
Statistische steekproefmethodes

Een handleiding tot statistisch verantwoorde selectie

Ine Coessens en Wim Heirman

1.	Inleiding	2018-3
2.	Definitie	2018-8
3.	Casus	2018-10
4.	Voorwaarden/criteria	2018-13
5.	Betrouwbaarheid van de steekproef	2018-16
5.1	Bepalen van het betrouwbaarheidsinterval	2018-17
6.	Steekproefmethodes en toepassing	2018-21
6.1	Enkelvoudige steekproefmethodes	2018-21
6.1.1	Enkelvoudige aselecte steekproef	2018-22
6.1.2	Enkelvoudige systematische aselecte steekproef	2018-23
6.1.3	Semi-systematische steekproef	2018-23
6.2	Gestratificeerde steekproef	2018-25
6.2.1	Casus: de verkoopfacturen van Bethune & Fils	2018-27
7.	Aandachtspunten	2018-30
7.1	Gekoppelde archiefstukken	2018-30
7.2	Steekproeven op onvolledig archief	2018-31
7.3	Combinatie van een aselecte steekproef met andere selectiemethodes	2018-32

2018-2 Statistische steekproefmethodes

8.	Besluit	2018-33
9.	Bibliografie	2018-34
10.	Appendix 1: Grootte van de steekproef	2018-36
11.	Appendix 2: De steekproef in Excel	2018-41

1. Inleiding

“Many case files are created, but few should be chosen. It is hoped that this article will help archivists to identify the few gems that have value and destroy the many which block their light from view.”¹ Met deze woorden besloot Terry Cook in 1991 zijn artikel over de waardering en selectie van grootschalige archiefreeksen. Deze gedachte vormt ook het uitgangspunt van onderstaand artikel, dat – net zoals de bijdrage van Cook – een instrument wil zijn voor de archivaris/records manager. Het richt zich in het bijzonder tot hen die binnen het scala aan selectiemethodes gekozen hebben voor de steekproef en deze op een statistisch verantwoorde manier wensen uit te voeren. Deze bijdrage wil een overzicht geven van wat een statistisch verantwoorde steekproef is, welke mogelijkheden er binnen deze methode zijn en hoe deze uitgevoerd moeten worden. Bovenal wil dit artikel echter een (wiskundig) inzicht in en begrip van de methode meegeven en de voor- en nadelen (en potentiële gevaren) van elke steekproefoptie op de archiefreeks schetsen aan de hand van praktische voorbeelden en grafische weergaven. Het artikel reikt zo een aantal basisprincipes en -inzichten aan aan de archivaris/records manager. Die zouden hem of haar in staat moeten stellen op een weldoordachte manier aan steekproefselectie te doen, rekening houdend met de impact die deze beslissing zal hebben op de voor steekproef in aanmerking komende archiefreeks.

Het spreekt uiteraard voor zich dat de keuze om slechts een deel van het geheel aan archiefstukken binnen een archiefreeks te bewaren, gebaseerd op statistische principes, gekaderd dient te worden binnen een ruimer selectie- en waarderingsproces. Binnen een dergelijk proces is zowel het af- en overwegen en/of combineren van selectiemethodes, al dan niet via een vooraf bepaalde en bedeneerde selectielijst, als het inwinnen van voorafgaande historische en contextuele informatie over de archiefreeks waarop de steekproef zal plaatsvinden, onontbeerlijk. De keuze zal uiteindelijk gemaakt worden op basis van een combinatie van intellectuele en praktische overwegingen. Bij de steekproef speelt vooral de aanzienlijke investering die de bewaring van grootschalige archiefreeksen vergt, evenals de intellectuele onoverzichtelijkheid van de massale hoeveelheid archiefstukken die, bij niet-handelen, historisch onderzoek op deze reeksen bemoeilijkt tot haast onmogelijk maakt, een doorslaggevende rol.² Een gedetailleerde omschrijving van waardering en selectie en het bijho-

1 Terry Cook, “‘Many are called, but few are chosen’: Appraisal guidelines for sampling and selecting case files,” *Archivaria* 32 (1991): 45.

2 Felix Hull, *The use of sampling techniques in the retention of records: a RAMP study with guidelines* (Parijs: United Nations Educational, Scientific and Cultural Organization, 1981): 7.

rende proces, evenals van het denken over deze begrippen en hun inhoud is echter niet de doelstelling van dit artikel.¹

Het artikel wil evenmin verderbouwen op het debat dat reeds over de steekproef als methode is gevoerd. Toch is het noodzakelijk om kort in te gaan op het betreffende debat en dit om het eigen standpunt terzake te verduidelijken en te beargumenteren. De uiteenzetting zal duidelijk maken waarom de steekproef als methode kan en moet beschouwd worden als een valabel instrument voor selectie. Bovendien zullen uit het debat en de literatuur een aantal noden blijken waaraan tot op heden onvoldoende aandacht werd besteed en waarop dit artikel een antwoord zal trachten te formuleren.

Wanneer we de literatuur over de steekproef als selectiemethode binnen de archivaliek erop nalezen, is de vraag of, en op basis van welke criteria, de steekproefmethode uitgevoerd moet worden, pertinent aanwezig. Het uitvoeren van een steekproef impliceert immers dat wat via deze methode geselecteerd wordt voor bewaring, het toekomstige totaal aan archiefmateriaal van een archiefreeks zal vertegenwoordigen. Het niet geselecteerde materiaal wordt onherroepelijk vernietigd en kan bijgevolg nooit meer geraadpleegd worden.² Dit vooruitzicht creëert angst en onzekerheid bij de archivaris/records manager, aangezien deze de volle verantwoordelijkheid draagt voor het verlies en de bewuste vernietiging van potentieel bruikbare en waardevolle informatie.³ Het hoeft dan ook niet te verwonderen dat de steekproefmethode door hen bestempeld wordt als “so difficult and dangerous an archival skill”.⁴

In lijn met de manier waarop de steekproef door de professional gepercipieerd wordt, is ook het debat over de methode, en dan vooral het debat over de invalshoek van waaruit deze zijn/haar selectie zou moeten uitvoeren, in de literatuur erg verdeeld.

In eerste instantie zijn er zij die zich principieel tegen de steekproef kanten, vanuit het idee dat elke impact op het door een organisatie of instelling gecreëerd archief omwille van onderzoeksdoeleinden neerkomt op een re-creatie of

1 Voor een uitgebreid overzicht van deze begrippen en een grondige analyse van de manier waarop over waardering en selectie werd gedacht en hoe de begrippen werden ingevuld doorheen de tijd en binnen een aantal toonaangevende landen, verwijzen we graag naar volgend werk: Paul Drossens, *Archief gewogen: een onderzoek naar de theorie en praktijk van waardering en selectie* (Brussel: Algemeen Rijksarchief, 2011).

2 Hull, *The use of sampling techniques in the retention of records*, 10.

3 G.M Baldé en H. Spijkerman, “De kunst van het selecteren,” *Nederlandsch Archievenblad* 1 (1990): 304.

4 Hull, *The use of sampling techniques in the retention of records*, 10.

een vervalsing van het betrokken archief. De archivaris/records manager dient zich hiermee in hun optiek per definitie niet in te laten.¹

Vervolgens zijn er diegenen die de archiefreeks zullen beschouwen vanuit haar functie, of vanuit de activiteiten/verantwoordelijkheden die ze van een organisatie of instelling weerspiegelt. De steekproef heeft tot doel een algemeen beeld te schetsen van wat deze activiteiten/verantwoordelijkheden inhouden en moet per definitie uitgevoerd worden door de archivaris. Alleen hij/zij is in staat om een correcte inschatting te maken van de grond en manier waarop de steekproef moet uitgevoerd worden. Hij/zij doet dit op basis van een grondige kennis van de instelling, haar activiteiten en verantwoordelijkheden, de daaruit voortvloeiende archieven en hun onderling belang, alsook op basis van beheersgerelateerde elementen zoals bewaarkost.²

De laatste groep zal zich tot slot concentreren op de inhoud van de archiefreeks en is met andere woorden vooral begaan met de secundaire waarde die de reeks heeft voor het publiek, in het bijzonder voor historici. Volgens hen zou een steekproef genomen vanuit het perspectief van de archiefvormers en hun activiteiten een onvolledig beeld geven van de werkelijkheid, en op die manier ook de benadering van de reeks vanuit toekomstige historische vraagstellingen belemmeren.³ De inbreng van historici over hoe de steekproef uitgevoerd zou moeten worden is wenselijk. De archivaris/records manager moet zich op zijn minst trachten voor te stellen welke mogelijke thema's of onderwerpen toekomstige historici zouden willen onderzoeken.⁴ Vanuit dit standpunt wordt de steekproef in extremis zelfs voorgesteld als een 'fopspeen' of 'schaamlap' voor de archivaris/records manager, een 'tussenoplossing' of ook een vorm van zelfbedrog voor hen die geen bewaar- of vernietigingsbeslissing durven te nemen. Hoe kan men immers een steekproef toepassen op een archiefreeks, zonder de toekomstige historische vraagstellingen te kennen?⁵

1 F.C.J. Ketelaar, "De archivaris en de toevloed van moderne archieven: 22^e internationale conferentie van de Table Ronde des Archives," *Nederlandsch Archievenblad* 3 (1984): 139-140.

2 Hull, *The use of sampling techniques in the retention of records*, 41.

3 Jan van Tol, "Verruiming van selectiecriteria en steekproefreductie van massabestanden," *Archievenblad* 8 (1998): 31.

4 P.J. van Strien, "Rationeel beslissen en historische fantasie," *Nederlandsch Archievenblad* 3 (1991): 90.

5 A.J.M. den Teuling, "Steekproeven, een uitdaging tot onderzoek," *Nederlandsch Archievenblad* 1 (1995): 42.

Op basis van bovenstaand debat, wensen we zelf de volgende kanttekeningen te maken:

1. De steekproef afzweren als selectiemethode is zowel vanuit institutioneel als historisch oogpunt onverantwoord. Dit impliceert immers dat enkel algehele bewaring of vernietiging nog mogelijk zouden zijn. Een dergelijke visie zou bij een algehele bewaring enorme bewaarkosten en intellectuele onoverzichtelijkheid en onbeheersbaarheid met zich meebrengen, iets waar noch de archiefvormer, noch de onderzoeker baat bij hebben. De keuze voor een algehele vernietiging zou voor de onderzoeker dan weer een enorm verlies aan studiemateriaal en inzicht in een bepaald maatschappelijk aspect betekenen. Een dergelijke drastische beslissing zou bovendien elke vorm van controle en transparantie onmogelijk maken, iets wat de archiefvormer aanzienlijke imagoschade zou kunnen berokkenen. Tot slot mag niet vergeten worden dat ook de archiefvormer baat heeft bij een gedeeltelijke bewaring van een bepaalde archiefreeks, als bron van inzicht in activiteiten en verantwoordelijkheden uit het verleden, die op hun beurt bouwstenen van de toekomst kunnen zijn.
2. De praktische uitvoering van een steekproef belemmeren op grond van het niet kennen van de toekomstige historische vraagstellingen is niet alleen een motie van wantrouwen ten overstaan van de archivaris/records manager en zijn/haar kunde en institutionele kennis, maar tevens een ernstige onderschatting van de historicus/historica en zijn/haar bewustzijn van en omgang met de manier waarop historische bronnen tot hem/haar komen. Historici zijn er zich immers niet alleen terdege van bewust dat bronnen in de loop der geschiedenis al dan niet bewust verloren zijn gegaan, ze beseffen bovendien als geen ander dat een aantal aspecten uit het verleden nu eenmaal nooit kenbaar zullen zijn, omdat hun hedendaagse vraagstellingen vreemd zijn aan de belangstelling van vroegere generaties. Het is voor hen ook vanzelfsprekend dat de manier van ordenen, beschrijven en bewaren van bronnen door archiefvormers en archiefinstellingen verre van neutraal is. In dit opzicht wordt een gedeeltelijke vernietiging als antwoord op de ongebreidelde massificatie van archivalia sinds de Tweede Wereldoorlog dan ook gezien als een noodzakelijk kwaad, mits deze gebaseerd is op een verantwoorde keuze.¹
3. Dat er überhaupt een aantal kanttekeningen ter verduidelijking gemaakt moeten worden om het debat in perspectief te plaatsen, benadrukt niet alleen de angst die er in de literatuur aanwezig is voor de steekproef als instrument voor selectie, maar bovenal het gebrek aan degelijke kennis over de methode. In de literatuur kwam meermaals concreet tot uiting dat de auteurs – en in ruimere zin de archivariissen/records managers –

1 Marc Boone, *Historici en hun métier: een inleiding tot de historische kritiek* (Gent: Academia Press, 2007): 43-47.

pseudoargumenten aanvoeren tegen de steekproef om op die manier te ontsnappen aan het gebruik van de methode en de verantwoordelijkheid voor de uitvoering ervan.¹ Toch is alleen de archivaris/records manager in staat om op basis van zijn/haar kennis over de organisatie en zijn/haar vertrouwde met het archiefmateriaal een correcte steekproef uit te voeren. De professional dient aldus de hand in eigen boezem te steken, zijn/haar gebrekkige kennis te (h)erkennen, en bij te spijkeren waar nodig. Angst voor of onwetendheid over de methode mag immers geen excuus zijn om de selectiemethode niet te gebruiken, noch om de term ‘steekproef’ – en zeker ‘statistisch verantwoorde steekproef’ – te misbruiken als dekmantel voor selectiepraktijken en -voorstellen die lijken op een steekproef, maar het niet zijn.²

Dit artikel hoopt dan ook de kennis en inzichten aan te reiken aan wie niet of onvoldoende vertrouwd is met de steekproefmethode, zodat toekomstige (statistische) steekproeven niet alleen correct zullen uitgevoerd worden, maar ook opdat de methode de erkenning en waardering zou krijgen die ze verdient.

