlaamse gemeenten. Al deze cijfers worden op een gelijkaardige manier afgeleid uit het

1. Wat is een statistiek?

Tabel 1.5: Tabelreeksen zijn verzamelingen van cijfertabellen met sterke overlap tussen de conceptuele en operationele definities.

Aantal inwoners naar geslacht en leeftijd in Vlaamse gemeenten in 2019.		
Volgens geslacht.		
Gemeente	Geslacht	Aantal
Aartselaar	man	7089
Aartselaar	vrouw	7204
Antwerpen	man	262921
Antwerpen	vrouw	263014
Boechout	man	6506
Boechout	vrouw	6760
Boom	man	9024
Boom	vrouw	9220
Borsbeek	man	5270
Borsbeek	vrouw	5584
...

Volgens leeftijdsgroep.		
Gemeente	Leeftijd	Aantal
Aartselaar	0-18 jaar	7089
Aartselaar	19-35 jaar	7204
Aartselaar	36-64 jaar	7089
Aartselaar	65+ jaar	7204
Antwerpen	0-18 jaar	262921
Antwerpen	19-35 jaar	263014
Antwerpen	36-64 jaar	262921
Antwerpen	65+ jaar	263014
Boechout	0-18 jaar	6506
Boechout	19-35 jaar	6760
...

Rijksregister. Het is dus efficiënter die afleiding op één en dezelfde plaats te documenteren en te beheren in plaats van voor elke cijfertabel apart.

Tabel 1.5 toont een voorbeeld van zo'n tabelreeks rond het aantal inwoners volgens geslacht en leeftijd. Deze vereenvoudigde tabelreeks bevat een overkoepelende conceptuele definitie als titel waarin nog steeds alle concepten aan bod komen, namelijk de grootheid het aantal inwoners, alle Vlaamse gemeenten en het jaartal 2019. Daarnaast vermeldt het ook de nieuwe concepten geslacht en leeftijd. De tabelreeks bestaat uit twee cijfertabellen, één voor het aantal inwoners volgens geslacht en één volgens leeftijd. Elke tabel heeft nog steeds haar eigen conceptuele definitie als titel aangevuld met de titel van de hele tabelreeks.

Door de overlap tussen definities van de cijfers in een tabelreeks kunnen we uiteraard ook de operationele definitie opnieuw generaliseren naar de hele tabelreeks. De operationele definitie van de tabelreeks in Tabel 1.5 luidt bijvoorbeeld: "De grootte van de wettelijke bevolking op 1 januari 2019 per gemeente volgens de NIS code indeling in 2019 volgens (1) geslacht zoals geregistreerd in het Rijksregister (man/vrouw), en (2) leeftijd zoals geregistreerd in het Rijksregister en opgedeeld in vier categorieën (0-18 jaar, 19-35 jaar, 36-64 jaar en 65+ jaar). De wettelijke bevolking omvat...". Merk op dat de operationele definitie een heel duidelijke opsomming geeft van de cijfertabellen in de tabelreeks, namelijk een tabel volgens geslacht en een tabel volgens leeftijd. Daarnaast is deze definitie nog steeds duidelijk expliciet over hoe de verschillende concepten worden gemeten en opgedeeld zoals geslacht en leeftijd. Uit de operationele definitie van de

tabelreeks moet je dan ook de operationele definitie van elke individuele cijfertabel in de reeks kunnen afleiden.

Het spreekt voor zich dat we nu ook attributen op een vierde niveau kunnen definiëren. Naast attributen van individuele cijfers, groepen cijfers en cijfertabellen, kunnen er immers ook attributen zijn die gelden voor alle cijfers in de hele tabelreeks. Het kan hierbij nog steeds gaan over alle soorten attributen zoals concepten in de conceptuele en operationele definities, kwaliteitsevaluatie, gebruikersbehoeften, en andere administratieve metadata. Net zoals de operationele definitie, moeten gebruikersbehoeften en kwaliteitsevaluatie nog steeds gedefinieerd zijn zodat ze alle cijfers in de tabelreeks rechtvaardigen en evalueren. In Tabel 1.5 zijn de grootheid aantal inwoners en het jaar 2019 attributen op het niveau van de tabelreeks, want alle cijfers in deze tabelreeks verwijzen naar aantallen inwoners in 2019.



