

## De klok en de klepel

*E.P. Martens, Statisticor, 2016*

*Opmerkingen over de studie "Studie kostprijs en hoeveelheid zwerfvuil in 2013", KplusV, Arnhem, 2015.*

### 1. Inleiding

In opdracht van de OVAM is door KplusV een studie uitgevoerd naar de kostprijs en de hoeveelheid zwerfvuil in Vlaanderen in 2013. Dit onderzoek is te downloaden via onderstaande link:

["http://www.ovam.be/sites/default/files/atoms/files/Studie-kostprijs-en-hoeveelheid-zwerfvuil-in-2013-DEF.pdf"](http://www.ovam.be/sites/default/files/atoms/files/Studie-kostprijs-en-hoeveelheid-zwerfvuil-in-2013-DEF.pdf). Vanwege het indienen van een offerte voor een ander onderzoek, was het noodzakelijk dit rapport te bekijken. Omdat hierin nogal wat mis gaat en het een openbaar rapport betreft met maatschappelijke relevantie, leek het goed deze kritiekpunten nader aan te duiden. De volgende paragrafen gaan hierover.

### 2. Conclusie

De conclusie die getrokken moet worden uit onderstaande paragrafen is dat aan de uitkomsten van het onderzoek "Studie kostprijs en hoeveelheid zwerfvuil in 2013" in zijn geheel geen waarde kan worden toegekend omdat:

- er vele fouten zijn gemaakt,
- een te beperkte hoeveelheid informatie is gebruikt en
- kansen zijn gemist.

De schattingen omtrent hoeveelheden en kostprijs zwerfvuil kunnen om die redenen rechtsreeks 'indevuilbak'. Zonde van het gemeenschapsgeld dat hieraan is besteed.

De essentiële kritiek is te vinden in de paragrafen 3 t/m 5. In paragraaf 6 wordt de toelichting op de steekproeftrekking besproken die zeer lachwekkend is en die de inspiratiebron vormde voor de titel van dit artikel.

### 3. Steekproeftrekking of integrale waarneming

Bij ieder onderzoek is de steekproeftrekking van groot belang voor het onderzoek, niet alleen de grootte maar ook de wijze van steekproeftrekking. De eerste vraag moet echter zijn: *is er een steekproeftrekking nodig*, of kan de gehele populatie worden onderzocht gegeven de onderzoeksmethode?

Bij dit onderzoek is het volkomen onduidelijk waarom slechts een beperkte steekproef is getrokken, terwijl de informatie digitaal is verzameld; naast enkele werksessies, zijn de uitnodigingen digitaal toegestuurd en moest de enquête digitaal worden ingevuld (pag. 21 rapport). Uiteraard levert de ver-

werking meer werk op, maar de betrouwbaarheid bij integrale waarneming is aanzienlijk groter. Bovendien zit in de beperkte steekproef ook nog non-respons waarmee niets is gedaan (selectiviteit).

#### 4. Stratificatie

Bij de steekproeftrekking is stratificatie toegepast door de gemeenten in vier groepen in te delen (zwerfvuilrisicoprofielen). Op zichzelf een goede keuze, die waarschijnlijk is overgenomen van de Netheidsindex die eerder door Statisticor is ontwikkeld. Echter, de steekproeffractie is niet aangepast aan de *heterogeniteit en de omvang* van het stratum, hetgeen tot gevolg heeft dat de betrouwbaarheid per stratum heel verschillend zal zijn. Dat is een ernstige zaak, omdat bij de extrapolatie met die verschillen in betrouwbaarheid niets is gedaan.

#### 5. Extrapolatie

De uiteindelijke uitkomsten van het onderzoek zijn bepaald door *extrapolatie van de steekproefuitkomsten* naar de gehele Vlaamse populatie van gemeenten. Voor de kwaliteit van deze uitkomsten is de wijze van extrapolatie van esstieel belang. En juist op dit punt gaat het behoorlijk mis.