Daarom start deze bijdrage met een duidelijke definiëring en kadering van wat een steekproefmethode binnen de statistiek is en welke mogelijkheden er binnen de methode zijn. Vervolgens wordt er in het derde onderdeel dieper ingegaan op de casus waarmee de mogelijkheden en beperkingen van alle ‘statistisch verantwoorde steekproefmethodes’ aangetoond en geïllustreerd zullen worden. Daarna zullen de criteria waaraan een archief moet voldoen om er een steekproef op uit te voeren aan bod komen, gevolgd door een uiteenzetting over hoe de grootte van een steekproef op een statistisch verantwoorde manier kan berekend worden, daarbij rekening houdend met de grootte en de betrouwbaarheid van de steekproef. Een belangrijk aspect binnen dit vijfde onderdeel is het betrouwbaarheidsinterval, waarvan de wiskundige achtergrond en de praktische berekening ervan voor een archief aan bod zullen komen in respectievelijk Appendix 1 en Appendix 2. Na het berekenen van de steekproefgrootte, zal het scala aan aselecte steekproefmethodes besproken en

- 1 Nog los van de aangewende argumentatie binnen het debat, werd in de literatuur ook meermaals de nood aan meer duiding en voorbeelden geëxpliciteerd. Of zoals Felix Hull het treffend samenvatte: “The problem therefore remains of attempting to find some common factors and procedures which can provide archivists with guidelines when attempting to put sampling projects into practice.” Hull, *The use of sampling techniques in the retention of records*, 45.
- 2 Een kritiek die onder meer Terry Cook reeds aankaartte: “Furthermore, as a term, ‘sampling’ is employed ambiguously in archival literature, for too often it is used to refer generically to any decision to retain less than the whole population of a given phenomenon.” Cook, “Many are called, but few are chosen,” 27.

geïllustreerd worden in onderdeel 6, gevolgd door een aantal aandachtspunten bij de praktische uitvoering van een steekproef in onderdeel 7. Tot slot zullen de belangrijkste bevindingen van het artikel gebundeld worden in het besluit, waarna een opsomming van alle geraadpleegde referenties volgt in de bibliografie.

2. Definitie

Vooraleer de archivaris/records manager een steekproef kan uitvoeren, moet hij of zij eerst een idee hebben van wat deze methode inhoudt en – in overeenstemming met de doelstelling van dit artikel – ook hoe deze methode op een statistisch verantwoorde manier kan uitgevoerd worden.

Binnen de statistiek wordt de steekproefmethode omschreven als het wetenschappelijke proces waarbinnen kwalitatieve en kwantitatieve informatie over het geheel gegeven wordt op basis van de analyse van een deel van dit geheel.¹ De informatie verkregen op basis van de analyse van dit deel wordt bijgevolg aanzien als representatief voor het geheel.² Vertaald naar de archivistiek omvat de steekproefmethode het proces waarbinnen kwalitatieve en kwantitatieve informatie over een archiefreeks gegeven kan worden op basis van de analyse van een deel van de archiefreeks, waarbij enkel dit deel in aanmerking komt voor permanente bewaring. Zoals reeds eerder aangegeven, zal de archivaris/records manager deze selectiemethode uitvoeren vanuit zijn/haar intentie om de massa aan archiefmateriaal binnen een bepaalde reeks te reduceren. De uitvoering van deze bewuste selectie vindt met andere woorden plaats op het ogenblik dat het einde van de levensloop van het archiefmateriaal bereikt is, ofwel op het ogenblik dat de intrinsieke waarde voor de instelling die het materiaal gecreëerd en/of ontvangen heeft, volledig verdwenen is.³

Het deel dat geselecteerd werd aan de hand van de steekproefmethode wordt de steekproef genoemd. Om een dergelijke steekproef te verkrijgen, zijn twee methodes mogelijk:

1. Selectie van een steekproef op basis van kansberekening.
2. Selectie van een steekproef op basis van subjectieve beslissingsfactoren.

1 Ranjan Kumar Som, *A manual of sampling techniques* (Londen: Heinemann Educational Books Ltd, 1973): 1.

2 Frank B. Evans, François-J. Himly en Peter Walne, *Dictionary of Archival Terminology* (München: K.G. Saur Verlag KG, 1984): 151.

3 Hull, *The use of sampling techniques in the retention of records*, 9.

1. Een steekproef geselecteerd op basis van kansberekening is een willekeurige of aselechte steekproef (Engels: *random sample*). Hierbij heeft elk archiefstuk een even grote kans om gekozen te worden. Dit maakt dat de variatie aan archiefstukken binnen de steekproef (indien deze voldoende groot is) gelijkwaardig is aan de variatie aan archiefstukken binnen het totale archief. Dit zal toelaten om later – op basis van enkel de steekproef – extrapolaties te maken en zo uitspraken te kunnen doen over het gehele archief.¹

Bovendien laat de statistiek toe een inschatting te maken van de betrouwbaarheid van een aselechte steekproef. Zo kan het betrouwbaarheidsinterval bepaald worden, ofwel de grenzen waarbinnen de foutenmarge op de latere extrapolatie met een zekere waarschijnlijkheid zal liggen. Tot slot is het ook mogelijk om op basis van deze methode een ‘gepersonaliseerde’ steekproef af te nemen, waarbij ofwel de grootte ofwel de betrouwbaarheid van de steekproef zal domineren op, of ondergeschikt zal zijn aan de andere factor, afhankelijk van de eisen die gesteld worden aan de betreffende steekproef.²

Voorbeeld: een archiefreeks bestaande uit 15.000 medische patiëntendossiers wordt aan een willekeurige steekproef onderworpen. De archivaris beslist dat hij/zij de kosten van bewaring van maximaal 1.000 dossiers kan verantwoorden. Per 15 dossiers kiest hij/zij er willekeurig één exemplaar uit voor bewaring, waarop de resterende dossiers vernietigd worden. Op basis van de bekomen steekproef kan achteraf een analyse gemaakt worden van het relatieve voorkomen van verschillende ziektebeelden in de dossiers behorende tot de steekproef, waarbij de frequentie van voorkomen van de verschillende ziekten in de steekproef geëxtrapoleerd kan worden naar het oorspronkelijke (volledige) archief.

2. Een steekproef niet geselecteerd op basis van kansberekening is een niet-willekeurige of selecte steekproef (Engels: *non-random sample*). Binnen deze methode wordt het selectieproces van de steekproef gedomineerd door subjectieve en persoonlijke beslissingsfactoren, wat het onmogelijk maakt de betrouwbaarheid van de betreffende steekproef te meten en bovendien correcte extrapolaties naar het originele archief belemmert.³

1 Een extrapolatie laat toe uitspraken te doen over datgene wat onbekend is, op basis van datgene wat wel gekend is. In het kader van dit artikel zullen de door middel van een steekproef bekomen archiefstukken (met gekende, beschikbare informatie) toelaten uitspraken te doen over de archiefstukken van het gehele archief (die werden vernietigd na het nemen van de steekproef, en waarvoor niet langer informatie beschikbaar is).

2 Som, *A manual of sampling techniques*, 4.

3 Som, *A manual of sampling techniques*, 4-5.

Voorbeeld: in de voornoemde archiefreeks met medische patiëntendossiers worden opnieuw 1.000 dossiers geselecteerd, zij het ditmaal omdat ze handelen over 1) bekende personen, of 2) een medisch uniek of interessant ziektebeeld. De resulterende steekproef zal achteraf geen historisch onderzoek naar de meest voorkomende ziektebeelden uit de desbetreffende periode toelaten, aangezien er sprake is van een oververteenwoordiging van 1) dossiers van bekende personen en van 2) dossiers handelend over zeldzame ziekten. Het hoge aantal en de concentratie aan dergelijke dossiers binnen de steekproef maakt betrouwbare uitspraken over de gehele archiefreeks aangaande de aangetroffen ziektebeelden als zodanig onmogelijk.

Beide methodes komen ook voor binnen de archivistiek en zijn afhankelijk van de intentie van de archivaris/records manager. Deze kan enerzijds beslissen om een onafhankelijke en objectieve selectie te maken binnen een archiefreeks (*random sampling*). Anderzijds is het evenzeer mogelijk een subjectieve keuze te maken binnen een archiefreeks, op basis van de voor de professional meest belangrijk/significant geachte elementen binnen de betreffende reeks (*non-random sampling*).¹ Daarnaast is het perfect mogelijk beide methodes te combineren binnen dezelfde archiefreeks. Voorwaarde daarbij is dat beide steekproeven afzonderlijk van elkaar genomen en gedocumenteerd worden en dat ze intellectueel gescheiden blijven om de statistische validiteit van de via kansberekening verkregen steekproef niet onderuit te halen.²

Ondanks de mogelijkheden die beide steekproefmethodes kunnen bieden aan de archivaris/records manager, zal hierna enkel de via kansberekening verkregen steekproef besproken worden, door ons omschreven als de ‘statistisch verantwoorde steekproef’.

3. Casus

Om de mogelijkheden en beperkingen aan te tonen van elke methode die onder de noemer ‘statistisch verantwoorde steekproef’ valt, zal gebruik gemaakt worden van een casus.

De casus in kwestie omvat alle verkoopfacturen opgemaakt in de periode 1737-1799 door de textielhandel Bethune & Fils, die gevestigd was te Kortrijk (West-Vlaanderen, België). De 15.143 facturen in kwestie werden samen met de rest van het handelsarchief van de familie Bethune uit de tweede helft van de 18^{de} eeuw in detail onderzocht door Annik Adriaenssens, in het kader van haar doctoraat.

1 Hull, *The use of sampling techniques in the retention of records*, 9-10.

2 Cook, “Many are called, but few are chosen,” 43.

Dit doctoraat – een analyse van de textielhandel Bethune & Fils tijdens de tweede helft van de 18^{de} eeuw – stelt ons in staat een grondige analyse van de reeks en van het archief in het algemeen te maken (bij gebrek aan een eigen fysiek of elektronisch archief waarop een hypothetische steekproef zou uitgevoerd kunnen worden), als voorbereiding op de eigenlijke steekproef.

De uitgekozen reeks verkoopfacturen maakt deel uit van een groter geheel aan privaat handelsarchief, eigendom van en bewaard door de familie Bethune op hun kasteel te Marke (West-Vlaanderen, België). Het geheel aan bewaarde archiefstukken vertelt het verhaal van het voormalige en naar de familie genoemde textielbedrijf, waarvan de oorsprong terug te brengen is tot de tweede helft van de 17^{de} eeuw. Aanvankelijk gevestigd door Joachim Van Dale in centrum Kortrijk (West-Vlaanderen, België), concentreerde het bedrijf zich op de verkoop van allerhande stoffen aan kleine handelaars en particulieren. Vanaf de overname op 1 oktober 1753 door Jean-Baptiste Bethune – getrouwd met een latere telg van de Van Dale familie – zou het bedrijf verder evolueren tot een gespecialiseerde linnenhandel. Samen met zijn zoon zette Bethune de zaak voort onder de naam ‘Bethune & Fils’ (vanaf oktober 1780), een naam die behouden zou blijven tot het bedrijf in 1850 werd opgedoekt door een nazaat van de familie.

Om de gebruikte metadata van de casus bij de illustraties van de steekproefmethodes verderop in het artikel beter te begrijpen, evenals hun onderlinge verdeling en betekenis binnen de steekproefmethodes, vereist de beslissing van Jean-Baptiste Bethune van 26 maart 1768 om tot een specialisatie in linnen over te gaan de nodige uitleg.¹ Zijn opleiding tot linnenspecialist zorgde er in eerste instantie voor dat hij een bijzondere affiniteit met en kennis over linnen bezat. Een dergelijke specialisatie bleek in zijn tijd geen overbodige luxe, aangezien het aanbod aan ‘wollegoederen’ – het geheel aan wollen en semi-wollen stoffen – die zijn voorganger verkocht in toenemende mate uitbreidde.² Dit was het resultaat van een samenleving waarin consumptie en mode meer en meer op de voorgrond traden en waarin niet alleen het aanbod, maar ook het aantal verkopers van dergelijke stoffen aanzienlijk steeg. De steeds toenemende onderlinge concurrentie resulteerde op zijn beurt in dalende winstmarges zodat de handel in deze wollegoederen steeds minder interessant werd. Bovendien was Bethune bij het kiezen van de perfecte stoffen binnen het verder uitbreidende aanbod vaak aangewezen op de expertise van tussenpersonen en leveranciers. Deze kennis en afhankelijkheid van derden kwamen met een

1 Annik Adriaenssens, “Van laken tot linnen: de textielhandel Bethune & Fils, tweede helft achttiende eeuw: een analyse op basis van het bedrijfsarchief” (PhD diss., Universiteit Gent, 2016): 2: 1199-1200.