Tabelreeksen zijn erg geschikt voor de weergaven van meervoudige concepten. Naast concreet en abstract, kan je concepten ook opsplitsen in enkelvoudige en meervoudige concepten.

- *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 met één cijfer 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.*

Zowel enkelvoudige als meervoudige concepten kunnen concreet en abstract zijn, al zullen meervoudige concepten in de praktijk doorgaans vaker abstract zijn dan concreet.

- *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.*

1. Wat is een statistiek?

- *Woningkenmerken (bv. oppervlakte, aantal kamers, type dak) is een concreet en meervoudig concept. De verschillende woningkenmerken zijn stuk voor stuk direct 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 enkele vraag: "Hoe tevreden bent u over uw werk?". Met zo'n vraag beschouw je het als een enkelvoudig concept, al kan je discussiëren of zo'n enkele vraag de nuances in arbeidstevredenheid wel voldoende vat.*
- *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.*

1.3 EXPERIMENTELE STATISTIEKEN

Om steeds beter en tijdig in te spelen op de behoeften van gebruikers is het voor een statistiekinstelling belangrijk steeds onderzoek te blijven uitvoeren naar nieuwe data verzamelings- en analysemethoden. Uit de toepassing van zo'n nieuwe methodes kunnen nieuwe cijfers of statistieken ontstaan die nog niet geofficialiseerd werden (VSP-lijst met bijhorende afspraken over de operationalisatie). De kwaliteit van zo'n statistieken is echter niet altijd meteen duidelijk en moet eerst diepgaander onderzocht worden. Dit soort statistieken worden daarom *experimentele statistieken* genoemd (**Eurostat2025Experimental**).

Eerder in dit hoofdstuk zagen we al dat de kwaliteit van een statistiek wordt bepaald door de verhouding tussen gebruikersbehoeften en de conceptuele en operationele definities. Experimentele statistieken kunnen daarom in meerdere vormen voorkomen. Soms ontstaan ze spontaan vanuit nieuwe data verzamelings- en analysemethoden maar is het niet duidelijk of ze relevant zijn omdat er geen duidelijke behoefte is aan de cijfers. Zo publiceren we bijvoorbeeld cijfers over [voorbeeld toevoegen] het maar weten we niet wie deze cijfers gebruikt en voor welk doeleind. Soms ontstaan experimentele statistieken ook als alternatief voor reeds

bestaande officiële statistieken (VSP-lijst) met duidelijk geformuleerde gebruikersbehoeften. In dat geval rijst steeds de vraag of de nieuwe statistieken kwalitatief beter zijn dan de bestaande statistieken op het vlak van, onder andere, accuraatheid of tijdigheid. Zo onderzochten we bijvoorbeeld of de analyse van Twitter-berichten een alternatief kan vormen voor burgerbevragingen om sociaal sentiment in Vlaanderen te meten. Beide methodes hebben uiteraard hun voor- en nadelen en onderzoek moet uitwijzen welke methode de beste resultaten oplevert.

Wat precies onder “nieuwe data verzamelings- en analysemethoden” valt, is trouwens rekbaar. Vaak denken mensen bij dit begrip aan recente, innovatieve, complexe top notch data science technieken. Een experimentele statistiek kan echter ook worden ontwikkeld met behulp van oude en eenvoudige methoden. Zo kan een eenvoudige vraag in een bestaande bevraging al gelden als een nieuwe databron, en kunnen hieruit nieuwe cijfers worden berekend en gepubliceerd waarvoor de gebruikersbehoeften nog niet duidelijk in kaart werden gebracht. Ook in deze situatie spreken we van een experimentele statistiek.

Bij experimentele statistieken is het belangrijk om steeds op voorhand een duidelijk evaluatiemoment in te plannen, bijvoorbeeld vijf jaar na de eerste publicatie. Tijdens dit evaluatieonderzoek worden onder andere volgende vragen gesteld:

- Wordt de statistiek effectief gebruikt, en zo ja, waarvoor?
- Komt het gebruik en de interpretatie overeen met de operationele definitie?
- Zijn er suggesties van gebruikers om de statistiek beter af te stemmen op hun noden?

Op basis van de resultaten van dit onderzoek kan de statistiek:

- worden geüpgraded tot een officiële statistiek;
- worden aangepast om beter aan te sluiten bij gebruikersbehoeften;
- worden stopgezet indien de kwaliteit onvoldoende blijkt.