Allereerst heeft dat al te maken met de keuze een (beperkte) steekproef te trekken terwijl de methode van verzamelen volkomen elektronisch is gedaan. Met een iets grotere inspanning op dit punt had een veel grotere kwaliteit kunnen worden geleverd. Een suggestie waarop in het onderzoek en in het rapport bezuinigd had kunnen worden, kan eenvoudig gegeven worden: de gebakken lucht over de "Activity Based Costing" op pag. 17 en 18 had best mogen ontbreken.

Ten tweede is *niets met mogelijke selectiviteit gedaan* die door de non-respons kan zijn veroorzaakt. Door de volledig elektronische wijze van gegevens verzamelen kon worden verwacht dat er aanzienlijke non-respons zou optreden, hetgeen ook het geval was. Deze was 28% en in 1 stratum zelfs 39% (zie pag. 22). Omdat enige informatie over deze gemeenten aanwezig is, had het weergeven van de mate van selectiviteit en enige correctie op zijn plaats geweest.

Ten derde is het zo dat ondanks de stratificatie die is toegepast, de *steekproeffractie niet is aangepast aan de heterogeniteit en de omvang* van het stratum. Dat is typisch een geval van een gemiste kans, die zonder veel inspanningen gegrepen had kunnen worden. Dat heeft tot gevolg dat de betrouwbaarheid per stratum heel verschillend zal zijn.

Als al het voorgaande even terzijde wordt geschoven, is er nog een essentieel punt dat bij de extrapolatie over het hoofd is gezien. Er is namelijk in dit onderzoek slechts *lineair naar inwonertal* geëxtrapoléerd. Dat is beschreven op pag. 24 ("*Voor de berekeningen zijn de verkregen gegevens omgerekend naar gemiddelde kengetallen (bijvoorbeeld VTE), kosten en gemiddelde hoeveelheden per inwoner per stratum. Deze gemiddelden zijn vervolgens opgeschaald door het aantal inwoners per stratum te vermenigvuldigen met het gemiddelde.*").

Deze extrapolatie is *verkeerd en zeer beperkt*. De extrapolatie is verkeerd, omdat de relatie tussen hoeveelheden/kosten en inwonertal helemaal niet

lineair verloopt. Zo is er altijd sprake van vaste kosten, en is een dergelijke relatie bij toeristengemeenten theoretisch al helemaal onzinnig. Zie hiervoor ook het rapport Kostenonderzoek zwerfafval in Nederland van Deloitte uit 2010.

De extrapolatie is zeer beperkt, omdat door toevoeging van enkele extra factoren (naast inwonertal), een grote verbetering kon worden bereikt. De verschillen binnen een stratum zijn namelijk bijzonder groot (zie bovengenoemd rapport van Deloitte) en de steekproeffractie in dit onderzoek is beperkt, hetgeen een grote onzekerheid *binnen strata* oplevert (die bovendien sterk verschilt *tussen strata*).

Al met al zijn voldoende inhoudelijke argumenten en redenen aangegeven waarom aan de uitkomsten van de extrapolatie en dus de uitkomsten van het onderzoek geen waarde dient te worden toegekend.

## 6. Toelichting steekproeftrekking

Het leukste hebben we voor het laatste bewaard. Dat is de verwijzing naar en de toelichting op de steekproeftrekking.

In het rapport zelf wordt naar de steekproeftrekking in de bijlage verwezen (pag. 21, moet overigens bijlage 2 zijn ipv bijlage 3). Daarin lezen we: *"Als uitgangspunt voor de betrouwbaarheid van het beeld van de kosten en hoeveelheden voor alle Vlaamse gemeenten hebben we voor de steekproef een niveau van statistische significantie van 0.9 en een betrouwbaarheidsinterval van 10% aangehouden."*

Deze formulering was de inspiratiebron voor de titel van dit artikel. Mocht je nog denken dat deze formulering een 'slip of the pen' is, dan zie je al snel in Bijlage 2 dat dit niet het geval is. Sterker nog: die toelichting is werkelijk om je statistische vingers bij af te likken. Eerst even twee opmerkingen over bovenstaande formulering.