2 Adriaenssens, “Van laken tot linnen,” 1: 225.

prij斯卡artje en maakten bovendien een snelle en adequate opvolging van de aankoop van deze stoffen haast onmogelijk. Toen de nodige winst – ondanks de eis tot scherpere prijzen bij zijn leveranciers – steeds meer uitbleef, besliste hij uiteindelijk de wolgoederen uit het aanbod te schrappen.¹ Alleen over de destijds florerende linnenhandel bezat hij zelf de nodige expertise en deze was winstgevend genoeg om verder te zetten, waardoor zijn bedrijf zich vanaf 1768 uitsluitend zou specialiseren in lijnwaad.²

Bethune legde zich in hoofdzaak toe op de verkoop van huishoudlinnen, waarmee onder andere hemden en zakdoeken werden vervaardigd.³ Uitzonderlijk verkocht hij ook tafellinnen en lichtere linnenstoffen (*rollez*), zij het uitsluitend op bestelling.⁴ Binnen het eerder beperkte spectrum aan huishoudlinnen vervaardigd uit zuiver linnen werd het onderscheid hoofdzakelijk gemaakt aan de hand van de breedte van het stuk, en aan de hand van de geslotenheid van de weefstructuur. Andere onderscheidende factoren waren de kwaliteit en de fijnheid van het vlagaren, die de klasse en bijgevolg ook de prijs van het lijnwaad bepaalden.⁵

Binnen de vooropgestelde casus zullen echter twee andere types metadata een doorslaggevende rol spelen:

1. De lengte van het lijnwaad: doordat de breedte van het linnen vaak niet vermeld werd in de factuur, kon in het doctoraat op basis van een vergelijking van de lengte van de stukken alsnog een onderscheid gemaakt worden tussen de verkochte linnensoorten.⁶
2. De kleur van het lijnwaad: afhankelijk van de wensen van de klant, werd het linnen in zijn natuurlijke kleur/onbewerkt (*écru*), geverfd of gebleekt verkocht. De goedkoopste, minderwaardige linnenstoffen werden doorgaans geverfd, het betere lijnwaad (inclusief het courante huishoudlinnen) daarentegen gebleekt.⁷ Twee bleekmethodes werden regelmatig toegepast op het door Bethune & Fils verkochte linnen. In eerste instantie werd er gebleekt met aslogen, wat ‘menagebleek’, ‘halve bleek’ of ‘waterbleek’ genoemd werd. Vroeg de klant echter om zuiver wit linnen, dan werd het linnen na het bleken met assen verder afgebleekt in zuurbaden

1 Adriaenssens, “Van laken tot linnen,” 1: 578-580.

2 Adriaenssens, “Van laken tot linnen,” 1: 583.

3 Adriaenssens, “Van laken tot linnen,” 2: 594.

4 Adriaenssens, “Van laken tot linnen,” 2: 596.

5 Adriaenssens, “Van laken tot linnen,” 2: 600.

6 Adriaenssens, “Van laken tot linnen,” 2: 632-633.

7 Adriaenssens, “Van laken tot linnen,” 2: 642-645.

van karnemelk. Deze agressievere bleekmethode stond bekend onder de naam ‘melkbleek’.¹

4. Voorwaarden/criteria

Na een grondige analyse van het voor steekproef in aanmerking komende archief, is het belangrijk om na te gaan of een wetenschappelijk verantwoorde steekproef überhaupt kan toegepast worden op de betreffende documenten. In het geval van onze casus: kan een dergelijke steekproef wel toegepast worden op de reeks verkoopfacturen, opgemaakt tussen 1737 en 1799?

Om op deze vraag een antwoord te kunnen bieden en dus een correcte inschatting van de selectiemogelijkheden op het archief te kunnen maken, moeten de volgende twee criteria/vragen in acht genomen worden:

1. Omvat het archief een homogene en afgesloten groep archiefbescheiden?
2. Is er minstens één betrouwbare *random variable* in de reeks aanwezig?

1. Omvat het archief een homogene en afgesloten groep archiefbescheiden?
Een steekproef heeft tot doel representatieve informatie over het geheel (het archief) te geven op basis van een deel van dit geheel (de steekproef). Dit betekent dat de informatie verkregen uit de steekproef representatief moet zijn voor het geheel, of nog: de informatie uit het door middel van een steekproef geselecteerde archief moet geëxtrapoleerd kunnen worden naar het gehele archief. Om een dergelijke extrapolatie van informatie mogelijk te kunnen maken, dienen de archiefbescheiden van het betreffende archief een afgesloten geheel te vormen én gelijkaardige informatie/metadatas te bevatten. Zeker in het geval van de wetenschappelijk verantwoorde steekproef is het belangrijk over een afgesloten groep van uniforme metadatas te beschikken, aangezien deze aan wiskundige berekeningen onderworpen zullen worden. Omdat de inherente kenmerken van een archiefreeks hieraan het best voldoen, is het toepassen van een wetenschappelijk verantwoorde steekproef vooral aangeraden voor dit type archief.² Dit betekent echter niet dat op andere archieven geen selectie kan toegepast worden, zij het dan meestal eerder op subjectieve gronden.

1 Adriaenssens, “Van laken tot linnen,” 2: 654-656.

2 Onder het begrip ‘archiefreeks’ of ‘archieiserie’ wordt een ‘archiebestanddeel bestaande uit formeel gelijksoortige archiefbescheiden’ begrepen. Deze gelijksoortigheid vertaalt zich in de functie, het ontwikkelingsstadium of de redactionele vorm van de archiefbescheiden. A.J.M. den Teuling, *Archiefterminologie voor Nederland en Vlaanderen* (s-Gravenhage: Stichting Archiefpublicaties, 2007): 107.

Casus: de gekozen reeks bevat een afgesloten geheel van eenzelfde type archiefbescheiden met gelijkaardige metadata – verkoopfacturen opgemaakt in de periode 1737-1799 door de textielhandel Bethune & Fils – en dit maakt het mogelijk om wiskundige berekeningen uit te voeren op het gegevensbestand van de reeks. Op basis van deze berekeningen zal het mogelijk zijn een steekproef te selecteren, waarvan de metadata representatief zullen zijn voor het geheel van de reeks.

2. Is er minstens één betrouwbare *random variable* in de reeks aanwezig?

Een *random variable*¹ is een type metadata, waarvoor elk archiefstuk een numerieke waarde heeft en waarop een wiskundige analyse kan toegepast worden. Op basis van deze analyse kan de interne variatie binnen dit type metadata berekend worden aan de hand van het gemiddelde en de standaardafwijking. De interne variatie zal het op haar beurt mogelijk maken om de grootte en de kwaliteit van de uiteindelijke steekproef te bepalen. Correcte uitspraken op basis van deze interne variatie zijn echter alleen mogelijk wanneer er een betrouwbare *random variable* aanwezig is, een type metadata waarvan de waarden uniform, onderling vergelijkbaar en aldus extrapoleerbaar moeten kunnen zijn naar de rest van het archief. Indien het uitgekozen type metadata niet voldoet aan deze vereisten, en bijgevolg niet betrouwbaar is, dan zal deze extra variatie bevatten bovenop de berekende interne variatie. Deze extra variatie zal betrouwbare uitspraken op basis van de interne variatie bemoeilijken, en zelfs onmogelijk maken.

Casus: binnen het gegevensbestand gerelateerd aan de reeks verkoopfacturen werden meerdere *random variables* aangetroffen. Een van deze variabelen is de lengte van een stuk linnen. Zoals eerder aangegeven, bevatten alle facturen een lengtemaat die toelaat de breedte van de stukken linnen (een type metadata waarvoor de waarden geregeld ontbraken in de verkoopfacturen) te berekenen. Elk archiefstuk bevat een of meerdere numerieke waarden voor dit type metadata, wat het mogelijk maakt de interne variatie van de lengtematen te berekenen. Vooraleer de interne variatie kan berekend worden, moeten de lengtematen echter uniform en vergelijkbaar zijn om te voldoen aan de vereisten van een betrouwbare *random variable*. In het gegevensbestand gerelateerd aan de reeks – zoals opgemaakt door Annik Adriaenssens – werden de lengtes vermeld in elke factuur alvast opgesplitst per linnensoort, waarbij het onderscheid gemaakt werd door middel van een combinatie van de breedte en de kleur van het linnen (voorbeeld: menagegebleekt linnen 2/3 versus menagegebleekt linnen 3/4). Omdat de lengte van een stuk linnen binnen de verkoopfacturen echter zowel in Franse als in Vlaamse el werd uitgedrukt, werden de lengtematen uit het gegevensbestand eerst omgerekend. Aangezien de Vlaamse el het vaakst als lengtemaat gebruikt werd in de originele stukken, dienden enkel de waarden in Franse el

1 Som, *A manual of sampling techniques*, 5.

omgerekend te worden. Indien er meerdere stukken van eenzelfde linnensoort werden aangekocht (voorbeeld: 11 stukken melkgebleekt linnen 3/4), dan werden in het gegevensbestand van de reeks enkel de minimale en maximale lengtes van alle stukken opgegeven (voorbeeld: 59 tot 63 Vlaamse el). Om de lengtematen in het gegevensbestand zo uniform als mogelijk met elkaar te kunnen vergelijken, werd voor elke factuur het gemiddelde van beide waarden aanzien als de lengte van alle stukken (voorbeeld: 61 Vlaamse el). Wanneer de omrekeningen uitgevoerd en de gemiddeldes vastgesteld zijn, wordt een onderlinge vergelijking tussen de lengtematen mogelijk en zal de interne variatie berekend kunnen worden. Ondanks de invloed van de omrekeningen en de genomen gemiddelde lengtematen als extra variatie, zullen op basis van de bekomen interne variatie alsnog betrouwbare uitspraken gedaan kunnen worden over de grootte en de kwaliteit van de steekproef.

Een andere mogelijke *random variable* binnen ons gegevensbestand is de prijs van het verkochte linnen. Die zou in vergelijking met de lengtemaat een minder betrouwbare uitkomst geven, maar is gezien het type archiefbescheiden (verkoopfacturen) waarschijnlijk meer voor de hand liggend. Net als in het geval van de lengtematen, zijn er voor elke verkoopfactuur in het databestand een of meerdere numerieke waarden voor dit type metadata. Voor elke verkoopfactuur is er, zoals aangegeven, een opsplitsing per linnensoort, waardoor een prijswaarde per linnensoort onderscheiden kan worden. Om de verschillende waarden met elkaar te kunnen vergelijken, zou echter een omrekening van alle bedragen naar één muntstelsel doorgevoerd moeten worden. Het linnengoed werd immers in twee verschillende eenheden aangerekend (Vlaamse ponden en Brabantse gulden). Bovendien zou ook de gedurende de periode variërende wisselkoers bij deze omrekening in rekening gebracht dienen te worden, net als de invloed van inflatie en deflatie op de prijzen van het lijnwaad, zoals aangerekend op de verkoopfacturen uit de periode 1737-1799. Elk van de voornoemde ingrepen is niet alleen erg moeilijk te becijferen, maar zou er bovendien voor zorgen dat de impact van de extra variatie zodanig groot wordt, dat uitspraken op basis van deze *random variable* erg moeilijk, zo niet onmogelijk zouden zijn.

Hoewel de impact van de extra variatie op elke potentiële *random variable* niet onderling met elkaar vergeleken kan worden, moge het duidelijk zijn dat het bestaan van meerdere potentiële variabelen niet noodzakelijk betekent dat elk van deze variabelen even bruikbaar/betrouwbaar is. Een ideale *random variable* bestaat niet, maar bovenstaande voorbeelden hebben toch aangetoond dat een grondige analyse van de types metadata aanwezig in de archiefbescheiden de archivaris/records manager in staat kunnen stellen een correcte inschatting van de minder of meer betrouwbare opties te maken. Wanneer een relatief betrouwbare variabele werd aangeduid, zou deze relatief betrouwbare uitspraken over de kwaliteit en de grootte van de steekproef mogelijk moeten kunnen maken.

5. Betrouwbaarheid van de steekproef

De uitvoerder van een statistisch verantwoorde steekproef dient twee essentiële vragen te beantwoorden:

1. Hoe groot dient een steekproef te zijn?
2. Hoe groot dient een archief te zijn om een statistische steekproef mogelijk te maken/het nemen van een nuttige statistische steekproef toe te laten?

Idealiter vormt de steekproef een zo nauwkeurig mogelijke benadering van het gehele archief, waarbij een zo groot mogelijke steekproef vaak als de beste garantie hiervoor beschouwd wordt (en in extremis, de bewaring van het gehele archief). Praktische en financiële bezwaren leggen echter meestal een bovengrens op aan de te bewaren hoeveelheid archiefbescheiden. De archivaris/records manager zal bij het bepalen van de steekproefgrootte dus een afweging moeten maken tussen de betrouwbaarheid van de steekproef enerzijds, en de omvang anderzijds.

De literatuur biedt echter geen duidelijkheid omtrent de kwantitatieve relatie tussen de steekproefgrootte en de betrouwbaarheid ervan. Cook stelt dat de betrouwbaarheid afhangt van het absolute aantal stukken in de steekproef (voorbeeld: 1.000 stukken), eerder dan van de relatieve grootte (voorbeeld: 1% van het totale archief).¹ Voor grote archieven klopt deze stelling: vanaf een bepaalde steekproefgrootte is het waarschijnlijk dat van alle verschillende types archiefstukken voldoende voorbeelden in de steekproef zijn opgenomen. Op dat moment maakt het dus niet meer uit of het totale archief bijvoorbeeld één miljoen of tien miljoen stukken bevat, en of de steekproefverhouding 0,1% dan wel 0,01% bedraagt. Voor kleinere archieven raadt Cook de tabel opgesteld door Bell Telephone aan, die voor een aantal archiefgroottes de gewenste steekproefgrootte weergeeft.² Opnieuw is er echter geen kwantitatieve bepaling van de steekproefkwaliteit; de tabel biedt enkel de optie om voor een *low*, *average* of *high* steekproefgrootte te kiezen. Problematischer is echter dat deze tabel niet opgesteld werd om representatieve steekproeven mogelijk te maken, maar wel voor een heel ander specifiek probleem, namelijk het zoeken naar defecte of minderwaardige onderdelen bij de levering van een

1 Cook, "Many are called, but few are chosen," 36.

2 "The Bell Telephone MIL-STD 105D Sampling Plan", gereproduceerd in Cook, "Many are called, but few are chosen," 46.

groot aantal producten.¹ De tabel geeft aan hoeveel onderdelen getest moeten worden om met voldoende zekerheid te kunnen zeggen dat alle onderdelen van de levering het gewenste kwaliteitsniveau halen. De statistiek die gebruikt werd om deze tabel op te stellen gaat dus uit van een zeer klein aantal afwijkende elementen (de defecte producten), in een verder uiterst homogene groep (de producten van aanvaardbare kwaliteit). Wanneer we deze statistische formules zomaar toepassen op een heel andere situatie, namelijk die van archiefstukken met uiteenlopende eigenschappen waarop we een steekproef met representatieve verdeling hopen toe te passen, is hun geldigheid niet langer gegarandeerd. We begrijpen dat Cook zich, bij gebrek aan rekenkracht om zelf de nodige statistische analyses uit te voeren, destijds baseerde op bestaande (al dan niet van toepassing zijnde) tabellen, maar voor een hedendaagse archivaris die beschikt over een overdaad aan rekenkracht² en standaard softwarepakketten zoals Microsoft Excel mag dit hoegenaamd geen excuus meer zijn.

5.1 Bepalen van het betrouwbaarheidsinterval

In dit artikel stellen wij voor om de betrouwbaarheid van de steekproef wiskundig te bepalen door middel van het betrouwbaarheidsinterval. Hiervoor maken we gebruik van de gekozen metadata (*random variable*) en een afgeleide eigenschap (meestal het gemiddelde) die we kunnen uitrekenen voor de steekproef en voor het gehele archief. De wiskundige achtergrond en formules die toelaten om het betrouwbaarheidsinterval te berekenen, hebben we opgenomen in Appendix 1. In deze tekst beperken we ons tot een beschrijving van het betrouwbaarheidsinterval en hoe het te gebruiken. Appendix 2 bevat instructies die de archivaris/records manager zullen toelaten om, met een computerprogramma zoals Microsoft Excel, betrouwbaarheidsintervallen te berekenen voor zijn/haar eigen archief.

- 1 De MIL-STD 105 standaard werd aanvankelijk in het leven geroepen voor gebruik bij militaire aanbestedingen en is een vorm van aanvaardingsbemonstering. Hierbij wordt een levering van goederen onderworpen aan een steekproef en worden enkel de producten in de steekproef uitvoerig getest. Wanneer deze allemaal van voldoende kwaliteit bevonden worden, gaat men ervan uit dat de gehele levering geen (of een aanvaardbaar klein aantal) defecte onderdelen bevat en wordt de betreffende levering aanvaard. De grootte van de steekproef wordt bepaald door de tabellen opgegeven in de MIL-STD 105 standaard en hangt af van de grootte van de levering (totaal aantal producten) enerzijds en van de gewenste kans om toch een defect onderdeel toe te laten (uitgedrukt in *low*, *average* en *high* klassen).
- 2 De gemiddelde laptop van vandaag heeft immers heel wat meer rekenkracht dan de supercomputer uit 1963 die nodig was om de tabellen van de MIL-STD 105D standaard op te stellen.

Het betrouwbaarheidsinterval voor een gegeven steekproef zal aangeven tussen welk interval (minimum- en maximumwaarden) het gemiddelde van onze gekozen metadata zal liggen voor het gehele archief. Dit interval drukt met andere woorden de kwaliteit uit van de extrapolatie, die een latere onderzoeker zal kunnen maken op basis van de steekproef.

Voor onze casus weten we bijvoorbeeld dat de gemiddelde lengte van de verkochte linnenstukken 74 Vlaamse el bedraagt. Wanneer we een steekproef uitvoeren op 1.000 willekeurig gekozen facturen, bekommen we een betrouwbaarheidsinterval dat stelt dat “de gemiddelde lengte, met 95% zekerheid, tussen 73 en 75 el valt”.

In de bovenstaande uitspraak zijn drie elementen van belang:

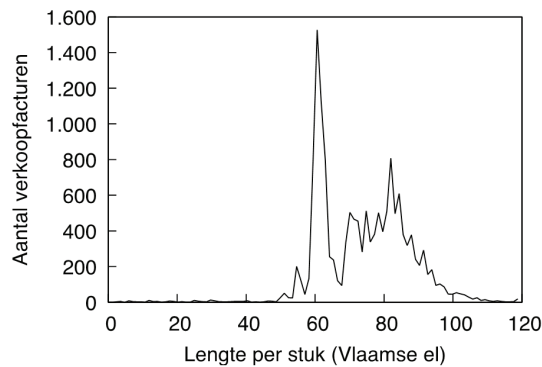
1. De grootte van de steekproef, in dit geval 1.000 verkoopfacturen.
2. Het interval zelf, hier 73-75 el. Meestal wordt de grootte van dit interval bekeken in verhouding tot de centrumwaarde, hier dus 1 el boven of onder het gemiddelde van 74 el, wat een afwijking van 1,35% inhoudt. Dit wil zeggen dat we, op basis van de steekproef, de gemiddelde lengte van de verkochte linnenstukken in het geheel aan verkoopfacturen kunnen voorspellen met een fout van hoogstens 1,35%.
3. De waarschijnlijkheid, hier 95%, ofwel de kans dat de uitspraak correct is, en dat de gemiddelde lengte van de linnenstukken vermeld in het geheel aan verkoopfacturen effectief binnen de opgegeven grens valt. Meestal wordt door statistici een waarde van 95% aangenomen als een aanvaardbare kans. Hogere waarden, zoals bijvoorbeeld 99%, komen voornamelijk voor in toepassingen waarbij grote afwijkingen gevaarlijke situaties met zich mee kunnen brengen, zoals het berekenen van de sterkte van bouwkundige constructies. Dergelijke hoge waarschijnlijkheidspercentages zijn binnen de archivaliek over het algemeen echter niet nodig.

Dit betrouwbaarheidsinterval kan de archivaris/records manager een richtlijn bieden om een afweging te kunnen maken over de kwaliteit van de steekproef, op basis van objectieve criteria. De grootte van de steekproef heeft hij/zij zelf in handen en kan desgewenst aangepast worden totdat het betrouwbaarheidsinterval binnen voor hem/haar aanvaardbare grenzen valt.

Een belangrijke opmerking over de berekening van het betrouwbaarheidsinterval is dat de grootte van het interval niet enkel afhangt van de grootte van de steekproef, maar ook van de interne variatie tussen de verschillende archiefbescheiden (de standaardafwijking σ in Appendix 1): hoe groter deze interne variatie, hoe groter het betrouwbaarheidsinterval en dus ook hoe onzekerder de kwaliteit van de extrapolatie zal zijn, die mogelijk zou moeten zijn

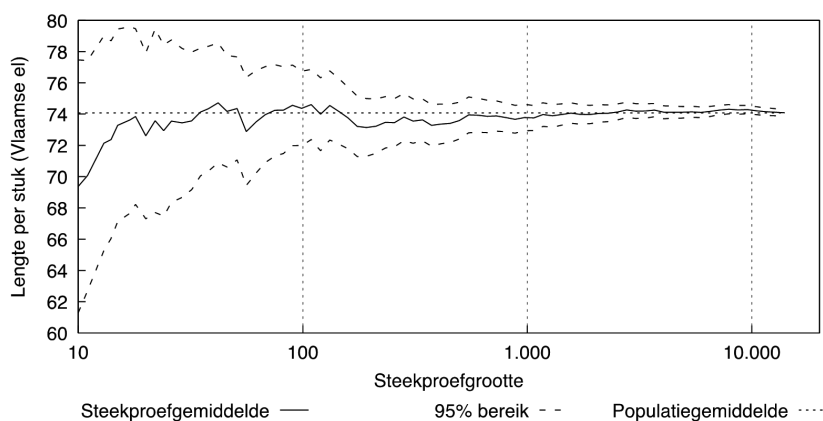
op basis van de steekproef. Hieruit volgt dat, voor een gelijke steekproefgrootte, de kwaliteit van de steekproef slechter zal zijn indien het archief een grote interne variatie vertoont. Een grotere steekproef zal in dat geval wenselijk zijn om de kwaliteit van de betreffende steekproef op het gewenste niveau te brengen.

Gewapend met de techniek van het betrouwbaarheidsinterval, kunnen we nu het effect van verschillende steekproefgroottes illustreren aan de hand van onze casus, de linnenverkopen van textielhandel Bethune & Fils tussen 1737 en 1799. In totaal werden 15.143 verkoopfacturen geanalyseerd. Per factuur weerhielden we het type stof en de lengte van het verkochte stuk. Binnen het geheel aan verkoopfacturen zagen we de verdeling weergegeven in Figuur 1.



Figuur 1 Verdeling van de stuklengte

Vervolgens simuleerden we een aantal mogelijke steekproeven en bepaalden we hun verwachte nauwkeurigheid aan de hand van hun betrouwbaarheidsinterval. In de praktijk werden dus steeds een aantal willekeurig geselecteerde facturen gekozen, waarop het betrouwbaarheidsinterval (95%) werd berekend en uitgezet op een grafiek (Figuur 2). De grootte van de steekproef lieten we starten op tien facturen, en verder toenemen tot ze uiteindelijk de gehele populatie omvatte. De horizontale stippellijn op de grafiek toont de gemiddelde lengte over het gehele archief, 74 el. De curve aangeduid met een volle lijn toont het steekproefgemiddelde bij een gegeven steekproefgrootte. De curves aangeduid met stippellijnen bakenen tot slot het betrouwbaarheidsinterval af.



Figuur 2 Betrouwbaarheidsinterval bij verschillende steekproefgroottes

Uit de grafiek leiden we af dat voor een kleine steekproef, bijvoorbeeld minder dan 100 facturen, het steekproefgemiddelde nog sterk varieert tussen de verschillende steekproeven. De variatie tussen de facturen is immers zo dat, bij een klein aantal geselecteerde bescheiden, de kans reëel is dat de gegevens van een extra geselecteerd archiefstuk sterk zullen afwijken van de gegevens van de reeds gekozen stukken. Het betrouwbaarheidsinterval bevestigt deze bezorgdheid in wiskundige zin. We kunnen dus stellen dat de steekproef waarschijnlijk niet representatief zal zijn voor het geheel aan archiefbescheiden. Bij een steekproefgrootte van 500 verkoopfacturen is het betrouwbaarheidsinterval echter sterk gekrompen, tot ± 1 el. Dit wil zeggen dat bij een steekproef van 500 facturen we (met een zekerheid van 95%) kunnen stellen dat het steekproefgemiddelde met ten hoogste 1 el zal afwijken van het gemiddelde van de totale populatie.

Voorgaande grafiek werd opgemaakt met kennis van het volledige archief. Het was met andere woorden mogelijk het reële gemiddelde te kennen, alsook om verschillende (ook zeer grote) steekproefverhoudingen uit te testen, iets wat in de praktijk niet altijd mogelijk zal zijn. Bedenk echter dat het betrouwbaarheidsinterval *uitsluitend* berekend wordt op basis van de gegevens uit de genomen steekproef, niet op basis van het totale archief. Dit wil zeggen dat het niet noodzakelijk is metadata te verzamelen voor alle archiefstukken, maar enkel voor de in de steekproef opgenomen stukken. Afhankelijk van de gekozen metadata kan dit een aanzienlijke tijds winst opleveren. Bovendien kan men beginnen met een relatief kleine steekproef, hierop de metadata bepalen en het betrouwbaarheidsinterval uitrekenen, en vervolgens beslissen of de bekomen

kwaliteit al dan niet voldoende is. Een dergelijke werkwijze laat een incrementele aanpak toe, waarbij de steekproef progressief vergroot wordt totdat de gewenste kwaliteit bekomen is (of totdat het te bewaren volume haar praktisch toelaatbare maximale grootte bereikt heeft).

6. Steekproefmethodes en toepassing

Nu blijkt dat het archief wel degelijk geschikt is om een statistische steekproefmethode op toe te passen en we bovendien weten hoe we kunnen nagaan of een steekproef groot genoeg is om betrouwbaar te zijn, kunnen we overgaan tot het effectief uitvoeren van de steekproef. Indien we een statistisch verantwoorde steekproef wensen uit te voeren, wat zijn dan de mogelijkheden en hoe dienen deze uitgevoerd te worden?