- Significantieniveaus liggen over het algemeen op 0,01-0,05 (ook wel tussen 1% en 5%) en niet op 0,9 (of 90%). De klok en de klepel is hier goed van toepassing.
- Datzelfde geldt voor de formulering 'betrouwbaarheidsinterval van 10%'. Een x%-betrouwbaarheidsinterval rondom de schatting van een parameter is een verzameling getallen die met een betrouwbaarheid van x% allemaal de werkelijke waarde zouden kunnen zijn. Een betrouwbaarheidsinterval kan niet simpelweg als '10%' worden aangeduid.

## Bijlage 2

Bijlage 2 van dit rapport is echt om te smullen. Daarin staat de toelichting op de steekproeftrekking op pag. 67-68. Er is weinig statistische kennis voor nodig om te zien dat dit van een zeer bedenkelijk niveau is; de klok klepelt er hier lustig op los. De fouten in de tekst en in de gebruikte formule stapelen zich op. Het begint al direct met de eerste zin: *"De voor het onderzoek gewenste steekproefomvang werd bepaald door het gewenste statistische significantieniveau van 0.9 en betrouwbaarheidsinterval van 10%."* Daarover is hierboven al het nodige gezegd, omdat een dergelijke formulering ook in de tekst zelf voorkwam.

Daarnaast wordt met een dergelijke formulering gesuggereerd dat de steekproefomvang in een onderzoek wordt bepaald door alleen deze twee factoren, hetgeen uiteraard onjuist is.

Het belangrijkste om een verantwoording omtrent de steekproefomvang mee te beginnen is *op welke parameter* de berekening gedaan wordt (kostprijs, kostprijs per inwoner, hoeveelheid zwerfvuil). Vervolgens dienen dan de (realistische) waarden genoemd te worden waarvan bij de berekening wordt uitgegaan (bijv. uit ander onderzoek, of o.b.v. aannemelijke schattingen). Daarna moet dan aangegeven worden hoe betrouwbaar de opdrachtgever deze parameter in het onderzoek geschat zou willen hebben. In dit onderzoek is 'kosten per inwoner' gebruikt, maar is gezien de doelstellingen van het onderzoek niet de juiste parameter om de berekening op te doen.

#### Formule N

Hieronder is de door KplusV gehanteerde formule voor de steekproefomvang weergegeven.

$$N = \left( \frac{SD}{SE} \right)^2$$

Allereerst is nergens te vinden op welke parameter de SD en SE betrekking hebben. Daarnaast is deze 'formule' vooral lachwekkend. Een bekende formule in de statistiek is dat de standaardfout (SE) van een parameter kan worden geschat m.b.v. de standaarddeviatie (SD) en het aantal waarnemingen (N), ofwel  $SE = SD/\sqrt{N}$ . Als dit in bovenstaande 'formule' wordt ingevuld, dan staat er om precies te zijn  $N=N$ . Het hoeft geen toelichting dat de informatiegraad daarvan precies 0 is. In ieder willekeurig statistiekboek kan nagekeken worden wat de juiste formule voor de berekening van de steekproefomvang zou moeten zijn.

Een ander zwak punt is er geen eindigheidscorrectie is toegepast. Aangezien de geplande steekproef fractie in dit onderzoek  $85/300 = 0,28$  (ruim  $> 0,10$ ) is, is een dergelijke correctie noodzakelijk.

#### Formule SD

Hoewel een formule voor de standaarddeviatie (te vinden in elk statistiekboek) volkomen zinloos bij een toelichting over steekproefomvang, wordt deze toch gegeven:

$$SD = \sqrt{\frac{\sum_{i=1}^x (i \cdot i - \bar{x})^2}{N_x}}$$

Zelfs dat bleek te lastig om die correct weer te geven. De relevante standaarddeviatie haal je namelijk direct als getal uit een ander rapport, of daarvan wordt een aannemelijke schatting gemaakt. Door KplusV is deze uit een klasse-indeling berekend uit het rapport Kostenonderzoek zwerfafval in Nederland van Deloitte uit 2010. Echter is er geen rekening gehouden met de frequenties en met het feit dat het om een steekproef gaat ipv een populatie. De schatting van € 5,36 is daarom duidelijk te laag en had € 6,55 moeten zijn.