In dit artikel behandelen we enkel de klasse van aselechte steekproeven. Zoals reeds aangegeven in onderdeel 2. Definitie, zijn dit steekproefmethodes waarbij het al dan niet selecteren van een bepaald archiefstuk puur op toeval gebaseerd is. Alternatieve types steekproeven zijn de selecte methodes, waarbij de steekproef niet puur volgens toeval wordt samengesteld. Selecte steekproefmethodes kunnen niet uitsluiten dat de steekproef een over- of ondervertegenwoordiging bevat van bepaalde afwijkende types documenten en zijn dan ook niet geschikt om een extrapolatie naar het geheel archief te doen. Historisch onderzoek op basis van dergelijke selecte steekproeven is dan ook haast onmogelijk.

Tot de groep aselechte steekproeven behoren zowel de enkelvoudige steekproefmethodes als de gestratificeerde steekproef, die beiden in dit onderdeel aan bod zullen komen.

6.1 Enkelvoudige steekproefmethodes

Binnen de enkelvoudige steekproefmethodes bespreken we drie mogelijkheden: de enkelvoudige aselechte steekproef, de enkelvoudige systematische steekproef, en de semi-systematische steekproef.

In de tekst die volgt, gebruiken we volgende symbolen:

n	Steekproefgrootte
N	Grootte van de gehele populatie

p	Steekproefverhouding n/N ; de kans van een stuk om deel uit te maken van de steekproef
k	'Skip' $1/p$; bij een steekproefverhouding van $p = 1/k$ zal voor elke k stukken er één gekozen worden

6.1.1 Enkelvoudige aselechte steekproef

Het meest eenvoudige type steekproef is de enkelvoudige aselechte steekproef. Bij deze steekproefmethode worden alle potentiële archiefstukken gelijk behandeld (er is dus slechts één klasse van archiefstukken, 'enkelvoudig'), en heeft elk archiefstuk een gelijke kans om gekozen te worden ('aselect'). Bij een homogeen archief is dit type steekproef het meest geschikt en biedt het ook de grootste kans op representativiteit.

Eens we ervoor gekozen hebben om een aselechte steekproef uit te voeren met een vooropgestelde steekproefverhouding, stelt zich de vraag hoe we deze in de praktijk zullen uitvoeren. Laat ons uitgaan van een steekproefverhouding p , berekend als de gewenste steekproefgrootte n gedeeld door de grootte van het totale archief N (de populatiegrootte). De verhouding $p = n/N$ is een getal tussen 0 en 1, wat de kans (*probability*) uitdrukt dat een gegeven stuk gekozen wordt. Zo zal bij een totaal archief van 100 archiefstukken en een steekproefgrootte van 5 archiefstukken, de kans p gelijk zijn aan $5/100 = 0,05$, ofwel 5%.

Tijdens de uitvoering van een aselechte steekproef is het belangrijk toevallige selectie voldoende kans te geven en dus geen mechanisme te gebruiken dat enige vorm van koppeling tussen archiefstukken ('correlatie') mogelijk maakt. De kans p dat elk archiefstuk heeft om gekozen te worden, moet met andere woorden volledig onafhankelijk van alle andere archiefstukken staan. Dit kan door aan elk archiefstuk een willekeurig getal tussen 0 en 1 toe te kennen en door het archiefstuk te selecteren als het toegekende getal kleiner is dan p . Een andere manier om toevallige selectie toe te laten, bestaat erin de archiefstukken te sorteren volgens een willekeurige volgorde¹ en de eerste n stukken te kiezen (waarbij n de grootte van de steekproef is). Deze laatste methode heeft bovendien als voordeel dat het een eenvoudige toepassing van de incrementele methode toelaat, door eerst de eerste n stukken te kiezen, vervolgens het betrouwbaarheidsinterval te berekenen en indien nodig extra stukken van de reeds bestaande en gesorteerde lijst toe te voegen.

1 Deze sortering is puur intellectueel (bijvoorbeeld in een Excellijst waarin de archiefstukken worden opgelijst); er dient dus niet overgegaan te worden tot een fysieke herordening van het archief.

Voorgaande techniek vereist dat aan elk archiefstuk een willekeurig getal toegerekend kan worden. Tegenwoordig is dit met computerprogramma's zoals Microsoft Excel eenvoudig uit te voeren (zie Appendix 2 voor een uitgewerkt voorbeeld). In het verleden was het echter moeilijk grote hoeveelheden willekeurig geselecteerde getallen te genereren. Als oplossing maakte men gebruik van lijsten met willekeurige getallen, met het gevaar dat dezelfde lijst meerdere malen gebruikt werd, wat een bepaalde systematiek in de selectie bracht.

6.1.2 *Enkelvoudige systematische aselechte steekproef*

Een populaire selectiemethode die dikwijls in de literatuur beschreven wordt, is de systematische steekproef.¹ In tegenstelling tot de aselechte steekproef, waarbij voor elk archiefstuk individueel beslist wordt – op basis van puur toeval – om het al dan niet toe te voegen, gebruikt de systematische steekproef een bepaalde systematiek om de stukken te selecteren. De meest gebruikte methode zal, bij een steekproefverhouding 1 op k , elk k -de stuk kiezen vertrekkend van een willekeurige startpositie. Voor een steekproefverhouding van bijvoorbeeld 1 op 10 en een willekeurige startpositie 3, zullen de archiefstukken met rang 3, 13, 23, etc. geselecteerd worden.

Het grote gevaar bij deze methode schuilt er echter in dat voor documenten die enige systematiek vertonen, de mogelijkheid bestaat dat de steekproef niet langer representatief is. Bij een archief dat bestaat uit boekhoudkundige dagoverzichten bijvoorbeeld, met één archiefstuk per weekdag, zou de selectie van stukken 3, 13, 23, etc. enkel stukken in de steekproef opnemen die opgemaakt zijn op een woensdag – waardoor niet langer sprake is van een representatieve steekproef. Dit fenomeen wordt in de statistiek *aliasing* genoemd. Merk op dat dit fenomeen vooral gevaarlijk is wanneer de systematiek in het archief niet meteen duidelijk is. Wanneer men weet dat opeenvolgende archiefstukken betrekking hebben op opeenvolgende wekdagen, kan men uiteraard vermijden een steekproefverhouding te gebruiken die een veelvoud is van 5. Maar is de archivaris/records manager er zeker van dat er geen ander systematisch gedrag is, waarvan hij/zij zich niet bewust is?

6.1.3 *Semi-systematische steekproef*

Het nadeel van de puur aselechte methode bestaat er volgens Cook in dat er grote 'gaten' in de selectie kunnen voorkomen (door hem *missing pockets*

1 Cook, "Many are called, but few are chosen," 38.

genoemd).¹ Aangezien de selectie puur willekeurig verloopt, is het statistisch gezien inderdaad niet onmogelijk dat voornamelijk stukken uit het begin en het einde van de reeks opgenomen worden en bijgevolg haast niets uit het midden. Systematische selectie heeft dit probleem uiteraard niet, aangezien het aantal opeenvolgende niet-geselecteerde stukken altijd gelijk is aan $k - 1$.

Het moge echter duidelijk zijn dat – ten gevolge van de *aliasing* problematiek – de puur systematische selectie geen betrouwbare oplossing kan zijn. Het grote voordeel van systematische selectie, namelijk de eenvoudige uitvoering zonder het noodzakelijk gebruik van lange lijsten met willekeurige getallen, is dankzij de moderne rekenkracht niet langer van doorslaggevend belang. En ook het zogenaamde ‘grote gaten’ probleem in de geselecteerde stukken valt in de praktijk wel mee. Enerzijds gaan we al uit van de veronderstelling dat het archief homogeen is. Of we dan stukken kiezen uit het begin, dan wel uit het midden, zou op zich dus niet mogen uitmaken. Anderzijds is de kans op ‘gaten’ zeer klein: bij een steekproefverhouding van 1 op 10 is de kans dat er een gat valt van 100 of meer opeenvolgende stukken kleiner dan 1 op 35.000.²

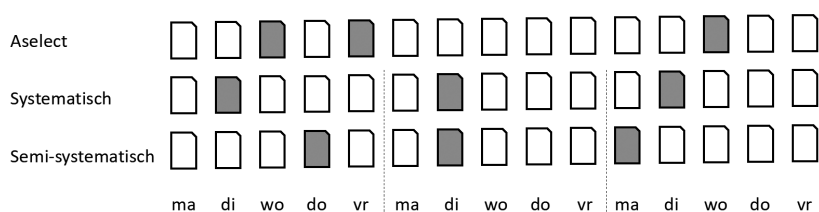
Toch kan het in sommige gevallen handig zijn om lange opeenvolgende periodes waarin er geen stukken geselecteerd zijn a priori uit te sluiten, omdat het archief bijvoorbeeld een tijdsreeks omvat waarin er geen grote periodes zonder informatie mogen vallen. Voor deze toepassing kan een variant van de systematische steekproef gebruikt worden die veel minder last heeft van het *aliasing* probleem. Bij deze variant wordt het archief opgedeeld – opnieuw voor een steekproefverhouding 1 op k – in groepjes van k opeenvolgende stukken. In elk groepje wordt nu telkens een willekeurig stuk geselecteerd, dat voor elk groepje anders is (en dus – in tegenstelling tot de systematische steekproef – niet vastligt op basis van één enkele startpositie). Voor elk groepje heeft men dus een willekeurig getal nodig tussen 1 en k . Deze methode heeft als voordeel dat er maximaal $2 \cdot (k - 1)$ opeenvolgende niet-geselecteerde stukken kunnen zijn (dit is het geval wanneer in één groepje het eerste en in het daaropvolgende groepje het laatste stuk gekozen wordt), wat een grens oplegt aan de maximale grootte van de *missing pockets*.

In tegenstelling tot de puur systematische steekproef heeft deze aangepaste vorm geen periodiek gedrag, zodat mogelijks periodiek gedrag in het archief

1 Terry Cook, *The archival appraisal of records containing personal information: a RAMP study with guidelines* (Parijs: United Nations Educational, Scientific and Cultural Organization, 1991): 64.

2 De willekeurige selectie van archiefstukken met elk een kans p , is een Bernoulli proces. De kans dat n opeenvolgende experimenten negatief zijn, is $(1 - p)^n$. Met $p = 0,1$ en $n = 100$ is deze kans $1/37,649$.

dat gelijkloopt met de steekproefverhouding geen probleem meer oplevert. Deze methode valt echter moeilijker te implementeren in een computerprogramma als Microsoft Excel. Bovendien leent deze methode zich minder voor incrementele selectie.



Figuur 3 Schematische weergave van de enkelvoudige steekproefmethodes

Figuur 3 illustreert tot slot de behandelde enkelvoudige steekproefmethodes, telkens bij een steekproefverhouding van 1 op 5. Bij de aselechte methode heeft elk archiefstuk een gelijke kans; de stukken zijn met andere woorden volledig willekeurig gekozen. Bij de systematische en semi-systematische methodes worden de archiefstukken eerst opgedeeld in groepen van vijf stukken. Binnen de systematische methode wordt enkel de startpositie willekeurig bepaald (hier het tweede stuk) en wordt de steekproef dus gevormd door elk tweede stuk uit alle groepen van vijf te selecteren. Alle gekozen stukken blijken echter op een dinsdag te vallen, wat het risico op *aliasing* aantoont. De semi-systematische methode daarentegen kiest voor elke groep van vijf opeenvolgende archiefstukken één willekeurig stuk en heeft daarom geen last van *aliasing*.

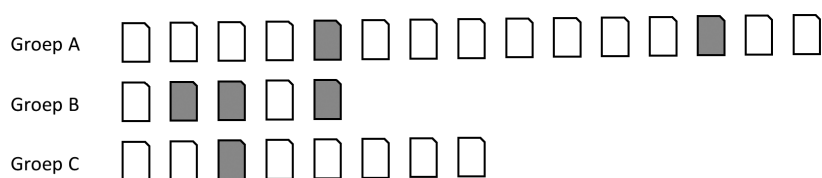
6.2 Gestratificeerde steekproef

In een homogeen archief is er relatief weinig variatie tussen de verschillende archiefstukken en het betrouwbaarheidsinterval zal bij de toepassing van een incrementele selectie relatief snel verkleinen tot een aanvaardbare waarde. In de praktijk is het echter dikwijls zo dat er een aantal sterk verschillende klassen van archiefstukken te onderscheiden zijn, die intern toch homogeen zijn. In het archief kunnen bovendien ook bepaalde stukken aanwezig zijn die weinig voorkomen en bijgevolg weinig kans maken bewaard te worden, ondanks de waardevolle informatie die ze bevatten (voorbeeld: medische dossiers over zeldzame ziekten).

In deze gevallen kan men ervoor kiezen een gestratificeerde aselechte steekproef toe te passen. Hierbij wordt de populatie eerst onderverdeeld in klassen

(groepen of strata),¹ op basis van een door de archivaris/records manager gekozen parameter (voorbeeld: het ziektebeeld). Daarna wordt op elke klasse afzonderlijk een enkelvoudige aselechte steekproef toegepast. Ook het bepalen van de grootte van de steekproef gebeurt voor elke klasse afzonderlijk en kan opnieuw via incrementele selectie gebeuren, daarbij geleid door het betrouwbaarheidsinterval dat in elke klasse bekomen wordt. De steekproefverhouding kan bijgevolg anders zijn voor elke klasse, afhankelijk van de onderlinge variatie van de archiefstukken binnen hun respectievelijke klasse. Bij een goed gekozen classificatieparameter zal elke klasse op zich veel homogener zijn dan de gehele populatie. Zo kan een veel kleinere steekproefverhouding volstaan om toch een representatief beeld te krijgen van elke klasse. Door slim te stratificeren kan het immers zijn dat meerdere kleine steekproeven een hogere betrouwbaarheid hebben dan één grote, zelfs al is het totale aantal geselecteerde dossiers van de verschillende kleine steekproeven kleiner dan het aantal stukken in de grote steekproef.

Belangrijk is wel om tijdens het documenteren van de selectieprocedure duidelijk aan te geven hoe groot elke klasse was binnen het gehele archief en dus welke steekproefverhouding gebruikt werd voor elk van de klassen. Binnen de steekproef zal namelijk een overrepresentatie zijn van stukken uit klassen waarvoor een grotere steekproefverhouding gebruikt werd. Als men bijvoorbeeld de dossiers van alle zeldzame ziekten selecteert en slechts 1/10 van de ‘gewone’ ziekten, dan mag men bij later historisch onderzoek op basis van deze steekproef niet tot de (verkeerde) conclusie komen dat de archiefvormer haast uitsluitend zeldzame ziekten behandelde.



Figuur 4 Illustratie van de gestratificeerde steekproef

Figuur 4 illustreert de gestratificeerde steekproef. Het archief is onderverdeeld in drie groepen: A, B en C. Groep A bevat 15 stukken en is relatief homogeen; hierop wordt een steekproefverhouding van 2/15 toegepast. Groep B is kleiner

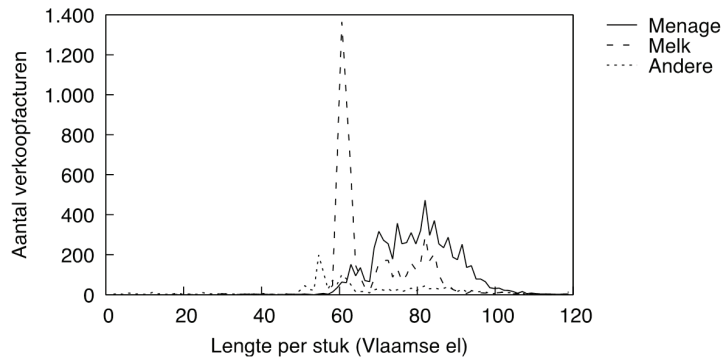
1 Een stratum is een groep uit de populatie die intern homogeen is, maar waarbij de verschillende strata onderling heterogeen kunnen zijn. Som, *A manual of sampling techniques*, 97.

en bevat slechts 5 stukken, maar is veel minder homogeen, waardoor gekozen wordt voor een steekproefverhouding van 3/5. Groep C omvat 8 stukken en is opnieuw zeer homogeen, waardoor er met een verhouding 1/8 geselecteerd wordt. De steekproef zal nu 2 stukken uit groep A, 3 stukken uit groep B, en 1 stuk uit groep C bevatten. Uiteraard moet de originele grootte van elke klasse duidelijk gedocumenteerd worden in het selectiedossier, opdat een latere onderzoeker zou weten dat groep A, en niet groep B het meeste voorkwam, wat uit de relatieve omvang van groep B in de steekproef (foutief) besloten zou kunnen worden.

6.2.1 Casus: de verkoopfacturen van Bethune & Fils

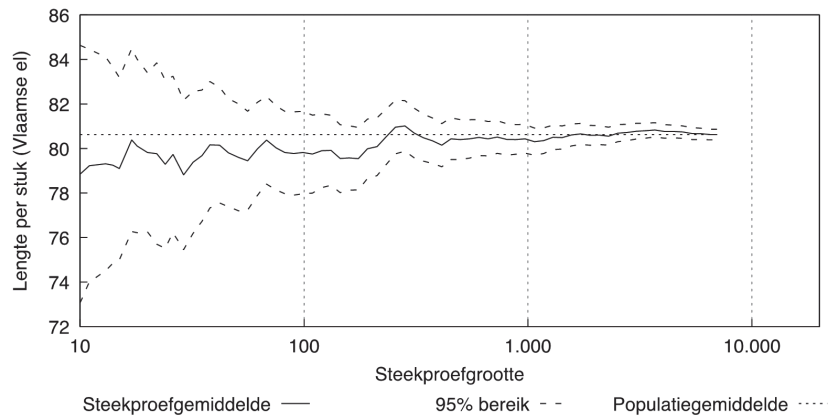
Ter illustratie van de gestratificeerde steekproef door middel van incrementele selectie, keren we terug naar de linnenverkopen van textielhandel Bethune & Fils tussen 1737 en 1799. In Figuur 1 was te zien dat de verdeling van de lengte van de verkochte stukken een scherpe piek vertoonde op 60 el en een bredere piek tussen 70 en 90 el. Men kan dus stellen dat de stukken niet bijzonder homogeen verdeeld zijn, wat zich uit in een relatief groot betrouwbaarheidsinterval wanneer minder dan 1.000 archiefstukken tot de steekproef toegelaten worden. Daarom proberen we een stratificatie van de stukken door te voeren en kiezen we er in dit geval voor om het verkochte linnen, beschreven in de verkoopfacturen, in klassen op te delen. We onderscheiden ‘menage’ linnen (goed voor 46% van alle verkopen), ‘melk’ linnen (43% van alle verkopen), en ‘ander’ linnen (11% van alle verkopen). Deze laatste categorie bestaat uit een combinatie van voornamelijk *écru* en afgewerkte tafellinnenproducten als *nappe* en *serviette*.

Bekijken we opnieuw de verdeling van de lengte, nu opgesplitst voor elk van onze drie klassen, dan bekommen we Figuur 5. Opvallend is dat enkel de verkoopfacturen van de klasse ‘melk’ verantwoordelijk zijn voor de piek rond 60 el, een lengtemaat die voor de andere klassen nauwelijks voorkomt. Anderzijds komt de bredere piek tussen 70 en 90 el voornamelijk voort uit de ‘menage’ archiefbescheiden.

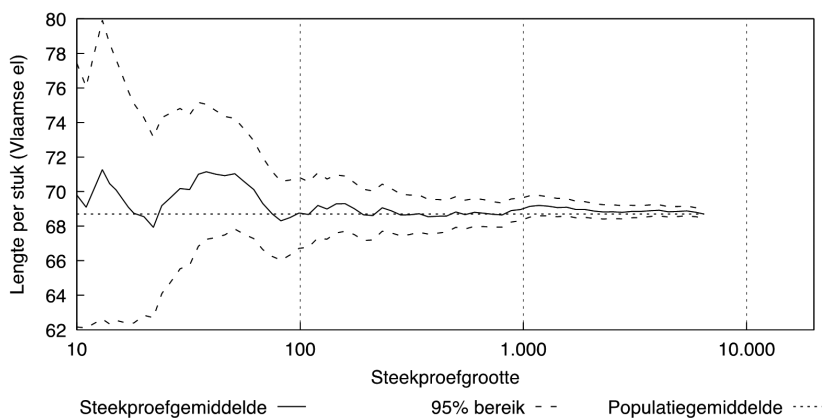


Figuur 5 Verdeling van de stuklengtes, opgedeeld per klasse

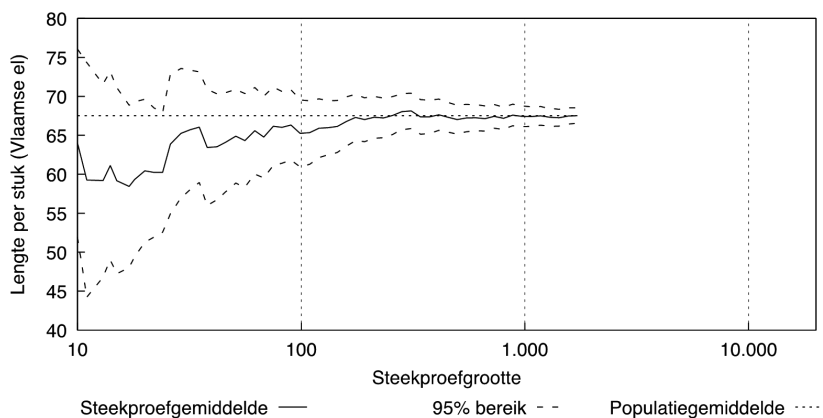
Vervolgens passen we een incrementele selectie toe, nu op elk van de klassen afzonderlijk. We beginnen opnieuw met een steekproefgrootte van 10 verkoopfacturen (per klasse), en laten de grootte van de steekproef toenemen terwijl we voor elke steekproefgrootte het betrouwbaarheidsinterval berekenen.



Figuur 6 Betrouwbaarheidsintervallen voor de verkoopfacturen van menagelinnen



Figuur 7 Betrouwbaarheidsintervallen voor de verkoopfacturen van melklinnen



Figuur 8 Betrouwbaarheidsintervallen voor de verkoopfacturen van ‘ander’ lin-
nen

Stellen we als grens een lengteverschil van 1 el vast voor het betrouwbaarheidsinterval, dan zouden we voor de enkelvoudige aselechte steekproef in totaal 500 archiefstukken moeten bewaren (zie Figuur 2). Bij toepassing van een stratificatie zullen voor de klassen ‘menage’ en ‘melk’ elk een 300-tal stukken nodig zijn, aangezien deze klassen intern minder variatie vertonen. De ‘andere’ klasse is echter veel diverser en heeft bij een selectie van 1.000 stukken nog steeds een betrouwbaarheidsinterval dat een fout van meer dan 1 el aangeeft.

In dit geval kunnen we dus stellen dat de gestratificeerde selectie een grotere algehele steekproef suggereert van in totaal 1.600 stukken, die echter, omwille van de vooraf gekozen klassen, beter in staat zal zijn om voor elk van de verschillende types archiefstukken een zinvolle extrapolatie toe te laten.

7. Aandachtspunten

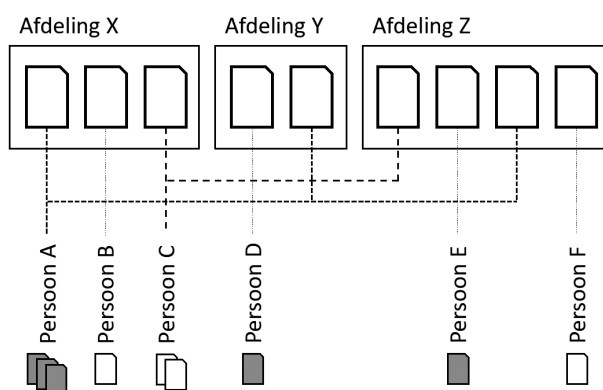
Tijdens de praktische uitvoering van de statistisch verantwoorde steekproef dient de archivaris/records manager rekening te houden met een aantal relatief vaak voorkomende aspecten. Wat dient hij/zij immers te doen wanneer er een onderlinge relatie tussen verschillende archiefstukken van een archief bestaat, wanneer een deel van het archief verloren is gegaan, of wanneer deze de steekproefmethode met andere selectiemethodes wenst te combineren?

7.1 *Gekoppelde archiefstukken*

De essentie van de aselechte steekproef is dat elk archiefstuk een gelijke kans heeft om in de steekproef opgenomen te worden. Vaak bestaat er echter een onderlinge relatie tussen bepaalde stukken, wat het zinvol maakt alle stukken waartussen deze relatie bestaat bij te houden. Men zou in de verleiding kunnen komen om een willekeurig aantal dossiers te kiezen en om vervolgens de overige, hieraan gerelateerde dossiers er eveneens aan toe te voegen. Dit maakt echter dat archiefbescheiden geen gelijke selectiekans meer hebben, wat de steekproef ongeldig maakt. Het is immers zo dat een archiefstuk met veel onderlinge relaties meer kans heeft om deel uit te maken van de steekproef, dan een stuk zonder of met slechts enkele onderlinge banden. Gaan we uit van een steekproefverhouding met kans p voor elk stuk, dan zal een stuk met onderlinge relaties zowel met zijn eigen kans p opgenomen kunnen worden, als ook met een extra kans p voor elk van de met dit stuk gerelateerde archiefstukken. Het betreffende stuk met verwijzingen ('links') naar andere stukken zal in de praktijk dus een kans $p \cdot (l + 1)$ krijgen. Een betere werkmethode voor dit soort archief is het archief eerst op te delen in clusters, waarbij alle stukken die met elkaar verbonden zijn in eenzelfde cluster opgenomen worden. Deze clusters hebben geen onderling verband, waardoor we de steekproef kunnen uitvoeren op het niveau van de clusters in plaats van op het niveau van de individuele archiefstukken. Omdat elke cluster een gelijke kans heeft om gekozen te worden, is de steekproef wel geldig en laat ze toe correct te extrapoleren naar het originele archief.

Een voorbeeld van gekoppelde archiefstukken kunnen we vinden in het archief van een ziekenhuis, waarbij elke afdeling per patiënt een dossier bijhoudt. Wanneer we de

steekproef toepassen op individuele dossiers is het mogelijk dat – voor patiënten die op meerdere afdelingen behandeld werden – bepaalde dossiers van deze patiënt bijgehouden zullen worden en andere niet, wat een nadeel kan zijn voor later historisch onderzoek indien dit een volledig ziektebeeld wil analyseren van specifieke patiënten. Als we de overige dossiers van een patiënt echter zomaar zouden toevoegen, dan zouden de dossiers van patiënten die meerdere afdelingen bezochten oververtegenwoordigd zijn, wat de steekproef niet langer representatief zou maken. Een correcte uitvoering van de aselecte steekproef in dit geval zou eruit bestaan de dossiers eerst te groeperen per patiënt, om vervolgens een selectie door te voeren op basis van patiënten, die elk op hun beurt opnieuw een gelijke kans hebben gekozen te worden.



Figuur 9 Schematische weergave van gekoppelde archiefstukken. De archiefreeks bestaat uit een aantal dossiers van verschillende afdelingen. Eén persoon kan een dossier in meerdere afdelingen hebben. De dossiers worden eerst geclusterd per persoon, waarna een aselecte steekproef wordt uitgevoerd op het niveau van de personen

7.2 Steekproeven op onvolledig archief

Wanneer de archivaris beschikt over een onvolledig archief, kunnen we hierop dan nog steeds een steekproef toepassen? Uiteraard is het mogelijk om de populatie te (her)definiëren als het overgebleven archiefdeel en hierop – via de methodes beschreven in dit artikel – een steekproef toe te passen om dit deel verder te verkleinen. Op basis van de steekproef kan dan later een extrapolatie gemaakt worden naar het overgebleven archiefdeel.

Kunnen we echter stellen dat het onvolledige archief op zich een steekproef vormt van een initieel volledig archief, en – op voorwaarde dat we de grootte van het originele archief kennen – opnieuw extrapolatie kunnen gebruiken om een beeld te krijgen van het volledige archief? Het antwoord zal in de meeste gevallen negatief zijn. Om te kunnen spreken van een geldige steekproef is het immers noodzakelijk dat de selectie (in dit geval het al dan niet verloren gaan) *aselect* gebeurt en elk archiefstuk dus evenveel kans heeft verloren te gaan, zonder systematische effecten en zonder correlatie die een vorm van koppeling introduceert tussen het al dan niet verloren gaan van meerdere stukken.

Wanneer we kijken naar de mechanismen die verantwoordelijk zijn voor het verloren gaan van een deel van het archief, stellen we vast dat deze meestal niet voldoen aan de definitie van de *aselecte* steekproef. Meestal gaat archief verloren in grote groepen, die dikwijls thematisch gekoppeld zijn (voorbeeld: correspondentie met een specifieke contactpersoon die bewaard werd op een andere plaats, of waterschade aan de stukken bewaard op de onderste plank en handelend over patiënten met beginletters T tot Z) wat duidelijk een correlatie introduceert. Ook moedwillige vernietiging, waarbij dikwijls een aantal specifieke archiefstukken verdwijnen, kan moeilijk aangemerkt worden als een *aselect* proces, waarbij elk stuk een even grote kans op bewaring heeft.

7.3 *Combinatie van een aselechte steekproef met andere selectiemethodes*

Tot slot kunnen we opmerken dat het gebruik van een statistisch verantwoorde steekproefmethode het gebruik van andere selectiemethodes niet in de weg hoeft te staan. Het is perfect mogelijk de steekproef die bekomen werd door middel van een enkelvoudige of gestratificeerde *aselecte* steekproef aan te vullen met bijkomende archiefbescheiden, die op basis van andere (subjectieve) criteria geselecteerd werden (voorbeeld: medische dossiers aangaande zeldzame ziekten, of handelend over bekende personen).¹ Voorzichtig zijn is echter de boodschap. Een correct uitgevoerde *aselecte* steekproef is immers representatief voor de populatie en bevat normaliter geen oververtegenwoordiging van bepaalde klassen (of is, in het geval van de gestratificeerde *aselecte* steekproef, duidelijk gedocumenteerd betreffende het voorkomen van verschillende klassen in het originele archief). Voor de archiefstukken toegevoegd op basis van een niet-statistische methode kan men dit echter niet garanderen. In tegendeel, selectie volgens bijvoorbeeld medische uitzonderlijkheid zal vooral dos-

1 Dit volgt de werkwijze van Ketelaar, die start met de 'betrouwbare steekproef' en als 'restcategorie' dossiers van 'historisch belang' toevoegt. F.C.J. Ketelaar, "Archiefselectie en historisch onderzoek," *Nederlandsch Archievenblad* 95 (1991): 119-122.

siers toevoegen die net niet representatief zijn voor de populatie en deze mogen dus niet gebruikt worden in latere extrapolaties door een onderzoeker.

Het is daarom uitermate belangrijk dat de archiefbescheiden van de statistisch verantwoorde steekproef, en die geselecteerd op basis van subjectieve selectiecriteria, volledig afzonderlijk geselecteerd en gedocumenteerd worden. Een intellectuele scheiding kan volstaan, waarbij in een lijst met alle geselecteerde archiefstukken een kolom wordt voorzien die aangeeft of een welbepaald stuk al dan niet tot de statistisch verantwoorde steekproef behoort. Later historisch onderzoek kan dan – op basis van een goed gedocumenteerd selectiedossier – een correcte extrapolatie maken naar de gehele populatie. In het statistisch onderzoek worden dan enkel die stukken in aanmerking genomen die deel uitmaakten van de steekproef; de bijkomend geselecteerde stukken worden dan enkel gebruikt in case studies. Een dergelijke werkwijze maakt veelzijdig historisch onderzoek mogelijk en weerlegt de kritiek die in de literatuur aangetroffen werd over steekproefmethodes, namelijk dat op basis van de onmogelijkheid van anticipatie op toekomstige historische onderzoeksvragen, de praktische uitvoering van een steekproef ernstig belemmerd, zo niet onmogelijk gemaakt wordt.

8. Besluit

Deze bijdrage werd opgesteld als een handleiding voor de archivaris/records manager in relatie tot de statistisch verantwoorde steekproef, waarin de aan deze selectiemethode verbonden theoretische principes en het daarbij noodzakelijke praktische advies, voorbeelden en illustraties ter uitvoering van de betreffende methode samenkomen.

Binnen deze opzet hebben we getracht antwoorden te formuleren en duiding te geven bij aspecten van de methode waarvoor de bestaande literatuur geen of onvoldoende antwoord bood. Een belangrijk onderdeel in dit verband betreft de uiteenzetting over de relatie tussen de grootte en de kwaliteit van de steekproef, die in de bijdrage op wiskundige basis beschreven werd door middel van het betrouwbaarheidsinterval. Via incrementele selectie kan een correcte afweging tussen de grootte en de kwaliteit van de steekproef gemaakt worden. In combinatie met andere factoren zoals beschikbare opslagruimte en budget, zal een dergelijke aanpak een aanvaardbare verhouding tussen de kwaliteit en de grootte van de steekproef bepalen. Ook zal deze aanpak toelaten de aan deze keuze verbonden gevolgen te documenteren om latere extrapolatie op basis van de steekproef geldig te maken.

Bovendien kwamen binnen het scala aan wetenschappelijke steekproefmethodes de voor- en nadelen van elke afzonderlijke methode aan bod, evenals de manier waarop elke methode praktisch uitgevoerd kan worden, met aandacht voor de onderliggende wiskundige principes (Appendix 1) en voor de manier waarop de uitvoering geautomatiseerd kan plaatsvinden, ter ondersteuning van de professional (Appendix 2).

Een laatste cruciale boodschap binnen de bijdrage is dat het nemen van een wetenschappelijk verantwoorde steekproef geen ‘of/of’, maar wel een ‘en/en’ verhaal kan zijn. Selectie op basis van statistische principes kan in combinatie met andere (subjectieve) selectiemethodes en -beslissingen plaatsvinden, zolang er maar een intellectuele scheiding gemaakt wordt tussen elk van de toegepaste selectiemethodes en elk van de genomen beslissingen afzonderlijk gedocumenteerd wordt om later gebruik van en onderzoek op het resterende archief mogelijk te maken.

Het artikel in zijn geheel beschouwd, hebben we getracht aan te tonen dat de statistische steekproefmethode wel degelijk een bruikbaar en waardevol instrument kan zijn in handen van de archivaris/records manager, waarbij we de nodige basis aanreikten om de betreffende methodes correct en met kennis van zaken uit te voeren. Het citaat van Cook aan het begin van deze bijdrage indachtig, hopen we dan ook dat dit artikel de archivariissen/records managers in staat zal stellen de waardevolle schatten van hun archieven te identificeren, teneinde een succesvolle documentbeheersstrategie te kunnen voeren op korte en lange termijn.

9. Bibliografie

Adriaenssens, Annik. “Van laken tot linnen: de textielhandel Bethune & Fils, tweede helft achttiende eeuw: een analyse op basis van het bedrijfsarchief.” 2 vols. PhD diss., Universiteit Gent, 2016.

Baldé, G.M. en H. Spijkerman. “De kunst van het selecteren.” *Nederlandsch Archievenblad* 1 (1990): 302-312.

Boone, Marc. *Historici en hun métier: een inleiding tot de historische kritiek*. Gent: Academia Press, 2007.

Cook, Terry. “‘Many are called, but few are chosen’: appraisal guidelines for sampling and selecting case files.” *Archivaria* 32 (1991): 25-50.

Cook, Terry. *The archival appraisal of records containing personal information:*

a RAMP study with guidelines. Parijs: United Nations Educational, Scientific and Cultural Organization, 1991.

Drossens, Paul. *Archief gewogen: een onderzoek naar de theorie en praktijk van waardering en selectie*. Brussel: Algemeen Rijksarchief, 2011.

Evans, Frank B., François-J. Himly en Peter Walne. *Dictionary of archival terminology*. München: K.G. Saur Verlag KG, 1984.

Hogg, Robert Vincent en Allen Thornton Craig. *Introduction to mathematical statistics*. New York: Macmillan, 1978.

Hull, Felix. *The use of sampling techniques in the retention of records: a RAMP study with guidelines*. Parijs: United Nations Educational, Scientific and Cultural Organization, 1981.

Ketelaar, F.C.J. "De archivaris en de toevloed van moderne archieven: 22^e internationale conferentie van de Table Ronde des Archives." *Nederlandsch Archievenblad* 3 (1984): 134-145.

Ketelaar, F.C.J. "Archiefselectie en historisch onderzoek." *Nederlandsch Archievenblad* 95 (1991): 119-122.

Som, Ranjan Kumar. *A manual of sampling techniques*. Londen: Heinemann Educational Books Ltd, 1973.

van Strien, P.J. "Rationeel beslissen en historische fantasie." *Nederlandsch Archievenblad* 3 (1991): 86-91.

den Teuling, A.J.M. "Steekproeven, een uitdaging tot onderzoek." *Nederlandsch Archievenblad* 1 (1995): 34-43.

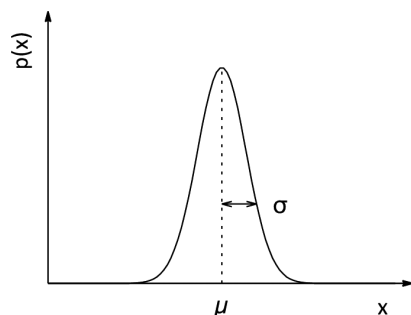
den Teuling, A.J.M. *Archiefterminologie voor Nederland en Vlaanderen*. 's-Gravenhage: Stichting Archiefpublicaties, 2007.

van Tol, Jan. "Verruiming van selectiecriteria en steekproefreductie van massabestanden." *Archievenblad* 8 (1998): 29-33.

10. Appendix 1: Grootte van de steekproef

x	Waarde van de <i>random variable</i> voor een gegeven archiefstuk
x_i	Waarde van de <i>random variable</i> voor het i -de archiefstuk uit de steekproef
μ	Gemiddelde waarde van x over het totale archief
σ	Standaardafwijking van x over het totale archief
n	Grootte (aantal archiefstukken) van de steekproef
\bar{x}	Gemiddelde waarde van x_i over de steekproef
s	Standaardafwijking van x_i over de steekproef

In deze appendix beschrijven we wiskundig hoe we de kwaliteit van een steekproef kunnen uitdrukken aan de hand van de gekozen *random variable*. Initieel gaan we uit van een normale verdeling. Dit wil zeggen dat de kansverdeling $p(x)$ om een bepaalde waarde x uit te komen voor de *random variable*, de vorm heeft van een Bell curve (Figuur 10).



Figuur 10 Kansverdeling van een normale distributie

Bekeken over de *populatie* (het volledige archief) centreert deze curve zich rond de gemiddelde waarde μ en heeft ze een breedte afhankelijk van de standaardafwijking σ , wat een maat is voor de spreiding.

Wanneer we uit deze populatie een steekproef nemen, dan kan van deze steekproef opnieuw de verdeling opgemeten worden, en kunnen we het gemiddelde \bar{x} en de standaardafwijking s berekenen via de volgende formule, met x_i als de verschillende meetwaarden en n het aantal archiefstukken in de steekproef:

$$\bar{x} = \frac{1}{n} \sum x_i$$

$$s = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$$

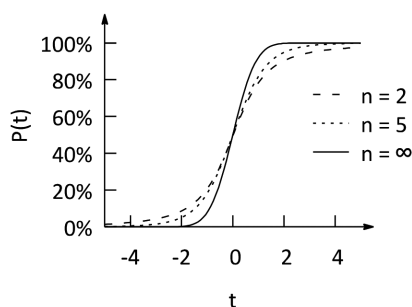
Opdat de steekproef een correcte representatie van de populatie zou weergeven en dus een goede steekproef zou zijn die extrapolatie door toekomstige onderzoekers mogelijk zal maken, willen we dat het gemiddelde van de steekproef \bar{x} het gemiddelde van de gehele populatie μ zo dicht mogelijk benadert. Volgens de centrale limietstelling zal dit gemiddelde \bar{x} opnieuw een normale verdeling volgen,¹ met gemiddelde μ en standaardafwijking σ/\sqrt{n} .²

Dit wil zeggen dat, als de steekproef voldoende groot is, het steekproefgemiddelde dat van de populatie zal benaderen. Bovendien laat het ons toe een kansverdeling op te stellen die aangeeft hoe ver het steekproefgemiddelde van het populatiegemiddelde ligt. Hiervoor kan de t-test gebruikt worden, welke uitgaat van de variabele t , bepaald door:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

De waarde t drukt met andere woorden uit hoe ver het steekproefgemiddelde \bar{x} verwijderd is van een onbekende, ‘correcte’ waarde μ , uitgedrukt als veelvoud van een factor s/\sqrt{n} die uit de steekproef afgeleid kan worden. De variabele t is op zijn beurt verdeeld volgens een Student’s t-distributie (Figuur 11).

- 1 De centrale limietstelling blijft geldig (\bar{x} blijft met andere woorden normaal verdeeld) – zelfs wanneer x zelf geen normale verdeling volgt – zolang de verschillende meetwaarden x_i maar onafhankelijk gekozen zijn.
- 2 Robert Vincent Hogg en Allen Thornton Craig, *Introduction to mathematical statistics* (New York: Macmillan, 1978): 173.



Figuur 11 Cumulatieve kansverdeling van t

In Figuur 11 zien we dat bij bijvoorbeeld $n = 5$, de kans dat $t \leq 2$ ongeveer 95% bedraagt. Wanneer we teruggrijpen naar de definitie van t , wil dit dus zeggen dat de kans dat $\bar{x} \leq \mu + 2 \cdot s/\sqrt{n}$ gelijk is aan 95%, of ook dat er een kans van 5% is dat de fout op het steekproefgemiddelde in positieve richting groter is dan $2 \cdot s/\sqrt{n}$. Meestal wordt de kans in zowel positieve als negatieve richting beschouwd en werkt men met een vooropgestelde totale kans, bijvoorbeeld 99%. Bij een steekproefgrootte van $n = 5$ zien we nu dat $t < -4,03$, met een kans $p = 0,5\%$ en dat $t > +4,03$ voor $p = 99,5\%$. Samen genomen wil dit zeggen dat we met een kans van 99% kunnen stellen dat $|t| < 4,03$.

Tot slot kunnen we de gevonden waarde van t gebruiken om de fout op het steekproefgemiddelde \bar{x} te bepalen. Hiervoor gebruiken we de t-test formule, omdat ze rechtstreeks de fout uitdrukt:

$$|\bar{x} - \mu| < t \cdot s/\sqrt{n}$$

We kunnen ook het betrouwbaarheidsinterval berekenen, ofwel de grenzen waartussen het steekproefgemiddelde \bar{x} zich met de gekozen kans zal bevinden:

$$\mu - t \cdot s/\sqrt{n} < \bar{x} < \mu + t \cdot s/\sqrt{n} \quad (1)$$

De waarde van t hangt af van de gekozen betrouwbaarheid (normaliter 95%) en van de grootte van de steekproef n . Het berekenen van t is complex en werd vroeger meestal in een tabel opgezocht,¹ maar kan tegenwoordig ook in Microsoft Excel berekend worden voor de arbitraire waarden van p en n , door

1 Zie bijvoorbeeld Som, *A manual of sampling techniques*, 339.

middel van de T.DIST (bepaalt de kans voor een gegeven t -waarde) of T.INV functies (wanneer t gezocht wordt, behorende bij een vooropgestelde kans).

Wanneer we de formule van het betrouwbaarheidsinterval opnieuw bekijken en ons afvragen wat we kunnen doen om de steekproef zo nauwkeurig mogelijk te maken, met andere woorden om t zo klein mogelijk te houden, dan zien we dat in deze formule de representativiteit van de steekproef (dus hoe dicht \bar{x} bij μ ligt) afhangt van de absolute grootte van de steekproef n (zowel rechtstreeks via de factor \sqrt{n} , als via de waarde van t), maar ook van de standaardafwijking op de steekproef s – en dus van de intrinsieke variabiliteit van de archiefstukken. Deze laatste factor is uiteraard geen parameter in de steekproef waar de archivaris/records manager vat op heeft, maar wel een externe eigenschap waarmee hij/zij rekening moet houden. Het komt er met andere woorden op neer dat voor archieven met een grote variatie tussen individuele documenten, een grotere steekproef nodig zal zijn om (met voldoende zekerheid) een representatief staal te kunnen overhouden. Uiteraard is dit geen verrassing, maar we kunnen nu wel wiskundig uitrekenen wat de te verwachten afwijking van de steekproef zal zijn. Op basis van deze berekening kan de professional beslissen of de mogelijke fout aanvaardbaar is (gelet op de hoeveelheid die hij/zij wil en praktisch kan bewaren), of dat de steekproef moet worden uitgebreid. De formule geeft met andere woorden een wiskundige uitdrukking voor de afweging tussen de grootte en de representativiteit van de steekproef – een keuze die uiteraard door de archivaris/records manager zelf gemaakt moet worden. Bij deze reiken we hem/haar een instrument aan dat toelaat deze afweging bewust te maken.

Merk op dat in de voorgaande uiteenzetting niet gesproken werd over de grootte van de totale archiefpopulatie of de steekproefverhouding, maar dat enkel de absolute grootte van de steekproef n aan bod kwam. De gebruikte statistiek gaat immers uit van een in essentie oneindig grote populatie, wat de wiskundige formulering sterk vereenvoudigd. Uiteraard is het mogelijk deze uiteenzetting te herhalen met een eindige populatie, ware het niet dat de wiskundige formulering een pak complexer wordt, zodat zelfs de meeste statistici zich er niet meer aan wagen.

Wel kunnen we enkele kwalitatieve uitspraken doen over de effecten van een eindige populatie. Enerzijds gaan voor zeer grote steekproefverhoudingen het verschil tussen de steekproefgrootte n en de populatiegrootte zodanig groot zijn, dat de aanname van een oneindig grote populatie nauwkeurig genoeg is, wat maakt dat het berekende betrouwbaarheidsinterval een goede benadering vormt. Anderzijds is het vanzelfsprekend dat, wanneer de steekproef zeer groot wordt ten overstaan van de populatie (en in extremis alle stukken geselecteerd worden), de fout van de steekproef nagenoeg nul zal zijn. Als we met

andere woorden de grootte van de steekproef gestaag laten toenemen, zullen steeds minder stukken niet geselecteerd worden (in extremis geen – alle archiefstukken werden geselecteerd), wat maakt dat de kans dat een sterk afwijkend stuk niet in de steekproef opgenomen wordt – de steekproeffout – daalt tot nul.

Bij gemiddelde steekproefverhoudingen is de situatie gematigder: het is nog steeds mogelijk om uitschieterende waarden te vinden in het deel van de populatie dat niet weerhouden werd, maar naarmate de steekproefverhoudingen vergroten daalt deze kans. In de praktijk kunnen we stellen dat Formule 1 dus een goede benadering geeft voor grote steekproefverhoudingen. Bij kleinere steekproefverhoudingen schept de formule een ietwat pessimistisch beeld, maar biedt desalniettemin een werkbare ondergrens voor de betrouwbaarheid van de steekproef (de steekproef zal in realiteit dus minstens zo goed zijn als wiskundig voorgesteld).

11. Appendix 2: De steekproef in Excel

Het berekenen van de betrouwbaarheidsintervallen en het toepassen van de selectie is relatief eenvoudig uit te voeren in een rekenblad in bijvoorbeeld Microsoft Excel:

1. De archivaris/records manager dient te beschikken over een lijst van alle archiefstukken, waarbij aan elk stuk een unieke code toegekend wordt, die het stuk identificeert. We gaan ervan uit dat deze code in kolom A opgenomen wordt.
2. Daarnaast hebben we voor elk stuk ook een numerieke waarde nodig (de *random variable*, in kolom B).
3. Vervolgens voegen we een kolom C toe, waarbij in elke cel de formule =RAND() opgenomen wordt, die aan elk stuk automatisch een willekeurig getal tussen 0 en 1 toekent.
4. Om te vermijden dat de willekeurige getallen veranderen als Excel het rekenblad opnieuw doorrekent, moeten we de kolom C kopiëren en *plakken als waarde* (*paste as values*) over de originele kolom.
5. Nu kunnen we het rekenblad sorteren volgens oplopende waarde van het willekeurig getal in kolom C.

Na het sorteren wordt de steekproef gevormd door de bovenste n stukken. De grootte van de steekproef n ligt voorlopig nog niet vast. We kunnen echter wel bepaalde steekproefgroottes simuleren, en voor elke grootte het betrouwbaarheidsinterval berekenen op de volgende manier:

1. Bereken eerst het gemiddelde en de standaardafwijking voor de gekozen steekproefgrootte (ofwel de bovenste n stukken). Voor $n = 100$ geeft dit de formules =AVG(B1:B100) voor het gemiddelde, en =STDEV(B1:B100) voor de standaardafwijking (kolommen E en F).
2. De waarde t kan berekend worden door middel van de t-distributie, de steekproefgrootte n en de onzekerheid (bij een betrouwbaarheid van 95% is de onzekerheid 5%, ofwel 0,05) via de formule =T.INV.2T(0.05, 100-1) (kolom G).
3. De fout wordt gegeven door $t \cdot s/\sqrt{n}$, of in Microsoft Excel met de steekproefgrootte, de standaardafwijking en de t-waarde van cellen D7, F7 en G7: =G7*F7/SQRT(D7) (kolom H).
4. Het betrouwbaarheidsinterval is nu gelijk aan het gemiddelde plus of min de fout =E7-H7 en =E7+H7 (respectievelijk de onder- en bovengrens in kolommen I en J).
5. De relatieve fout kan tot slot bepaald worden aan de hand van =H7/E7 (kolom K).

Merk op dat de waarde van de *random variable* enkel nodig is voor de archiefstukken die deel uit maken van de (potentiële) steekproef. Indien het veel werk

vergt om deze op te meten, is het voldoende de waarde enkel voor de eerste n stukken te bezitten. We kunnen, met andere woorden, beginnen de waarde voor pakweg de bovenste 100 stukken op te meten, de nauwkeurigheid te bepalen en te kijken of die voldoet. Indien deze waarde niet voldoet, verhogen we n naar bijvoorbeeld 200, en vullen we de waarde aan voor stukken 101 tot 200. Op die manier gebeurt niet enkel de selectie, maar ook de analyse van de archiefstukken op incrementele wijze.

	A	B	C	D	E	F	G	H	I	J	K
1	Dossier-nummer	Lengte	Willekeurig getal	Populatie	Gemiddelde	Standaardafwijking	t-waarde	Fout	Ondergrens	Bovengrens	Rel. fout
2	9881	84	0,0001495	n	74,07	13,27		16,13	60,12	92,38	21,16%
3	5982	83	0,0002044	15143				8,09	61,28	77,47	11,66%
4	10898	79	0,0002474					6,16	67,12	79,45	8,41%
5	6922	53,25	0,0003674	Steekproef	Gemiddelde	Standaardafwijking	t-waarde	Fout	Ondergrens	Bovengrens	Rel. fout
6	12498	82	0,0004078	n	76,25	12,99	2,78	16,13	60,12	92,38	21,16%
7	4953	64	0,0005148	5	69,38	11,31	2,26	8,09	61,28	77,47	11,66%
8	3014	62,5	0,0007487	10	73,28	11,13	2,14	6,16	67,12	79,45	8,41%
9	7769	61,5	0,0007685	15	72,61	11,34	2,09	5,31	67,30	77,92	7,31%
10	2118	62,5	0,0008038	20	74,32	11,88	2,01	3,38	70,94	77,69	4,54%
11	6126	62	0,0008259	50	74,35	11,86	1,98	2,35	72,00	76,71	3,16%
12	11963	77	0,0009720	100	73,74	13,52	1,96	0,84	72,90	74,58	1,14%
13	7686	79	0,0009779	1000							
14	13252	88	0,0009906								
15	7287	75,5	0,0010857								
16	9185	86	0,0011589								
17	998	62	0,0012145								
...											
15144	3604	62	0,9997647								

Tabel 1 Voorbeeld van een Excel werkblad voor het uitrekenen van betrouwbaarheidsintervallen bij het uitvoeren van aselecte steekproeven van verschillende groottes

D	E	F	G	H	I	J	K
1	Populatie						
2	n	Standaardafwijking					
3	15143	=AVERAGE(B2:B15144)	=STDEV(B2:B15144)				
4							
5	Steekproef						
6	n	Gemiddelde	Standaardafwijking	t-waarde	Fout	Ondergrens	Bovengrens
7	5	=AVERAGE(B2:B6)	=STDEV(B2:B6)	=T.INV.2T(0,05,D7-1)	=G7*F7/SQRT(D7)	=E7-H7	=H7/E7
8	10	=AVERAGE(B2:B11)	=STDEV(B2:B11)	=T.INV.2T(0,05,D8-1)	=G8*F8/SQRT(D8)	=E8-H8	=H8/E8
9	15	=AVERAGE(B2:B16)	=STDEV(B2:B16)	=T.INV.2T(0,05,D9-1)	=G9*F9/SQRT(D9)	=E9-H9	=H9/E9

Tabel 2 Excel formules gebruikt in het werkblad van Tabel 1